

Univerzita Karlova

Filozofická fakulta

Ústav translatologie

Diplomová práce

Yulia Skadchenko

**Strojový překlad a automatické tlumočení**  
**Machine Translation and Machine Interpreting**

## Poděkování

Ráda bych zde poděkovala PhDr. et Bc. Tomáši Svobodovi, Ph.D. za jeho cenné odborné rady při psaní této práce a za čas strávený jejím konzultováním.

Rovněž děkuji Prof. PhDr. Ivaně Čeňkové, CSc. za její ochotu tuto práci oponovat a za její energii, se kterou své znalosti a zkušenosti předává studentům ÚTRL.

Dále děkuji odbornému konzultantovi ze společnosti Baidu, který si nepřál být jmenován, za jeho spolupráci a ochotu odpovídat na mé dotazy.

Děkuji také všem účastníkům praktického testu, bez kterých by tato práce byla o poznání méně zajímavá, a také rodičům a všem přátelům, kteří mě svými všetečnými dotazy na stav práce donutili ji napsat.

### Prohlášení

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

V Praze dne

podpis

## Anotace

Cílem této diplomové práce je podrobněji prozkoumat oblast strojového překladu a automatického tlumočení, popsat vývoj těchto technologií, jejich současný stav, případnou pozici na trhu a jejich využití. Práce nabízí popis těchto technologií na několika úrovních, a to z pohledu technologického, teoretického, i praktického. Dále popisuje principy jejich fungování, kritéria pro jejich hodnocení, překážky, se kterými se mohou potýkat, a pokouší se vymezit rámec jejich využitelnosti. Součástí práce je praktický test s využitím softwaru Skype Translator od společnosti Microsoft. Cílem testu je zjistit, jaká je uživatelská zkušenost při jeho používání a především zda dokáže zajistit komunikaci mezi dvěma lidmi, kteří nehovoří stejným jazykem.

**Klíčová slova:** strojový překlad, automatické tlumočení, omezení strojového překladu, neurální překlad, Skype Translator, STACL, rozpoznávání řeči, syntéza řeči

## Abstract

This thesis aims to provide an in-depth overview of machine translation and machine interpreting, describing their history, development, and current state, as well as their place on the market and their potential use. The thesis describes machine translation and machine interpreting on theoretical, practical, and technological level, including the description of their basic principles, evaluation criteria, obstacles and challenges, and scope of their use. The thesis also includes a practical test of one of currently available machine interpreting programs – Skype Translator by Microsoft. The aim of the test is to determine whether the program can facilitate successful communication between two people who don't speak the same language, and to describe the user's experience.

**Keywords:** machine translation, machine interpreting, machine translation limitations, neural translation, Skype Translator, STACL, speech recognition, speech production

## Obsah

1	Úvod .....	6
2	Strojový překlad .....	7
2.1	Historie a vývoj strojového překladu .....	7
2.2	Současný stav strojového překladu .....	15
2.3	Aktuální lídři na trhu .....	17
3	Automatické tlumočení.....	20
3.1	Popis a zařazení.....	20
3.2	Princip fungování a nezbytné technologie.....	23
3.2.1	Automatické rozpoznávání řeči.....	23
3.2.2	Syntéza řeči .....	23
3.3	Historie a vývoj automatického tlumočení .....	24
3.4	Aktuálně dostupná řešení .....	27
3.4.1	Mobilní aplikace a software pro počítače .....	27
3.4.2	Přenosná zařízení .....	31
3.4.3	Ostatní zařízení a systémy.....	33
3.5	Automatické tlumočení versus lidské tlumočení.....	34
3.5.1	Model úsilí Daniela Gila.....	35
3.5.2	Model úsilí vs STACL.....	36
3.5.3	Strategie a postupy v simultánním tlumočení .....	37
4	Využití, omezení a hodnocení strojového překladu a automatického tlumočení .....	43
4.1	Využití strojového překladu a automatického tlumočení.....	43
4.2	Omezení automatického tlumočení a strojového překladu .....	47
4.3	Hodnocení strojového překladu a automatického tlumočení .....	50
5	Empirická část práce – praktický test programu Skype Translator .....	57
5.1	Cíl testu a metodologie .....	57
5.2	Hodnocení softwaru Skype Translator z pohledu účastníků testu .....	59
5.3	Hodnocení programu autorkou práce .....	65
5.4	Analýza chyb.....	68
5.5	Závěrečné shrnutí hodnocení .....	69
6	Omezení výzkumu a výhledy do budoucna.....	71
7	Závěr.....	73
8	Bibliografie .....	75
9	Přílohy .....	81

# 1 Úvod

Cílem této diplomové práce je seznámit čtenáře s automatickým tlumočením a strojovým překladem, a to jak se současným stavem, tak i s historickým vývojem těchto technologií. Práce je rozdělena na část teoretickou a část empirickou.

V teoretické části práce se budeme věnovat popisu strojového překladu a automatického tlumočení – popíšeme jejich historii, aktuálně dostupné technologie, možnosti využití těchto technologií i jejich omezení, a kritéria hodnocení jejich výstupů. U automatického tlumočení se pak zaměříme na popis jeho principů a technologií, na které se spoléhá, a pokusíme se na něj rovněž aplikovat některé postupy, které používají profesionální tlumočníci. Rovněž se pokusíme určit, zda je automatické tlumočení vůbec možné za tlumočení považovat, a pokusíme se ho i zařadit z hlediska rozdělení „lidského“ tlumočení na konsekutivní a simultánní.

V empirické části práce pak představíme test s využitím softwaru Skype Translator a jeho výsledky, které budou prezentovány ve formě zpracovaných dotazníků od účastníků testu a rovněž formou hodnocení zpracovaného autorkou práce. V rámci testu je program hodnocen jak z hlediska procesu (hodnotiteli jsou účastníci testu), tak z hlediska výsledného produktu (hodnotitelem je autorka práce). Veškeré materiály související s experimentem budou připojeny ve formě příloh k této práci a v případě digitálního materiálu dostupné na zaheslovaném úložišti.

Úvodem je také nutné říci, že co do zkoumaného tématu je tato práce jedinečná, minimálně pokud jde o vysokoškolské závěrečné práce v českém prostředí. Přestože se nám při průzkumu podařilo najít několik prací zabývajících se například nástroji CAT či různými aspekty strojového překladu – především se jedná o práce hodnotící jeho výstup, případně porovnávající výstupy z více strojových překladačů – nesečkali jsme se s prací, která by se tématu strojového překladu věnovala hlouběji, nemluvě o automatickém tlumočení.

Vzhledem k tomu, že jde o oblast, která se především v posledních letech neustále vyvíjí, museli jsme se potýkat s nedostatkem odborných zdrojů, neboť velká část toho, co bylo o tématu napsáno, již není aktuální, a v případě nejnovějších pokroků naopak veřejně dostupné odborné materiály nejsou k dispozici vůbec či jen omezeně. V této práci proto čerpáme z dostupné literatury, kterou jsme získali vyhledáváním v databázi BITRA a na doporučení odborných pracovníků, z informací od výrobců popisovaných nástrojů a technologií, které jsou často publikovány prostřednictvím takzvaných developer blogů, z konzultací s odborníky a také z vlastních zkušeností.

## 2 Strojový překlad

Strojový překlad (SP) je automatický překlad z jednoho přirozeného jazyka do druhého pomocí výpočetní techniky. Na začátek je třeba zdůraznit, že je nutné rozlišovat mezi strojovým překladem a počítačem podporovaným překladem. Zatímco v druhém případě je stále přítomna osoba překladatele, který používá nástroje počítačem podporovaného překladu (tzv. CAT nástroje) k tomu, aby zefektivnil svou práci, v prvním případě se jedná o zcela automatický počítačový proces, do kterého člověk nijak nezasahuje. Samostatnou kapitolou je pak post-editace, tedy následná úprava počítačem vytvořeného překladu. O té se ještě okrajově zmíníme v kapitole věnované využití strojového překladu.

V první podkapitole se budeme věnovat historii strojového překladu od jeho počátků zhruba do přelomu milénia. Zmíníme také ty nejvýznamnější systémy, které se v průběhu vývoje objevily – některé z nich fungují dodnes a staly se základem pro překladače, které známe a používáme v současnosti.

V oddílu věnovaném strojovému překladu budeme vycházet z informací od významných autorů či z uznávaných publikací – především využijeme dílo Johna W. Hutchinse, odbornou publikaci „Routledge Encyclopedia of Translation Technology“, odborné články publikované na webu Asociace počítačové lingvistiky (ACL), databázi odborných článků ArXiv, kterou spravuje Cornellova Univerzita, výstupy z odborných konferencí, informace od výrobců popisovaných technologií, poskytovatelů jazykových služeb, a další relevantní zdroje a publikace.

### 2.1 Historie a vývoj strojového překladu

V této kapitole se budeme opírat především o informace dostupné v odborné publikaci „Routledge Encyclopedia of Translation Technology“ a dílo Johna W. Hutchinse.

Přestože první myšlenky o strojovém překladu vyslovil už v 17. století René Descartes, první skutečné pokusy o moderní strojový překlad přišly až v roce 1933, kdy byly vydány dva patenty – jeden v Rusku a jeden ve Francii. Stroje, jichž se tyto patenty týkaly, fungovaly spíše jako mechanické slovníky, nicméně právě s nimi se objevily první návrhy a myšlenky o kódování a interpretaci gramatických funkcí s využitím symbolů založených na esperantu.

První myšlenky o využití tehdejší novinky – počítačů – pro účely strojového překladu se objevily po druhé světové válce, v roce 1947. Jako první s nimi přišli britský vědec Andrew D. Booth a Warren Weaver, který pak v roce 1949, když už se myšlenky o automatickém překladu (jak se strojovému překladu tehdy říkalo) objevovaly nezávisle na sobě v různých koutech světa, přišel se čtyřmi základními myšlenkami ohledně strojového překladu, které popsal ve svém memorandu „Translation“.

V roce 1952 pak proběhla na půdě Massachusettského technologického institutu (MIT) první konference věnovaná strojovému překladu. Svolal ji Jehošua Bar-Hillel, který se o rok dříve stal prvním vědcem z MIT, který se naplno věnoval výzkumu strojového překladu. Již tehdy bylo

dosaženo závěru, že plná automatizace překladu a kvalitní výstup jsou naprosto neslučitelné, a že bude proto vždy potřeba lidského zásahu ve formě pre-editace a post-editace. Konference se proto věnovala především ideám a myšlenkám v tomto směru, mezi účastníky se ale našli i skeptici. Jedním z nich byl Léon Dostert, který tvrdil, že pro získání financování nezbytného pro další výzkum oblasti strojového překladu je nutné ukázat veřejnosti, že strojový překlad má potenciál.

K tomu došlo o dva roky později, v lednu 1954. Systém IBM, na jehož vývoji se podílel i sám Dostert, předvedl překlad z ruštiny do angličtiny na vzorku pečlivě vybraných 49 vět, které se skládaly z 250 slov a obsahovaly jen 6 různých gramatických jevů. Přestože tedy nebyl systém příliš sofistikovaný, jeho ukázka přilákala tolik mediální pozornosti, že způsobila boom ve financování výzkumu strojového překladu v USA a inspirovala podobné projekty i v jiných zemích (SSSR).

V období mezi lety 1956 a 1966 se také objevily počátky tří základních přístupů ke strojovému překladu, které převládaly až do nástupu korpusových metod na konci 80. let. Souhrnně se pak tyto tři přístupy řadí do kategorie „pravidlových metod strojového překladu“ (neboli rule-based machine translation).

Prvním byl tzv. model „přímého překladu“ který spočíval ve vývoji programovacích pravidel pro překlad z jednoho konkrétního jazyka do druhého, avšak téměř bez jakékoliv analýzy a bez ohledu na syntax. V praxi to znamenalo zjednodušování bilingvních slovníků v naději, že jeden překladový ekvivalent slova ze zdrojového jazyka pokryje v cílovém jazyce většinu možných situací. Slovosled zdrojového jazyka byl ponechán i v jazyce cílovém, bez jakékoliv obecné sémantické a syntaktické analýzy. Tomuto přístupu se také říká „slovníkový“, jelikož právě slovník je zde tím nejdůležitějším zdrojem.

Druhým přístupem byl tzv. „interlingvní“ model (známý také jako „znalostní přístup“<sup>1</sup>). Ten se zakládá na tom, že pro všechny cílové i zdrojové jazyky je definováno jakési univerzální vyjádření, neboli interlingva (jinak také metajazyk, ústřední jazyk), a překlad tak probíhá ve dvou krocích: prvním je převod ze zdrojového jazyka do metajazyka, druhým převod z metajazyka do jazyka cílového. Tento metajazyk se skládá například ze symbolů nebo kódů, ale může to být i umělý nebo přirozený jazyk. Interlingvní model je považován za nejlepší pro systémy strojového překladu, které pracují s více než dvěma jazyky a můžeme ho vidět i u současných překladačů, jako je například Bing, které jako metajazyk spojující desítky jiných jazyků nejčastěji volí angličtinu.

Posledním byl tzv. „transferový přístup“. U toho byl v převodu přítomen krok tzv. konverze, kdy je pro oba jazyky definován systém přechodových struktur. Překlad pak probíhá ve třech krocích: prvním je analýza zdrojového textu a převod do přechodové struktury zdrojového jazyka, poté následuje převod mezi přechodovou strukturou zdrojového jazyka do přechodové struktury cílového jazyka, a z této je nakonec vytvořen cílový text.

V průběhu 60. let se výzkum v této oblasti nadále rozvíjel a rozšiřoval i za hranice USA a SSSR a v polovině 60. let už byly výzkumné skupiny rozšířené po celém světě – existovaly téměř ve všech zemích Evropy, ale také v Číně, Mexiku či Japonsku. Přístupy k výzkumu byly v té době

---

<sup>1</sup> Z anglického „knowledge-based approach“



rozdělené především na empirické, které k hledání gramatických a lexikálních pravidelností používaly především statistické metody, a teoretické, které položily základ současné počítačové lingvistiky. Mnoho vědců a výzkumníků tehdy navíc nemělo příliš důvěry k soudobým lingvistickým teoriím, a proto hledali a vyvíjeli vlastní metody pro fungování strojového překladu založené například na korpusové analýze. Oproti optimismu 50. let však bylo zřejmé, že překážek na cestě k plně automatizovanému a kvalitnímu překladu, který by byl k nerozeznání od lidského, je stále víc a víc – hovořilo se o tom, že výzkum narazil na nepřekonatelnou „sémantickou bariéru“. V roce 1960 Jehošua Bar-Hillel snahy o vytvoření dokonalého automatizovaného systému pro překlad kritizoval – tvrdil, že je z principu nemožné, aby strojový překlad někdy dosáhl stejné kvality, jako překlad vypracovaný člověkem, a doporučoval, aby si výzkum kladl o něco méně ambiciózní cíle. Výzkum strojového překladu však pokračoval dál a začaly se objevovat první funkční systémy – ty s povděkem kvitovali uživatelé, pro které byly hlavní především okamžité výsledky, nikoliv kvalita.

Vidina kvalitního strojového překladu se však stále více vzdalovala a v roce 1964 byl v USA založen ALPAC, neboli poradní výbor pro automatické zpracování jazyka, jehož cílem bylo prozkoumat vyhlídky bádání v oblasti strojového překladu. V roce 1966 pak tento výbor předložil zprávu, která mimo jiné došla k závěru, že strojový překlad je v porovnání s překladem lidským pomalejší, méně přesný, dvakrát dražší, a že „v současné době neexistují žádné vyhlídky na užitečný strojový překlad“. Zpráva, která dále doporučovala přesun financování z výzkumu strojového překladu do oblastí jako základní lingvistický výzkum či vývoj pomůcek pro překladatele, měla na celou oblast obrovský dopad. Jejím výsledkem bylo prakticky úplné zastavení výzkumu strojového překladu v USA na celé desetiletí, a zpomalení či omezení výzkumu v jiných zemích, například v SSSR.

I přes útlum způsobený zprávou ALPAC však výzkum v 70. letech pokračoval dál, přestože se poněkud změnil a přizpůsobil se potřebám konkrétních zemí. V USA byl prioritou překlad mezi ruštinou a angličtinou, v Kanadě pak mezi francouzštinou a angličtinou, a v Evropě začínala čím dál víc sílit potřeba po překladech mezi všemi evropskými jazyky. Hlavními centry výzkumu strojového překladu se tak staly právě Kanada a Evropa. Změnu zaznamenal i přístup ke strojovému překladu – stále více se upouštělo od modelu „přímého překladu“, který dominoval mezi lety 1956 a 1966, a místo toho se výzkum stále více přesouval k „nepřímým“ modelům založeným na metajazycích či konverzi. V polovině 70. let však začal být interlingvní přístup zpochybňován kvůli jeho rigiditě – jedna chyba v analýze totiž znamenala celkové selhání překladu – a ztrátě informací o povrchové formě textu ve zdrojovém jazyce, které by mohly pomoci s vhodným výběrem výrazů a vhodnou větnou konstrukcí v cílovém jazyce.

V této době také vznikly první významné systémy – jmenovat můžeme hned několik.

V Kanadě byl v roce 1976 výzkumnou skupinou TAUM (Traduction Automatique à l'Université de Montréal) vyvinut prototyp systému pro překlad předpovědí počasí **Méteo**. Kanadské ministerstvo životního prostředí pak systém používalo od roku 1981 až do roku 2001, kdy ho nahradil konkurenční produkt. Díky tomu, že byl systém navržen speciálně a pouze pro tento účel, tedy s ohledem na omezenou slovní zásobu i syntax předpovědí, dosahovala kvalita výstupu v roce 1985 hodnoty 85 % (procentuální hodnocení je založeno na počtu zásahů, které do výstupu vnášeli lidští editoři). Zajímavá je také informace o ceně překladu s využitím tohoto

systemu – díky nízkému počtu lidských zásahů do výstupu (publikace S. Nirenburga „Progress in Machine Translation“ uvádí, že se revize dotýkaly dokonce pouhých 4 % výstupů) byl systém vysoce produktivní a zároveň tedy extrémně levný – průměrné náklady na přeložení 1000 slov činily pouhých 5 USD<sup>2</sup>.

Dalším příkladem je **Systran** – systém, který vznikl v roce 1968 a překládal mezi ruštinou a angličtinou, později také mezi francouzštinou a angličtinou a postupně i v dalších jazykových párech. První verze Systranu byla letectvem Spojených států amerických využívána již od roku 1970. S americkým ministerstvem obrany pak společnost Systran dodnes spolupracuje v oblasti vývoje nových jazykových párů, v souladu s geopolitickými potřebami. Kromě amerického letectva využívá tento software také například NATO či velké společnosti jako General Motors, Xerox a další. Dříve se mezi uživatele Systranu řadila také Evropská komise, která však později přešla na systém MT@EC založený na toolkitu MOSES a v současnosti využívá systém s názvem eTranslation – ten mohou bezplatně využívat orgány státní správy členských zemí EU, Islandu a Norska. Dnes patří Systran k jedním z hlavních a nejstarších hráčů na poli strojového překladu – z původně armádního softwaru je dnes komerční produkt, o kterém se ještě rozepíšeme v kapitole věnované aktuálně dostupným technologiím strojového překladu.

Dalším systémem, který stojí za zmínku, je **Logos**, donedávna hlavní konkurent Systranu, který ve svých počátcích sloužil k překladu mezi angličtinou a vietnamštinou a byl hojně využíván v 70. letech. I on navíc přetrval až do dnešních dnů v podobě svého nástupce OpenLogos, který byl vyvinut pro operační systémy založené na linuxovém jádru a dokáže překládat z angličtiny a němčiny do většiny „velkých“ evropských jazyků.

Od poloviny 70. let navíc začal výzkum v oblasti strojového překladu znovu ožít po útlumu, který způsobila zpráva ALPAC z roku 1966. V 80. letech vznikla takzvaná druhá generace transferových systémů založených na lingvistických principech, jejichž „matkou“ se stal systém s názvem Ariane, který se vyznačoval flexibilitou a modulárností. Přestože se systém Ariane nikdy nedočkal uvedení do provozu a jeho vývoj byl na konci 80. let ukončen, stal se v 90. letech součástí projektu Eurolang.

Druhá polovina 80. let se vyznačovala obnoveným zájmem o tzv. interlingvní systémy, se kterými se pojilo využívání znalostních metod vycházejících z výzkumu umělé inteligence. Právě do té mnozí výzkumníci vkládali velké naděje v očekávání, že jazykový výzkum v kontextu umělé inteligence přinese zvýšení kvality strojového překladu. Větší pozornost byla věnována také vytváření specializovaných systémů pro strojový překlad, které se měly zaměřovat například na konkrétní obor či sloužit specifické skupině uživatelů.

V průběhu 80. let bylo v oblasti strojového překladu obzvlášť aktivní Japonsko, kde společnosti jako Hitachi, Toshiba či Fujitsu vyvíjely software pro strojový překlad. Ten však spoléhal na značnou pre-editaci vstupních textů a post-editaci výstupů a byl často úzce oborově zaměřený. V 80. letech také ožil výzkum v SSSR, který probíhal především na půdě „Všesvazového centra pro překlad“ v Moskvě. Vznikaly zde systémy založené na přímém modelu překladu (AMPAR pro překlad mezi angličtinou a ruštinou, NERPA pro překlad mezi němčinou a ruštinou), ale i

---

<sup>2</sup> NIRENBURG, Sergei. Progress in machine translation. Tokyo: Ohmsha, 1992. ISBN 978-90-5199-074-4.

systemy založené na teorii „smysl-text“ profesora Melčuka – ten však v roce 1977 emigroval do Kanady, a výzkum v této oblasti proto dále vedl profesor Apresjan.

Celkově byla 80. léta charakterizovaná boomem nejrůznějších funkčních systémů pro strojový překlad. Kromě obecně zaměřených systémů, jako byly Logos nebo Systran, vznikaly v 70. a 80. letech i systémy úzce specializované. Jako příklad můžeme uvést již zmiňovaný systém Méteo a také SPANAM a ENGSPAN – dvojici programů pro překlad ze španělštiny do angličtiny a z angličtiny do španělštiny, která byla vyvinuta Panamerickou zdravotnickou organizací. Oba programy vyžadovaly lidskou post-editaci, ale nikdy nepoužívaly pre-editaci, s výjimkou kontroly formátování vstupních dokumentů.

Dalším významným projektem byl systém Mu, který vznikl na Kjótské univerzitě a který uměl mimo jiné analyzovat gramatické pády a pracovat se závislostními stromy. V roce 1986 přestal být program pouhým prototypem a stal se funkčním softwarem, který používalo Japonské informační centrum pro vědu a techniku.

Ambiciózním počinem byl také projekt Evropských společenství **Eurotra**, jehož cílem bylo vytvoření pokročilého vícejazyčného překladového systému, který měl pokrýt všechny jazyky Evropských společenství. Přestože byl tento projekt na konci 80. let z důvodu absence funkčního prototypu ukončen, podařilo se mu naplnit alespoň svůj druhotný cíl – tím byla podpora mezinárodního výzkumu v oblasti počítačové lingvistiky, a navíc zájem o tuto oblast podpořily i nové poznatky z oblasti umělé inteligence. Dvěma nejvýznamnějšími systémy, které díky tomu vznikly, byly DLT, který jako „mezijazyk“ používal upravenou formu esperanta, a Rosetta, která se pokoušela aplikovat do strojového překladu tzv. Montagueho gramatiku (tedy tezi, že je možné stejným způsobem přistupovat k přirozenému i formálnímu jazyku, a že mezi syntaxí a sémantikou existuje systémový vztah).

V 90. letech se vývoj strojového překladu posunul od snahy vytvořit dokonalou imitaci lidského překladu k vývoji systémů, které naplňovaly specifické požadavky určitých odvětví. Strojový překlad se začal rozšiřovat stále rychleji, především pak v prostředí komerčních agentur, nadnárodních společností a vládních struktur, kde byla vysoká poptávka po velkoobjemových překladech především v oblasti technických textů a dokumentace – právě tyto firmy a instituce se tak staly hlavním trhem pro dříve zmiňované systémy jako Systran či Logos. Rapidně vzrůstal i objem strojově přeloženého textu: odhady hovoří o tom, že již v roce 1995 se v těchto společnostech a organizacích strojově přeložilo zhruba 300 milionů slov ročně<sup>3</sup>. Významně k rozvoji strojového překladu přispěl svými vysokými požadavky na rychlost překladu dokumentace i softwarový průmysl – dokumentace totiž musela být dostupná ve chvíli vydání softwaru a díky tomu, že tyto texty obsahovaly velké množství opakování, byl strojový překlad tím nejlepším řešením.

Snadnější dostupnost osobních počítačů a mikropočítačů navíc otevřela pro strojový překlad nový komerční trh, na který mohly vstupovat levné systémy strojového překladu, a trh tak byl

---

<sup>3</sup> HUTCHINS, John. Machine translation: a concise history. In: John Hutchins. Publications on machine translation, computer-based translation technologies, linguistics and other topics [online]. [cit. 2019-04-11]. Dostupné z: <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>

zaplaven celou řadou programů a „gadgetů“ od firem z celého světa, především ze Severní Ameriky, Evropy, Japonska, nebo Číny.

Příchod 90. let s sebou přinesl, kromě rapidního rozšiřování strojového překladu, také nové metody. Zatímco až do konce 80. let byly systémy strojového překladu v zásadě postaveny na nejrůznějších lingvistických pravidlech, v 90. letech se začaly objevovat nové metody a přístupy, které se dnes souhrnně označují jako „korpusové“. Nejvýraznějším posunem bylo oživení statistických metod, které byly používány už v počátcích strojového překladu, ale nepřinášely v té době uspokojivé výsledky. Průkopníkem se v této oblasti stal projekt IBM s názvem **Candide** který byl výjimečný tím, že k analýze i produkci používal výhradně statistické metody, bez jakýchkoliv lingvistických pravidel. Statistické metody totiž fungují pouze na základě analýzy paralelních textů (čím víc, tím lépe) ve dvou různých jazycích, která slouží k rozpoznání a pochopení nejrůznějších vzorců. Neřídí se žádnými pravidly a veškeré závěry činí počítač, bez lidského zásahu, pouze na základě statistik. Když byl pak systém Candide testován pomocí obsáhlého francouzsko-anglického korpusu, výsledky byly skutečně překvapivé – téměř polovina vět buď přesně odpovídala větám v korpusu, předávala stejný význam za použití trochu jiných slov, nebo nabízela jiný, avšak zcela legitimní překlad. Právě tento úspěch zapříčinil to, že se statistický strojový překlad stal hlavním předmětem bádání většiny vědců. Zatímco dříve se největší význam přikládal korelaci mezi slovy, v 90. letech došlo v tomto směru k posunu ke korelaci mezi „frázemi“ (tedy sekvencemi slov, nemuselo nutně jít o věty v tradičním smyslu). V současnosti jsou hlavními centry výzkumu statistických metod strojového překladu univerzity v Edinburghu a Cáchách a univerzita Jižní Kalifornie. Mezi nejambicióznější projekty v oblasti statistického strojového překladu patřil **Euromatrix** (a jeho nástupce EuroMatrixPlus), který byl financován Evropskou unií. Tento projekt byl spuštěn v roce 2006 a jeho cílem bylo vyvinout systémy strojového překladu na statistické bázi (nicméně vzhledem ke komplexnosti problematiky projekt nevyklučoval využití jiných přístupů, pokud se ukázaly jako vhodnější) mezi všemi jazyky EU, obzvláště pak pro tzv. „malé“ jazyky jako je estonština, slovinština a další. Původní projekt Euromatrix trval do roku 2009 a poté na něj navázal EuroMatrixPlus, který trval až do roku 2012. Oba projekty přinesly řadu nástrojů, například program Moses, korpus zápisů z jednání Evropského parlamentu od roku 1996 až do současnosti Europarl Corpus, nebo česko-anglický korpus CzEng 0.7, který vznikl na půdě Ústavu formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy.

Druhým přístupem, který dominoval výzkumu v 80. a 90. letech, je tzv. „příkladový“<sup>4</sup> přístup, který se na konci 80. let rozšířil z Japonska, kde ho už v roce 1981 poprvé navrhl Makoto Nagao – ten mimo jiné vedl také již zmiňovaný projekt Mu, první funkční systém strojového překladu mezi angličtinou a japonštinou. Základem příkladového přístupu je myšlenka, že člověk při překladu rozebírá větu na jednotlivé fráze, ty překládá, a nakonec tyto kousky řadí zpět tak, aby výsledná věta dávala ten správný smysl, přičemž tyto fráze jsou překládány na základě analogie s předchozími překlady. Stejně pak při použití příkladového přístupu postupuje i počítač, který své překlady zakládá na databázi překladů vytvořených profesionálními lidskými překladateli. Hlavním problémem je zde právě ono seřazení frází zpět do koherentního celku, hlavní výhodou (ve srovnání s metodami založenými na pravidlech<sup>5</sup>) je pak zajištění

---

<sup>4</sup> Z anglického „example-based“ approach

<sup>5</sup> Z anglického „rule-based“ approach

idiomatickosti překladu právě díky tomu, že jako základ slouží překlady vyprodukované lidmi. V porovnání s metodami založenými na pravidlech, které vyžadovaly nadefinování velkého množství pravidel a výjimek, aby vůbec mohly fungovat, se však jednalo o průlomovou novinku.

Další novinkou 90. let byl také takzvaný „řízený jazyk“. Jde o metodu, kterou používaly především velké společnosti disponující specializovaným softwarem pro strojový překlad, jejichž hlavním cílem bylo pomocí strojového překladu vytvářet publikovatelné texty. To ale až do té doby znamenalo, že výstup musel být vždy kontrolován a revidován lidskými překladateli či editory se znalostí obou jazyků. Tzv. „řízený jazyk“ byl používán s cílem zlepšit kvalitu výstupu strojového překladu a znamenal nastavení určitých omezení v oblasti slovní zásoby nebo například větné skladby, aby se co možná nejvíce předešlo možné mnohoznačnosti. Společnost Xerox tento přístup používala již od konce 70. let a vznikly dokonce společnosti, které se na nastavení tohoto „řízeného jazyka“ pro jiné firmy přímo specializovaly a specializují dodnes.

V průběhu 90. let se také objevily první systémy strojového překladu pro osobní počítače, především díky jejich zvyšujícímu se výpočetnímu výkonu. Mezi první volně prodejné software tohoto druhu patřily například programy PC-Translator nebo Power Translator, nicméně postupně se jich na trhu objevily desítky a začaly se objevovat i systémy pro do té doby opomíjené jazyky. Jako příklad můžeme uvést systémy ProMT, Transcend, ruský Stylus, japonský ATLAS, LogoVista, Sakhr pro arabštinu, Transtar pro čínštinu a mnoho dalších. Stejně tak se začaly pomalu objevovat i systémy pro překlad jazykových párů bez angličtiny, například mezi italštinou a španělštinou či japonštinou a čínštinou. Je však nutné zmínit, že zatímco výkon počítačů se od jejich vzniku neustále zvyšoval, o kvalitě strojového překladu to již říci nelze, přestože co do výkonu tyto komerční systémy z 90. let často své předchůdce předčily. Měly ale i řadu nedostatků – většina jich stále fungovala na zastaralém principu transferového nebo dokonce přímého překladu a jejich slovníky byly povětšinou nedostačující. Kromě nových systémů se však pro domácnosti a jednotlivce staly dostupné i již zaběhnuté systémy, které byly dříve dostupné pouze pro firmy a organizace, například Systran, SPANAM nebo ENGSPAN. Navíc se na trhu objevily i takzvané kapesní překladače, které byly ve své podstatě elektronickou verzí papírových konverzačních příruček – ty však v dnešní době ustoupily do pozadí a nahradily je online překladače a chytré telefony.

Od poloviny 90. let se o slovo ve strojovém překládání začal hlásit i internet a objevily se první online překladače, které zpočátku fungovaly na základě různých forem poplatků a předplatného. Nejstarším představitelem online překladačů tak, jak je známe dnes – tedy zadarmo a s okamžitými výsledky – se stal v roce 1997 překladač **Babel Fish**<sup>6</sup>, který byl součástí tehdy oblíbeného vyhledávače AltaVista. Tento program navíc existuje dodnes, přestože byl nejdříve převeden pod společnost Yahoo! a poté pod Microsoft, který ho ale v roce 2012 nahradil překladačem Bing. V době svého prvotního fungování podporoval systém Babel Fish 13 jazyků v 36 různých kombinacích.

---

<sup>6</sup> Jméno překladače pochází z knihy spisovatele Douglase Adamse „Stopařův průvodce po galaxii“. V knize jde o fiktivního živočicha – rybku – která dokáže tlumočit mezi jakýmkoliv dvěma jazyky v galaxii; stačí si žlutou rybku pouze vložit do ucha.

Po úspěchu Babel Fish na sebe konkurenční programy nenechaly dlouho čekat – téměř každá společnost disponující „offline“ překladačem ho začala nabízet i v online formě. Pro spoustu uživatelů navíc představovaly online překladače naprostou novinku, se kterou se do té doby neměli možnost setkat, což znamenalo i poněkud skeptický přístup. Nedůvěřiví uživatelé si proto tyto programy s oblibou „zkoušeli“ na idiomatických, neurčitých, či víceznačných frázích, nebo dokonce na příslovích, či metodou „zpětného překladu“. Není proto překvapením, že nedostatky strojového překladu rychle objevili. I přesto je ale nutné říci, že čím menší znalost výchozího jazyka uživatel má, tím větší hodnotu pro něj výstup strojového překladu představuje a často mu umožní pochopit alespoň základní „jádro“ textu.

Za dobu své existence urazil strojový překlad značnou cestu a díky novým technologiím se vyvíjí i nadále (v posledních letech vzbudily velkou pozornost pokroky společnosti Google v oblasti využití neuronových sítí a hlubokého učení pro strojový překlad). V současné době je již zřejmé, že pro naplnění různých potřeb jsou nezbytné různé druhy strojových překladačů. Dnes tedy proto existují „tradiční“ strojové překladače, které jsou využívány velkými společnostmi a které jsou většinou chráněné, dále počítačem podporovaný překlad s možností přidávání modulů pro strojový překlad, nebo volně dostupné online překladače pro běžné uživatele, které jim slouží pro každodenní potřeby (například Google Translate). Současnému stavu strojového překladu se budeme věnovat v další kapitole této práce.

Poslední část kapitoly o historii bychom však ještě rádi věnovali významné české instituci, kterou je **Ústav formální a aplikované lingvistiky Univerzity Karlovy**. Ten vznikl v roce 1990, navazuje však na několik předchůdců, konkrétně na Oddělení algebraické lingvistiky a strojového překladu, které v roce 1959 založil profesor Petr Sgall na Filozofické fakultě Univerzity Karlovy, a na Laboratoř algebraické lingvistiky, která vznikla v roce 1968, nicméně v tomtéž roce byla zrušena a převedena pod Matematicko-fyzikální fakultu. Na půdě ÚFALu vznikl také Pražský závislostní korpus, který „*obsahuje velké množství českých textů doplněných rozsáhlou a provázanou morfologickou (2 milióny slovních jednotek), syntaktickou (1,5 miliónu slovních jednotek) a sémantickou (0,8 miliónu slovních jednotek) anotací; na sémantické rovině jsou navíc anotovány aktuální členění věty a koreferenční vztahy.*“<sup>7</sup> (Pražský závislostní korpus)

Tento korpus se využívá například jako základ pro programy zaměřené na strojový překlad či vyhledávání informací. V současné době existuje tento korpus v několika verzích<sup>8</sup>.

Profesor Sgall je pak autorem tzv. funkčně generativního popisu jazyka, který poprvé formuloval v roce 1967. Mezi další významné osobnosti české počítačové lingvistiky patří například profesorka Eva Hajičová, profesorka Jarmila Panevová, či profesor Jan Hajič.

V současné době spolupracuje ústav například s Univerzitou Jižní Kalifornie, a to na projektu MALACH, který poskytuje přístup k digitálním archivům historických rozhovorů. V roce 2016 byl Ústav nominován na cenu Innovation Radar Prize, za kterou stojí Evropská komise. Cílem této iniciativy je vyhledávat a podporovat inovace s vysokým potenciálem. ÚFAL byl na tuto

---

<sup>7</sup> Pražský závislostní korpus 2.0. Home page | ÚFAL [online]. Copyright © 2006 ÚFAL [cit. 11.03.2019]. Dostupné z: <https://ufal.mff.cuni.cz/pdt2.0/index-cz.html>

<sup>8</sup> Více na <https://ufal.mff.cuni.cz/pdt3.0>

cenu jmenován díky účasti na mezinárodním projektu KConnect, který se zaměřuje na vytváření systémů pro vyhledávání v lékařských textech.

## 2.2 Současný stav strojového překladu

Za více než 50 let vývoje urazil strojový překlad skutečně dlouhou cestu, od metod založených na pravidlech, přes příkladové a statistické metody, až k nejmodernějším metodám využívajícím neuronové sítě. Některé metody zmiňované v předchozí kapitole najdeme v softwaru pro strojový překlad i dnes – základem Google Translate je dodnes odnož statistické metody překladu založená na frázích, a než Google nebo Microsoft vyvinuli vlastní technologie pro všechny nabízené jazyky, používaly obě společnosti Systran. Naopak některé, jako například metoda přímého překladu, skončily na odkladišti dějin. V této kapitole se zaměříme především na poslední průlom v oblasti strojového překladu, tedy na neuronové sítě, a pokusíme se popsat jejich fungování a vznik v kontextu strojového překladu.

Počátky využití těchto technologií ve strojovém překladu najdeme v roce 2014, kdy vyšla studie o potenciálním využití neuronových sítí ve strojovém překladu. Autoři v ní navrhují využití rekurentních neuronových sítí (RNN) ve strojovém překladu – konkrétně jde o nový model „RNN Encoder-Decoder“, jehož využití společně se statistickými metodami strojového překladu by podle autorů značně zlepšilo jeho kvalitu. Hlavní výhodou RNN je totiž to, že si pamatují i předchozí výsledek, v případě překladu tedy předchozí slovo, což je vzhledem k roli kontextu v rámci každého jazyka naprosto zásadní, a postupem času se RNN technologie naučily pamatovat si i slovo následující po „hlavním výrazu“, čímž se jejich efektivita při překladu ještě zvýšila. V popisu principu, na kterém v článku popisovaný RNN Encoder-Decoder model funguje, pak nelze nespátřit podobnost s interlingvním přístupem, který se také pokoušel o vytvoření univerzálního metajazyka, avšak nebyl ve své době příliš úspěšný. Ve studii je pak fungování nového systému popsáno takto: Navrhovaná architektura neuronových sítí se skládá ze dvou rekurentních neuronových sítí, které fungují jako dvojice kodéru a dekodéru. Kodér zakresluje zdrojovou sekvenci o proměnlivé délce do podoby vektoru o pevně dané délce a dekodér pak zakresluje toto vektorové zobrazení do cílové sekvence o proměnlivé délce. Obě sítě se společně učí tak, aby byla maximalizována podmíněná pravděpodobnost hledané sekvence a zdrojové sekvence.<sup>9</sup>

Studie sice nezbudila žádnou větší pozornost veřejnosti, všiml si jí však Google, který na nic nečekal a pustil se do práce, aby o dva roky později, v roce 2016, představil Google Neural Machine Translation, tedy svůj neuronový strojový překladač založený právě na RNN. Prvním spuštěným jazykovým párem se navíc stala dvojice angličtina-čínština, která je pověstná svou náročností pro úspěšný strojový překlad; kombinace čeština-angličtina se pak dočkala podpory v roce 2017. Přestože Google v článku na svém blogu věnovaném novinkám ve světě umělé inteligence sám připouští, že strojový překlad ještě zdaleka není dokonalý a stále je schopen dopouštět se chyb, které se lidskému překladateli stát nemohou, jedná se bezpochyby o velký

---

<sup>9</sup> Srov. CHO, Kyunghyun et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation [online]. 2014 [cit. 2019-04-11]. Dostupné z: <https://arxiv.org/pdf/1406.1078.pdf>

krok vpřed. Za pouhé dva roky přinesl neurální překlad větší pokrok, než předchozích 20 let výzkumu – Google uvádí, že GNMT (Google Neural Machine Translation) snížil počet chyb při překladu ve „velkých“ jazykových párech o 55–85 %. Nástup neurálního překladu také znamenal průlom v překladu jazyků, mezi kterými neexistují společné slovníky – dříve totiž takové překlady probíhaly přes angličtinu (opět se vracíme k interlingvní metodě), což však znamenalo víc příležitostí pro chyby. Neurální překlad však tento přechodový jazyk nepotřebuje a k fungování mu stačí dekodér. Přelomový je také způsob, jakým GNMT přistupuje k překladu vzácných nebo neznámých slov, které nemá ve své „slovní zásobě“ – takové výrazy se totiž pokusí rozložit na kousky známých slov a správný překlad získat z nich.

V trojici nejznámějších online překladačů<sup>10</sup> Google Translate doplňují Yandex Translate a Microsoft Translator (dříve známý jako Bing) které, stejně jako Google Translate, používají neurální sítě – z trojice těchto hlavních hráčů ho jako poslední zavedl Microsoft, a to na podzim roku 2018 v obousměrných kombinacích němčiny a čínštiny s angličtinou.

Yandex Translate zavedl neurální překlad v roce 2017 v páru angličtina-ruština, ovšem od hlavních konkurentů ho odlišuje kombinace se statistickými metodami, které Yandex Translate využívá od svého spuštění v roce 2011. Ve své tiskové zprávě<sup>11</sup> společnost Yandex tento nový hybridní přístup vyzdvihuje a spojením těch nejlepších vlastností obou metod slibuje uživatelům ještě vyšší kvalitu překladu. Hybridní systém Yandex Translate totiž funguje tak, že zadaný text nejprve přeloží jak pomocí statistické, tak pomocí neurální metody, dále pomocí speciálního algoritmu CatBoost ohodnotí výstupy z obou modelů, a nakonec vybere ten kvalitnější překlad. Tím pádem skutečně využije silné stránky obou metod: neurální překlad totiž používá k překladu slov kontext, což z něj dělá mnohem lepší volbu pro překlad dlouhých vět i delších textových celků, zároveň ale má právě kvůli spoléhání se na kontext problém s kratšími, osamocenými větami či vzácnějšími výrazy. To jsou naopak silné stránky statistické metody, která v takovém případě dokáže rychle a snadno najít vhodný překlad pro slova či fráze, která se v trénovacích datech vyskytují méně často. V jazykové plynulosti výsledného překladu však statistická metoda zaostává kvůli tomu, že věty při překladu rozděluje na slova a fráze. Větší podrobnosti o novém principu fungování svého překladače zatím společnost Yandex nezveřejnila, přímo na webových stránkách společnosti je ale k dispozici detailní popis fungování jejich překladače využívajícího pouze statistický model.

Jak je zřejmé z příkladu tří největších hráčů na poli online strojového překladu, budoucnost oboru spatřují aktuální špičky právě ve strojovém učení a neurálních sítích. I přes velké pokroky posledních let je prostoru pro rozvoj stále dostatek a novou metou by se tak mohlo stát učení neuronových sítí bez použití paralelních textů, neboť i ty nejpokročilejší stále potřebují ke svému učení paralelní korpusy či bilingvní slovníky. Lidé si však mohou rozšiřovat slovní zásobu v cizím jazyce například čtením článků, aniž by bylo nutné je překládat do jejich rodného jazyka. Pokusy naučit neurální sítě právě tomuto, tedy učení bez použití paralelních korpusů, jsou zatím v plenkách, přestože nedostupnost objemných bilingvních korpusů

---

<sup>10</sup> V evropském prostředí

<sup>11</sup> Více na: Yandex — Company blog — One model is better than two. Yandex.Translate launches a hybrid machine translation system. Yandex [online]. Copyright ©. Dostupné z: <https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/>



v některých jazykových párech představuje pro rozvoj neurálního strojového překladu mezi těmito jazyky značnou překážku. První pokusy v tomto směru byly představeny na konferenci ICLR 2018<sup>12</sup> (International Conference on Learning Representations). Všechny jsou založeny na využití dvou jednojazyčných korpusů, kdy navrhované modely mapují věty z těchto korpusů do stejného prostoru, ze kterého je pak rekonstruují zpět, a tím se fakticky učí překládat bez jakýchkoliv paralelních dvoujazyčných dat. Výsledky prvních pokusů jsou zatím poměrně působivé a v některých jazykových kombinacích dokonce předčí existující paralelní metody. Lze proto předpokládat, že se další výzkum neurálního strojového překladu a automatického tlumočení bude ubírat právě tímto směrem, a že nové metody budou založeny především na pokročilejších technikách strojového učení.

### 2.3 Aktuální lídři na trhu

Od vzniku prvního bezplatného online překladače Babel Fish uplynulo již přes 20 let a nejstarší placený software pro strojový překlad Systran již oslavil 50 let své existence. Za tu dobu se trh s placenými i bezplatnými online překladači mnohonásobně rozrostl a neustále se proměňuje – sestavit aktuální seznam všech dostupných řešení je proto prakticky nemožné. Do roku 2010 takový seznam pravidelně sestavoval, aktualizoval a publikoval John Hutchins pod názvem „Compendium of Translation Software<sup>13</sup>“, nicméně v roce 2012 výbor Evropské asociace strojového překladu rozhodl, že kompendium již nadále nebude aktualizováno. Poslední dostupná verze je z března roku 2010 a obsahuje stovky položek, a přestože kompendium obsahuje kromě nástrojů strojového překladu i další pomocný software, jedná se o jasný důkaz toho, jak velký je o tento trh zájem, a zorientovat se ve stovkách překladačů ve formě webových aplikací, pluginů, mobilních aplikací či samostatného softwaru je čím dál složitější. Tuto kapitolu proto věnujeme online překladačům, jelikož jsou na tomto trhu nejviditelnější a zároveň o nich lze získat nejvíce informací – výrobci komerčních licencí totiž detaily o fungování svých produktů nezveřejňují. Z významných poskytovatelů však můžeme jmenovat kromě Systranu také Slate Desktop, DeepL nebo společnost SDL, která disponuje vlastní technologií strojového překladu. Řada programů pro počítačem podporovaný překlad pak nabízí integraci se službami strojového překladu od třetích stran – například MemoQ nabízí uživatelům možnost napojení na Google Machine Translation, Microsoft Bing, DeepL a řadu dalších systémů<sup>14</sup>.

Jednoznačně největším hráčem na poli online překladačů je **Google Translate**, který byl spuštěn v dubnu 2006 a v začátcích plnil svou databázi transkripty z OSN a EU. Podle dostupných údajů dnes využije bezplatnou službu Google Translate, nabízející momentálně 103 jazyky, denně přes 500 milionů lidí po celém světě, kterým Google Translate přeloží 143 miliardy slov<sup>15</sup>. Webový překladač dokáže kromě textu zadaného uživatelem překládat i celé

<sup>12</sup> Veškeré příspěvky je možné prohlédnout si zde: <https://iclr.cc/Conferences/2018/Schedule?type=Poster>

<sup>13</sup> Archiv všech vydaných edicí je dostupný na <http://www.hutchinsweb.me.uk/Compendium.htm>

<sup>14</sup> Viz <https://www.memoq.com/en/integration-with-machine-translation>

<sup>15</sup> Google CEO Sundar Pichai revealed a jaw-dropping fact about its translation app that shows how much money is still sitting on the table. Business Insider [online]. [cit. 2019-04-11]. Dostupné z: <https://www.businessinsider.com/sundar-pichai-google-translate-143-billion-words-daily-2018-7>

webové stránky – stačí zadat adresu webu do políčka pro výchozí text. Přeložený web se otevře na nové stránce a je možné ho naprosto nerušeně procházet bez nutnosti vždy každou stránku překládat zvlášť. Kromě webového překladače je navíc Google Translate dostupný také v rámci balíčku Google Translator Toolkit, který je k dispozici zdarma a nabízí některé funkce nástrojů CAT, a jako aplikace pro mobilní telefony se systémy Android a iOS, kde dokáže fungovat i bez přístupu k internetu. Aplikace pak umí přeložit nejen text, který uživatel „naťuká“ na klávesnici, ale dokáže pracovat i se vstupem v podobě hlasu, fotografie (stačí namířit kameru telefonu například na ceduli a na obrazovce se okamžitě zobrazí překlad), či ručního psaní, což se hodí například pro jazyky se složitějšími abecedami. Výsledný překlad pak dokáže překladač uživateli také přečíst. Jak aplikace, tak webové rozhraní umí také automaticky rozpoznat zdrojový jazyk, pokud si jím uživatel není jistý. Online překladač navíc dává uživatelům možnost výsledné překlady vylepšovat tak, že vyberou z nabízených variant tu, kterou považují za nejlepší, a pokud to samé udělá dostatek lidí, začne překladač tuto variantu používat vždy a označí ji speciální značkou. Nesprávné překlady je naopak možné nahlašovat. Samostatně pak funguje komunita překladače, kde uživatelé mohou překládat slovní spojení či kontrolovat správnost překladu a tím Google Translate vylepšovat. Kromě bezplatných služeb nabízí Google i placené nástroje v podobě cloudového řešení s API (rozhraní pro programování aplikací) a nástroje AutoML Translation, který umožňuje vytváření vlastních modelů strojového učení, takže si uživatel může „vycvičit“ překladač v souladu se svými konkrétními potřebami. Před pouhými několika týdny navíc Google spustil možnost integrace své služby Google Translate do jiných aplikací na mobilních zařízeních s operačním systémem Android – uživatelé si tak mohou například posílat SMS zprávy ve dvou různých jazycích a nechat si je překládat přímo v dané komunikační aplikaci, aniž by museli přepínat do samostatné aplikace Google Translate. Se službou „Tap to Translate“ stačí na zprávu v cizím jazyce klepnout a zobrazí se nejen překlad, ale i možnost přehrání cizojazyčné věty.

Největší konkurent Google Translate, překladač od Microsoftu s názvem **Microsoft Translator**, vznikl v roce 2007 (tehdy ještě pod názvem Windows Live Translator a později Bing Translator). Stejně jako Google Translate, i tento překladač fungoval do zavedení neurálního překladu na statistickém modelu. V současnosti nabízí Microsoft celou řadu produktů spojených s překladem. V první řadě je to Microsoft Translator web app, tedy webové rozhraní, které nabízí stejné funkce, jako Google Translate – překlad zadaných textů nebo celých webových stránek. Momentálně (dle údajů z února 2019) podporuje Microsoft Translator 67 jazyků a na svých stránkách nabízí u každého jazyka přehled služeb či technologií, které jsou ve zvoleném jazyce podporovány<sup>16</sup>. Kromě webového rozhraní je Microsoft Translator rovněž dostupný v podobě aplikace pro iOS, Android a Windows Phone. Mobilní aplikace pak nabízí možnost offline překladu, překladu textu z obrázků, databázi užitečných frází, ale také překlad dvojjazyčné konverzace s hlasovým vstupem (o této službě se v práci ještě zmíníme později) a dokonce překlad skupinových konverzací s textovým i hlasovým vstupem, o kterém se rovněž zmíníme v jedné z dalších kapitol. Mimo to nabízí i řadu dalších zajímavých služeb včetně vlastního cloudového řešení s API, službu Custom Translator kterou lze propojit s API a která umožňuje vytvoření neurálních překladových systémů naučených na

---

<sup>16</sup> Viz <https://translator.microsoft.com/help/languages/>

terminologii dle požadavků uživatelů (firem, vývojářů, či poskytovatelů jazykových služeb), lokalizační nástroj Multilingual App Toolkit, a také zajímavou službu Live Feature. Ta je dostupná v prohlížečích Microsoft Edge, Google Chrome, Mozilla Firefox a Safari a také v mobilních aplikacích Microsoft Translator a dle popisu na stránkách Microsoftu umožňuje až stovce lidí účastnit se jedné konverzace na různých zařízeních a vést ji s až desítkou jazyků, nezávisle na druhu vstupu (hlasový či textový). Přeložený text pak služba zobrazí na obrazovkách účastníků a přečte. Z dostupných informací zatím není zcela zřejmé, jakými možnostmi nastavení služba disponuje, nicméně se jí budeme věnovat podrobněji v kapitole o dostupných technologiích automatického tlumočení.

Mimo západní trh stojí za zmínku ještě dvě služby – těmi jsou ruský Yandex Translate a čínský překladač Baidu, přičemž obě služby spadají pod společnosti, které zároveň provozují nejoblíbenější internetové vyhledávače na svých domácích trzích.

**Yandex Translate** nabízí, stejně jako konkurence, webový překladač, rozšíření do prohlížečů, mobilní aplikaci, funkci překladu celých webů a také překlad obrázků, tedy například oskenovaných dokumentů, pomocí technologie pro optické rozpoznávání znaků (OCR). V současnosti podporuje 95 jazyků, hlasový vstup a v mobilní aplikaci nabízí také transliteraci z vybraných jazyků<sup>17</sup>.

Překladač **Baidu** byl spuštěn v roce 2011 a v roce 2017 už podporoval 28 jazyků a fungoval na základě neurálních sítí. Podle informací od zástupce společnosti Baidu je jejich překladač založen na technologii hlubokého učení a funguje mnohem lépe než překladače založené na statistickém modelu<sup>18</sup>. Kromě základního webového rozhraní nabízí také překlad obrázků a funkci „rozpoznávání objektů“ – stačí namířit fotoaparát telefonu na předmět, vyfotit ho, a aplikace předloží uživateli název předmětu v čínštině i dalším zvoleném jazyce. Dále společnost nabízí i placené služby v podobě předplatného založeného na objemu, podobně jako konkurence. Baidu se také angažuje v oblasti automatického tlumočení – v roce 2017 představila přenosný překladač, který dokáže zpracovat hlasový vstup a mluvit v několika jazycích, jehož cílovou skupinou jsou především čínští turisté, a STACL, systém pro „simultánní překlad“ se schopností předpovídat a s možností ovládat prodlevu<sup>19</sup>, o kterém se podrobněji zmíníme v dalších kapitolách.

---

<sup>17</sup> Viz <https://yandex.com/support/translate/index.html>

<sup>18</sup> Baidu Translate: The Inside Story | Slator. Slator | Language Industry Intelligence [online]. Copyright © 2019 [cit. 11.04.2019]. Dostupné z: <https://slator.com/technology/baidu-translate-the-inside-story/>

<sup>19</sup> Baidu Research Announces Breakthrough in Simultaneous Translation [online]. 24.10.2018 [cit. 2019-04-11]. Dostupné z: <http://research.baidu.com/Blog/index-view?id=107>

### 3 Automatické tlumočení

S prudkým rozvojem strojového překladu od 90. bylo zřejmé, že další metou se stane právě automatické tlumočení, které je se strojovým překladem úzce spjato. Dnes už můžeme na trhu najít řadu placených i bezplatných systémů v nejrůznější podobě, od mobilních aplikací přes součásti jiných komunikačních programů až po samostatná zařízení, a na nejrůznější úrovni, od jednoduchých zařízení, která lze srovnat s konverzačními příručkami, po ambiciózní projekty, které slibují všestrannost po stránce jazykové i tematické.

Na začátku je však potřeba zmínit, že pokud i v oblasti strojového překladu stále nebyly naplněny ambice vytvořit strojový překlad nejen stejně kvalitní, jako ten lidský, ale dokonce i kvalitnější, pravděpodobnost toho, že by se to povedlo v oblasti automatického tlumočení je ještě menší. I tak si automatické tlumočení zaslouhuje pozornost, a přestože podle názoru mnoha expertů i profesionálních tlumočnicků nikdy nedokáže plně nahradit roli člověka (především v prostředí konferenčního tlumočení, vysoké konsektivity a v dalších případech), to, že se stane běžným nástrojem pro usnadnění komunikace v každodenním životě už zdaleka tolik nepravděpodobné není. Důkazem toho, že můžeme v budoucnosti očekávat neklesající zájem o tyto technologie a jejich rozvoj, je i dokument americké vlády z roku 2009 s názvem „A Strategy for American Innovation“. Cílem této strategie je totiž vypořádat se s „největšími výzvami“ 21. století – a na úplném konci, mezi dlouhodobými cíli, najdeme i bod který hovoří o vývoji systému pro „automatizovaný a vysoce přesný překlad největších světových jazyků v reálném čase“<sup>20</sup> který by výrazně zjednodušil mezinárodní obchod a spolupráci.

V této části práce budeme čerpat z odborných publikací od autorů jako jsou Daniel Gile a Ivana Čeňková, ze sborníkových publikací Routledge Encyclopedia of Translation Technology a Routledge Encyclopedia of Interpreting Studies, z archivu odborných článků ArXiv spravovaného Cornellovou univerzitou, z publikací věnovaných jednotlivým zmiňovaným systémům, z developerských blogů, a rovněž z webů či blogů výrobců zmiňovaných technologií. V části porovnávající nový čínský systém STACL s prací Daniela Gila jsme pak rovněž využili asistence odborného konzultanta, který je členem týmu vývojářů STACL, nicméně si nepřál být výslovně jmenován. Pozornosti čtenáře můžeme rovněž doporučit publikaci TAUS věnovanou automatickému tlumočení, kterou si členové mohou na webových stránkách TAUS stáhnout bezplatně, nečlenové pak za poplatek ve výši 200 USD<sup>21</sup>.

#### 3.1 Popis a zařazení

Než se dostaneme k historickému vývoji a detailnějšímu zkoumání automatického tlumočení, je potřeba vymezit, co automatické tlumočení vlastně je. Podstatou procesu označovaného v angličtině jako „speech translation“ nebo „speech to speech translation“ je převod mluvené

---

<sup>20</sup> A Strategy For American Innovation: Driving Towards Sustainable Growth And Quality Jobs. 2009. s. 22.

Dostupné z: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a524201.pdf>

<sup>21</sup> Publikace je dostupná ke stažení či koupi na tomto odkazu: <https://www.taus.net/think-tank/reports/translate-reports/taus-speech-to-speech-translation-technology-report>

řeči z jednoho přirozeného jazyka do mluvené řeči v jiném přirozeném jazyce, a tím pádem zprostředkování komunikace mezi dvěma lidmi, kteří nehovoří stejným jazykem. Ve své nejjednodušší podobě by tedy systém pro automatické tlumočení mohl fungovat jako mluvící slovník s funkcí rozpoznávání hlasu tlumočící jednotlivá slova bez ohledu na gramatiku a jiné lingvistické aspekty. Takový systém by však měl velmi omezené možnosti využití, bez ohledu na to, že v dnešní době již máme mnohem vyšší nároky – očekáváme technologie, které si poradí s celými větami, promluvy nebo s běžnou konverzací, a které nám poskytnou výstup blížící se lidské řeči jak srozumitelností, tak přirozeností.

Při snaze blíže definovat automatické tlumočení pak lze polemizovat o tom, jestli je současné systémy vůbec možné označit pojmem tlumočení a zda jeho znaky skutečně naplňují. Ani terminologie v dostupných odborných publikacích není ve směru definice jednotného pojmu konzistentní a k označení těchto technologií se používají výrazy jako „speech translation“, „machine interpreting“ nebo „automatic interpreting“, aniž by byly tyto pojmy jasně vymezeny a definovány.

Jednota tlumočnicků a překladatelů ve své příloze k časopisu ToP z roku 2006 uvádí jednoduchou definici toho, co je a co není tlumočení:

*„Tlumočení je zprostředkování mluvené komunikace mezi lidmi hovořícími různými jazyky. Tlumočení není překladem slovo od slova, ale převedením smyslu neboli obsahu sdělení.“<sup>22</sup>*  
(Diabová, Kautský, Pošta, 2006)

Přestože je pravdou, že „zprostředkování mluvené komunikace“, tedy dorozumění, dnešní systémy automatického tlumočení do určité míry a v určitém kontextu zvládají, problém spočívá především ve druhé větě, která vyvolává otázku, zda tyto systémy skutečně dokáží předat smysl sdělení a nakolik jsou schopny provést analýzu originálního sdělení, která je k nalezení onoho smyslu nezbytná. Přestože tedy dokáží novodobé systémy automatického tlumočení pracovat s hlasovým vstupem a poskytnout (více či méně) přirozený hlasový výstup, v jejich jádru stále najdeme strojové překladače vyvinuté pro překlad textu, které pracují s celými větami. Rozložením celého procesu na jednotlivé kroky tak zjistíme, že systémy automatického tlumočení vlastně až tak působivé nejsou a velmi zjednodušeně se dá říci, že jejich jedinou přidanou hodnotou je eliminace nutnosti text na jedné straně psát a na druhé straně číst. To má bezpochyby své výhody – komunikace prostřednictvím zařízení automatického tlumočení je díky absenci textového zadávání rychlejší a subjektivně i příjemnější než nutnost si s komunikačním partnerem navzájem ukazovat displeje mobilních telefonů s přeloženým výsledkem, a jedná se i o velký posun z hlediska zpřístupnění těchto technologií pro uživatele s postižením.

Připustíme-li však, že se v případě systémů automatického tlumočení o tlumočení skutečně jedná, pak narážíme na problém s jeho zařazením, totiž zda ho přiřadit ke tlumočení konsekutivnímu nebo simultánnímu.

---

<sup>22</sup> DIABOVÁ, Amalaine, Petr KAUTSKÝ a Miroslav POŠTA. Tlumočení a jak na to, aneb, Chcete, aby vám rozuměli posluchači (a spolu s nimi i tlumočníci)? Praha: JTP, 2006. 23 s. ISBN 8086711986.

Hlavním aspektem, který tyto dva druhy tlumočení odlišuje, je souběžnost, a to jak ve smyslu souběžnosti řečnickovy i tlumočnickovy produkce, tak ve smyslu souběžnosti procesů poslechu a produkce v tlumočnickově mozku:

*„Pojmem simultánní tlumočení se obvykle rozumí takový druh tlumočení, při kterém produkce řeči v cílovém jazyce probíhá souběžně (nebo téměř souběžně) s percepcí řeči ve výchozím jazyce.“ (Černov, 2009, překlad Y.S.)<sup>23</sup>*

U konsektivního tlumočení souběžnost nenajdeme, jelikož se řečník s tlumočnickem střídají a prodleva mezi nimi může činit až několik minut. V případě simultánního tlumočení je však prodleva jak mezi procesy poslechu a produkce, tak mezi produkcí řečníka a tlumočnicka minimální, v řádu vteřin, a lze tedy hovořit o souběžnosti. Pokud se tedy podíváme na v současnosti dostupné systémy, je logičtější přiřadit je k tlumočení konsektivnímu, jelikož funkční systém naplňující rysy simultánního tlumočení – tedy především souběžnost, ale také předvídání, analýzu, využívání tlumočnických strategií a další – zatím, pokud je známo, nebyl sestrojen. Nejvíce se takovému systému momentálně přiblížila čínská společnost Baidu – o jejím nejnovějším prototypu se podrobněji zmíníme v následujících kapitolách (3.4.3 a 3.5). Lze ale o systémech automatického tlumočení skutečně říci, že naplňují znaky alespoň konsektivního tlumočení? Kromě značné prodlevy za řečníkem je totiž další vlastností a požadovaným aspektem kvalitního konsektivního tlumočení to, že je kratší než původní řečnickův projev<sup>24</sup>. Předpokladem pro zkracování je však opět analýza, pochopení toho, které informace jsou zásadní, které je možné zredukovat, a které části projevu je možné vypustit (například opakující se informace, redundance atd.). Je však taková analýza při současných systémech založených na korpusovém učení možná? Hlubkové prozkoumání této otázky by jistě mohlo posloužit jako zajímavé téma pro budoucí bádání.

Na základě výše uvedených argumentů se domníváme, že současné široce dostupné systémy automatického tlumočení znaky tlumočení tak, jak jsou popisovány v rámci obecně přijímaných definic, nenaplňují. Jako vhodnější termín navrhuje „hlasový strojový překlad“ a pro pokročilejší systémy jako Baidu pak „simultánní strojový překlad“. Pro účely této práce však budeme s ohledem na konzistentnost a srozumitelnost nadále používat k popisu daných technologií termín automatické tlumočení, je ale nutné konstatovat, že anglický výraz „speech translation“ považujeme za nejpřesnější označení.

Vedle technologií pro automatické tlumočení existují ještě jiné, blízké přidružené technologie. Mezi tyto můžeme počítat systémy pro překlad z textu do mluvené podoby (text to speech translation) a překlad z mluvené podoby do textu (speech to text translation), které najdeme například v mobilních aplikacích Google Translate či Microsoft Translator, nebo například automatické vytváření titulků na základě hlasového vstupu a další. Všechny tyto systémy spojuje to, že ke svému fungování vyžadují technologie pro rozpoznávání hlasu a pro hlasovou syntézu. Těmto technologiím se budeme blíže věnovat v následující kapitole.

---

<sup>23</sup> ČERNOV, Gelij. Teorija i praktika sinchronnogo perevoda. Moskva: LIBROKOM, 2014. ISBN 978-5-397-04244-4

<sup>24</sup> ČEŇKOVÁ, Ivana. Úvod do teorie tlumočení. Praha: Česká komora tlumočnicků znakového jazyka, c2008. ISBN 978-80-87153-74-1.

## 3.2 Princip fungování a nezbytné technologie

Současné systémy pro automatické tlumočení by nemohly existovat bez několika jiných technologií, které jsou pro jejich fungování nezbytné – jedná se o strojový překlad, automatické rozpoznávání řeči, a syntézu řeči. V této kapitole je, s výjimkou strojového překladu, kterému jsme se již zevrubně věnovali v předchozí části práce, stručně představíme, neboť seznámení s nimi je nezbytné pro pochopení a hlubší vhled do omezení, se kterými se automatické tlumočení potýká.

### 3.2.1 Automatické rozpoznávání řeči

Za první skutečný systém automatického rozpoznávání řeči je považována tzv. „Audrey“ (Automatic Digit Recognition machine), kterou v roce 1952 vytvořili v Bell Labs, a která dokázala rozpoznávat fonémy. Audrey dokázala rozpoznat vyslovené číslice od 1 do 9 až s 90 % úspěšností, pokud je vyslovil její hlavní vývojář, H.K. Davis, a se 70–80 % přesností v případě jiných určených mluvčích. Pokud je ale vyslovili pro systém neznámí mluvčí, byla úspěšnost značně nižší.

Další velký posun přišel v roce 1976, kdy byl díky financování americké Agentury ministerstva obrany pro pokročilé výzkumné projekty (DARPA) vytvořen program jménem Harpy, který jako první dokázal rozpoznávat celé věty a měl slovní zásobu okolo 1000 slov. Až do 90. let však téměř všechny tyto programy vyžadovaly aby řečník mluvil pomalu, artikulovaně, a po každém slovu udělal pauzu. Revoluci přinesl až první komerční produkt pro rozpoznávání hlasu Dragon Dictate z roku 1990, který pak v roce 1997 následoval Dragon NaturallySpeaking, první program schopný rozpoznávat plynulou řeč při tempu zhruba 100 slov za minutu, který je dodnes používán například lékaři k vytváření lékařských záznamů.

Současné systémy automatického rozpoznávání řeči už fungují na bázi strojového učení, a jsou proto schopny poradit si s nejrůznějšími hlasy, přízvuky, výslovností či rychlostí – například software od Googlu už je údajně schopen rozpoznávat mluvenou angličtinu s 95 % přesností<sup>25</sup>. Přesto však tato technologie stále není dokonalá, především v situacích, kdy je na pozadí přítomen jiný zvuk či další hlasy.

### 3.2.2 Syntéza řeči

Ze všech technologií, které jsou k fungování automatického tlumočení nezbytné, je syntéza řeči nejstarší. První syntetizátor byl totiž sestaven už v roce 1779 Christianem Kratzensteinem v Petrohradu – ten popsal fyziologické rozdíly mezi samohláskami a, e, i, o, u a vytvořil stroj, který byl schopen je vyprodukovat. Experimenty s nejrůznějšími mechanickými stroji pak pokračovaly až do 60. let 20. století, avšak nevedly k žádným větším úspěchům. První elektrický syntetizér byl vyroben v roce 1922 a první skutečný řečový syntetizér VODER (Voice

---

<sup>25</sup> Internet Trends Report 2017 | Kleiner Perkins. Kleiner Perkins | Make History [online]. Dostupné z: <https://www.kleinerperkins.com/perspectives/internet-trends-report-2017>



Operating Demonstrator) byl představen na Světové výstavě v roce 1939 a stal se důkazem toho, jaký má tento obor potenciál.

První systém pro syntézu řeči z textu vznikl v Japonsku v roce 1968 a po několika desetiletích experimentů se jednalo o první systém schopný produkovat sice monotónní, ale srozumitelnou řeč. V roce 1976 byl vytvořen první systém pro předčítání textu, který sloužil jako pomůcka pro nevidomé.

V moderní řečové syntéze jsou pak zásadní dvě vlastnosti: srozumitelnost, tedy „technická“ kvalita výstupu, a přirozenost, tedy nakolik dokáže výstup zachytit a předat emoce, rytmus řeči, výslovnost, intenci atd. Momentálně se pro syntézu řeči z textu používají dvě základní metody: konkatenativní, založená na skládání malých audio nahrávek do větších celků (která je velmi srozumitelná ale také velmi nepřirozená) a parametrická, která nepoužívá přednahrané zvuky, ale vytváří řeč na bázi určitých parametrů (výsledky bohužel nejsou příliš dobré, jelikož je obtížné vygenerovat přirozeně znějící výstup na základě pevných parametrů). Nejnovější technologie z dílny Googlu, představené během posledních dvou let, pak pracují se spektrogramy a jsou téměř nerozeznatelné od lidské řeči – to proto, že se dokáží učit na vzorcích lidské řeči a vlastnosti jako jsou intonace či pauzy aplikovat na syntetické výstupy<sup>26</sup>. V porovnání s lidskými nahrávkami pak dosahují MOS skóre (Mean Opinion Score) 4,53, zatímco profesionální nahrávky řeči dosahují skóre 4,58.

### 3.3 Historie a vývoj automatického tlumočení

Stejně jako v případě strojového překladu, první prototypy systémů pro automatické tlumočení byly skromné. Jedním z nich byl koncept společnosti NEC Corporation, která na konferenci ITU Telecom World v roce 1983 představila systém pro automatické tlumočení mezi španělštinou a angličtinou – ten uměl zpracovat pouze 500 slov a na každé potřeboval několik vteřin.

Systematický výzkum v této oblasti začal se založením ATR (Advanced Telecommunications Research Institute International) v Japonsku v roce 1986 a řada velkých projektů pak byla v 90. letech v spuštěna právě v Japonsku, ale také v Německu a v USA. Mezi nejvýznamnější projekty tohoto desetiletí patřil například ATR-MATRIX, který tlumočil z japonštiny do angličtiny a jeho slovní zásoba čítající 2000 slov byla zaměřená na přirozenou komunikaci týkající se hotelových rezervací. Starším bratrem tohoto programu byl projekt ATR-ASURA, který uměl rozpoznat hlasový vstup v japonštině a převést ho do němčiny a angličtiny – výstup byl podáván ve formě syntetického hlasu.

V Evropě, konkrétně v Německu, patřil k hlavním počínům v této sféře projekt **Verbmobil**, který se vyznačoval svou rozsáhlostí – pod záštitou a financováním Federálního ministerstva pro vzdělávání a výzkum Spolkové republiky Německo se na něm podílelo 19 akademických institucí a 4 komerční partneři a výzkum trval 8 let. Cílem projektu bylo vytvoření systému pro

---

<sup>26</sup> Skerry-Ryan, R. J., Eric Battenberg, Ying Xiao, Yuxuan Wang, Daisy Stanton, Joel Shor, Ron J. Weiss, Robert Clark and Rif A. Saurous. "Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron." 2018, ICML. Dostupné z: <https://arxiv.org/pdf/1803.09047.pdf>



automatické tlumočení po telefonu. Výsledný produkt pak fungoval v obousměrných kombinacích angličtiny, němčiny a japonštiny a tematicky pokrýval tři oblasti spojené s obchodem – plánování schůzek, plánování cest a vzdálenou údržbu PC. Při hodnocení finální verze Verbmobilu čítala jeho slovní zásoba cca 10 000 položek (což odpovídá slovní zásobě přibližně pětiletého dítěte<sup>27</sup>).

Přelomovost Verbmobilu spočívala především v tom, že dokázal pracovat s kontextem, s tím, co již bylo řečeno a se svou vlastní znalostí tématu, a za pomoci kontextu tak z dialogu vybrat v případě více možných variant překladu tu správnou. Z uživatelského hlediska se také jednalo o první takový systém, ve kterém (v jeho finální verzi) uživatelům stačilo pouze zvednout mobilní telefon a ovládat službu pomocí hlasových příkazů – nebylo tedy potřeba mít k dispozici počítač ani jakékoliv kabelové připojení. Navíc, přestože prvotním cílem Verbmobilu bylo vytvořit systém, který by sloužil pouze jako pomůcka při komunikaci tváří v tvář (s mobilními telefony v roli „kapesních tlumočnicků“), bylo možné ho používat i jako skutečné tlumočení po telefonu, aniž by se účastníci rozhovoru navzájem viděli.

Na svou dobu (výzkum probíhal v letech 1993–2000) se jednalo o revoluční projekt, který se dokázal vypořádat s řadou výzev a přinést řadu inovací. Jednalo se například o první systém, který nevyžadoval využívání tzv. „push-to-talk“ (tedy podržení tlačítka) k tomu, aby mu uživatel naznačil začátek a konec svého vstupu a tím tedy i to, co je nutné převést do jiného jazyka. Verbmobil dokázal tuto informaci získat z hlasového vstupu samostatně a byl technicky vybaven tak, aby dokázal pracovat i s nedokonalou zvukovou kvalitou vstupu. Dále byl Verbmobil vybaven dvoustupňovou analýzou, díky které dokázal zpracovat takové aspekty přirozené řeči jako přeřeknutí, opravy či falešné začátky, což mu umožňovalo přetlumočit to, co řečník skutečně chtěl říct, a ne pouze to, co skutečně vyslovil. Verbmobil se také dokázal přizpůsobovat řečníkovi (jeho rychlosti, hlasu, výslovnosti) a tím zlepšovat rozpoznání jeho řeči.

K úplnému ukončení Verbmobilu došlo v roce 2010 a přestože se z něj nikdy nestal komerční produkt, pomohl prošlapat cestu k mnoha jiným technologiím. Dnes je Verbmobil exponátem v Německém institutu pro umělou inteligenci (DFKI).

Dalším významným projektem devadesátých let byl systém **JANUS**, který vznikl díky spolupráci Univerzity Carnegieho–Mellonových v USA a Univerzity v Karlsruhe v Německu. I tento systém byl zaměřen na konkrétní témata a dokázal tlumočit spontánní konverzaci z němčiny, angličtiny a španělštiny do němčiny, angličtiny, španělštiny, japonštiny a korejštiny. Zajímavostí je, že k zobrazení řečníkem zamýšleného významu vstupu používal JANUS sémantický metajazyk, což je důkazem toho, jak blízko si jsou strojový překlad a automatické tlumočení a jak nekonkrétní dokáže být hranice mezi nimi. Ani JANUS se však nedostal dál než k prototypům, které sloužily k testování prostředí, ve kterém by systém mohl být využíván (např. pro videokonference o dvou lidech nebo jako přenosný mobilní tlumočnick).

Posledním významným projektem, o kterém se zmíníme podrobněji, je výzkumný projekt Evropské Komise s názvem **TC-STAR**. Ten byl spuštěn v roce 2004 s dlouhodobým cílem

---

<sup>27</sup> Vocabulary size and auditory word recognition in preschool children. National Center for Biotechnology Information [online]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5400288/>

posouvat výzkum ve všech oblastech a technologiích, které jsou pro fungování automatického tlumočení nezbytné (tedy například vytvoření lepší, „lidštější“ hlasové syntézy nebo spolehlivého systému pro rozpoznávání řeči při různých stylech, dialektech, nebo i zvukových podmínkách) a s ambicemi vytvořit průlomovou technologii, která by významně zmenšila rozdíl v kvalitě tlumočení<sup>28</sup> mezi strojem a člověkem. Projekt pracoval s výběrem ničím nsvázané přirozené řeči ve formě projevů z Evropského parlamentu a z vysílání zpráv, pokrýval tři jazyky (evropskou angličtinu, evropskou španělštinu a mandarínskou čínštinu) a účastnily se ho mimo jiné například společnosti Siemens, Nokia, Sony či IBM (ta o přesně 50 let dříve předvedla vůbec první prototyp strojového překladače). Za jedno z měřítek úspěchu bylo mimo jiné považováno také zapojení externích účastníků do vyhodnocovacích kampaní, které pak měly být prezentovány v podobě workshopů s cílem přilákat zájem vědců a technologických firem. Projekt TC-STAR byl po 3 letech fungování ukončen v březnu 2007 (navzdory původnímu plánu, který počítal s šestiletým projektem). V závěrečné prezentaci k projektu je mezi jeho úspěchy zmiňováno vytvoření infrastruktury potřebné k urychlení výzkumu v dané oblasti a zlepšení výkonu v jednotlivých oblastech o 40–60 %. Za pokračovatele TC-STARu můžeme považovat dva projekty zaměřené na strojový překlad, EuroMatrix (2006-2009) a EuroMatrixPlus (2009–2012), které si kladly za cíl pokrýt veškeré jazyky Evropské unie a zaměřovaly se tedy na unijní dokumenty a vnitřní komunikaci.

Na konci 90. let se do výzkumu automatického tlumočení zapojila také Čína, jejíž první systém LodeStar byl zaměřen na tlumočení mezi čínštinou a angličtinou či japonštinou v oblasti cestování. V rámci letních olympijských her v Pekingu v roce 2008 byl představen experimentální systém automatického tlumočení s podporou čínštiny, angličtiny a španělštiny, který měl pomáhat turistům.

Zhruba od roku 2000 se výzkum v oblasti automatického tlumočení začal postupně přizpůsobovat reálným komunikačním potřebám a scénářům. Jedním z jeho hlavních hybatelů se tak stala armáda, pro jejíž potřeby při zahraničních misích byl vytvořen nejen jeden systém automatického tlumočení. Agentura ministerstva obrany pro pokročilé výzkumné projekty (DARPA) financovala tři významné projekty známé pod názvy GALE, TRANSTAC a BOLT, které se zaměřovaly především na arabské jazyky a na nichž se podílely akademické instituce i soukromé společnosti z USA a západní Evropy. To také znamenalo nutnost zajistit dostupnost těchto systémů nejen na počítačích, ale i na menších, přenosných zařízeních. První komerční systémy se začaly v různých podobách (někdy se jednalo o prototypy, jindy o hotové produkty) objevovat od roku 2006.

V roce 2009 byl představen první funkční „automatický hlasový překladač“, který dokázal fungovat i bez připojení k internetu, s názvem Jibbig. Ten nabízel přes 20 jazyků<sup>29</sup>, dokázal si poradit s tématy jako cestování, vojenství, zdravotnictví a humanitární pomoc a existoval i v podobě mobilní aplikace. Fungoval na stejném principu, jako hlasový překladač Google Translate: uživateli stačilo držet tlačítko a mluvit do telefonu nebo napsat text ručně – překlad

---

<sup>28</sup> V této kapitole používáme pro zachování konzistentnosti pojem „tlumočení“, ve velkém množství projektů v oblasti „speech to speech“ si však lze všimnout, že samy sebe či procesy uvnitř svého systému označují slovem „translation“, tedy překlad.

<sup>29</sup> Jibbig Translator 2.0 3.2 Free Download [online] [cit. 09.04.2019]. Dostupné z: <https://jibbig-translator-2-0.soft112.com/>

se mu pak zobrazil na displeji a zároveň ho aplikace i přečetla. Demonstraci Jibbiga v praxi je dokonce možné vidět v epizodě amerického dokumentárního seriálu Nova s názvem „Smartest Machine on Earth“ z roku 2011<sup>30</sup>, kdy se jednalo o skutečně přelomovou technologii. V roce 2013 pohltit společnost, která tuto aplikaci vytvořila, Facebook, který technologii nyní používá pro překlad statusů<sup>31</sup> (dříve využíval Facebook pro tento účel Microsoft Bing). V současné době se zdá, že samotná aplikace Jibbiga již není několik let podporována a zmizela i možnost získat ji pro mobilní zařízení.

### 3.4 Aktuálně dostupná řešení

V této kapitole navážeme na historický exkurz z předchozí kapitoly a představíme ty nejzajímavější a nejpoblárnější současné produkty či prototypy, které byly představeny veřejnosti. Vzhledem k tomu, jak mlhavá je hranice mezi překladem a tlumočením ve světě technologií, zařadíme do této kapitoly jak produkty fungující v režimu „speech to speech“, tak i jeden velice zajímavý produkt, který sice zatím nedisponuje funkcí hlasové syntézy, ale z hlediska svých schopností je naplnění znaků tlumočení výrazně blíží, než neurální překladače rozšířené o schopnost rozpoznávání a syntézy hlasu. Informace budeme v této kapitole čerpat převážně z webových stránek jednotlivých výrobců či z jejich tiskových zpráv, jelikož odborné publikace pokrývající tyto nejmodernější technologie zatím téměř neexistují. Vzhledem k tomu, že sami výrobci z pochopitelných důvodů nezveřejňují podrobné informace o principu fungování svých technologií, zaměříme se především na popis funkcionality těchto zařízení a aplikací. Pro přehlednost rozdělíme tuto kapitolu na několik kategorií.

#### 3.4.1 Mobilní aplikace a software pro počítače

##### a) Google Translate

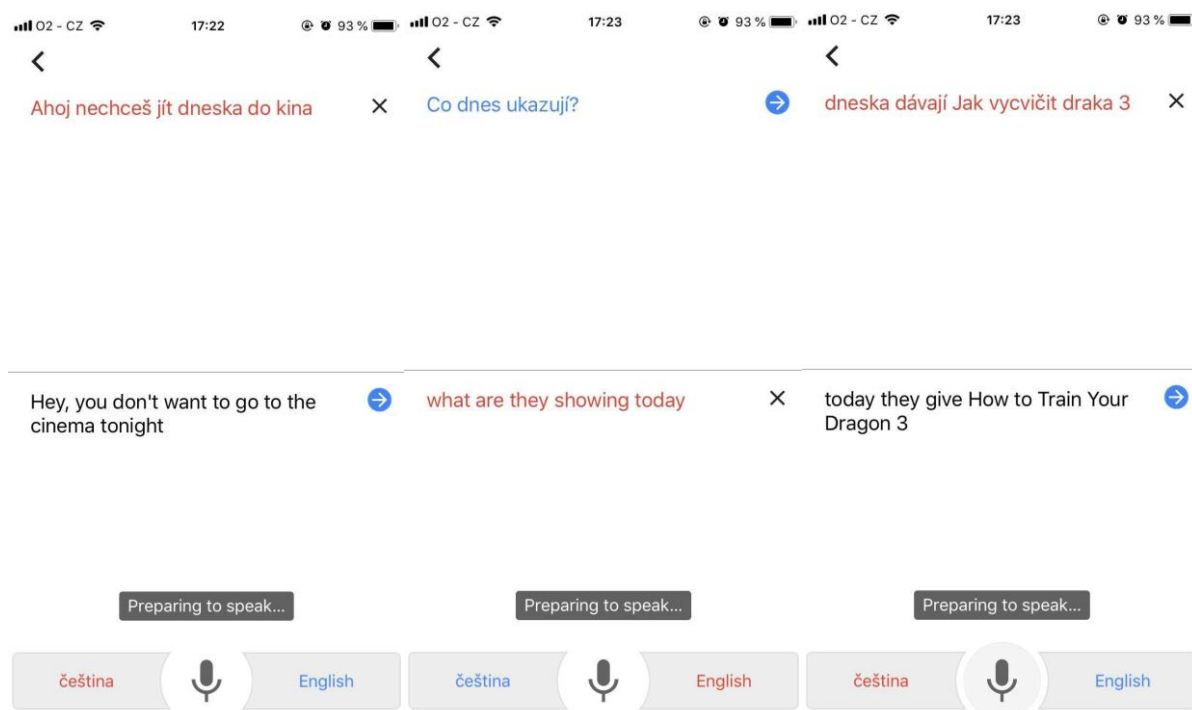
Službě Google Translate jsme se již poměrně podrobně věnovali v předchozích kapitolách (2.2, 2.3), zde proto pouze připomeneme, co všechno tento komplexní systém dokáže. V současnosti tedy Google Translate dokáže převádět výchozí jazyk do cílového v režimu text-to-text, text-to-speech, speech-to-text a speech-to-speech, a dále také dokáže přijímat vstupní informace v podobě ručně psaného textu, obrázků či fotografií. Hovořit o skutečném tlumočení však v případě Google Translate nelze, a ani sama společnost ve svých propagačních materiálech slovo tlumočení nepoužívá, čímž dává najevo, že ambicí Googlu je především usnadnit lidem komunikaci v běžném každodenním životě. Poslední novinkou jsou pak sluchátka Google Pixel Buds, do kterých je služba Google Translate integrována. Nejedná se však o žádnou novou samostatnou službu – jde pouze o způsob, jak uživatelům zpříjemnit používání překladače

---

<sup>30</sup> Nova Smartest Machine on Earth PBS Documentary. YouTube [online]. Dostupné z: <https://www.youtube.com/watch?v=yywl7ugBubA> čas: 33:20

<sup>31</sup> Facebook Acquires “Mobile Technologies”, Developer Of Speech Translation App Jibbiga – TechCrunch. TechCrunch – Startup and Technology News [online]. Dostupné z: <https://techcrunch.com/2013/08/12/facebook-acquires-mobile-technologies-speech-recognition-and-jibbiga-app-developer/?guccounter=1>

Google Translate, stále je však potřeba mít u sebe chytrý mobilní telefon s aplikací Google Translate. Uživatel se sluchátky pak slyší překlad přímo v nich a mikrofon ve sluchátkách zároveň zpracovává jeho hlas, zatímco protistrana používá mobilní telefon.

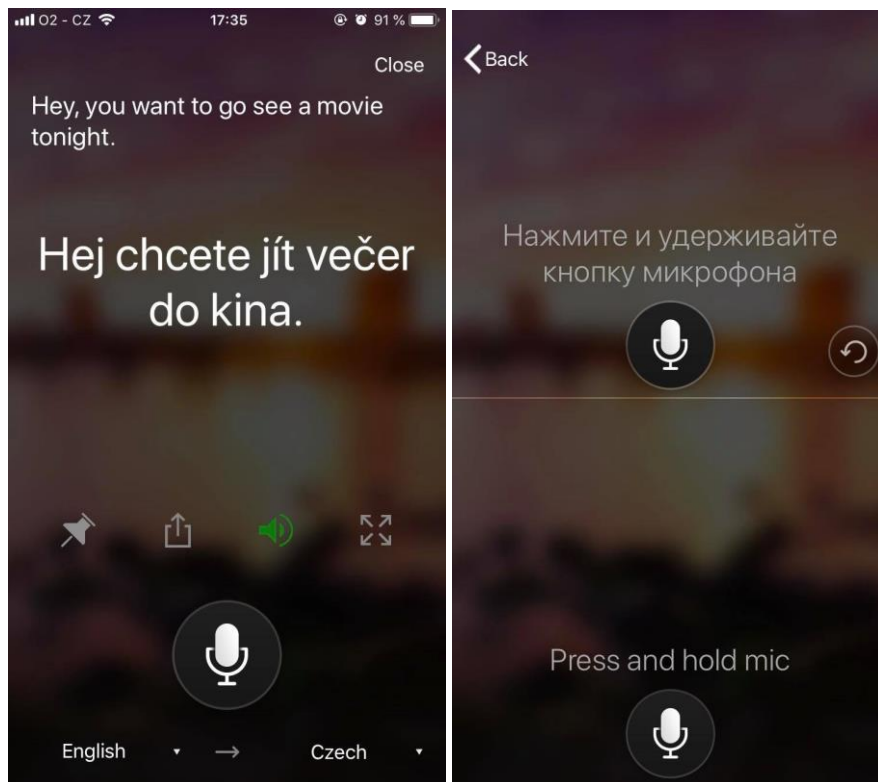


Obrázky 1-3: Screenshoty překladač mluveného dialogu pomocí Google Translate, vlastní zkouška programu.

## b) Microsoft Translator

Stejně jako v případě Google Translate, ani zde nebudeme opakovat informace, které již zazněly v části práce věnované strojovému překladu a pouze připomeneme, že co do funkcí je Microsoft Translator vybaven prakticky stejně, jako Google Translate. Jeho nejzajímavější funkcí je však tzv. **Live feature**, která funguje v aplikaci MS Translator na operačních systémech Android a iOS a na počítačích s Windows 10. Live feature slibuje tlumočené (speech-to-speech) hovory až pro 100 uživatelů s až 10 jazyky prostřednictvím tzv. virtuální místnosti<sup>32</sup>. Pokud tedy mluvčí prezentuje například v angličtině, jeho posluchači se mohou do této virtuální místnosti přidat prostřednictvím svých zařízení a nastavit si, ve kterém z podporovaných jazyků by chtěli mít překlad. Ten se jim pak zobrazuje na displeji a zároveň ho díky systému pro syntézu řeči mohou i slyšet. Posluchači zároveň mohou pokládat mluvčímu dotazy ve svém jazyce – Live feature mu je pak stejným způsobem předá na jeho zařízení do jeho jazyka. Služba je bezplatně dostupná veřejnosti a k jejímu využití stačí mít na svém zařízení nainstalovanou aplikaci Microsoft Translator. Mobilní aplikace pak, stejně jako konkurence, nabízí tzv. konverzační mód, který ale češtinu (zatím) podporuje pouze jako cílový jazyk, nikoliv však jako jazyk zdrojový (viz screenshot).

<sup>32</sup> Demonstrace MS Live feature při návštěvě čínských studentů ve výzkumném centru umělé inteligence společnosti Microsoft ke shlédnutí na YouTube: [https://www.youtube.com/watch?v=5tZn\\_oslNXw](https://www.youtube.com/watch?v=5tZn_oslNXw)



Obrázek 4: Uživatelské rozhraní konverzačního módu Microsoft Translator s angličtinou a češtinou.

Obrázek 5: Uživatelské rozhraní konverzačního módu Microsoft Translator s angličtinou a ruštinou.

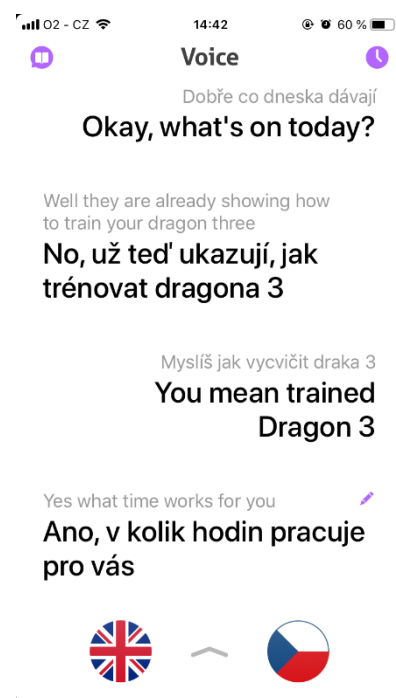
### c) Skype Translator

Vlastníkem populárního komunikátoru Skype je společnost Microsoft, a proto i překladač integrovaný do tohoto programu využívá stejné technologie, jako výše zmiňovaný Microsoft Translator. Skype Translator v současnosti podporuje 13 jazyků ve speech-to-speech režimu, tedy pro hovory, a 50 jazyků pro rychlé zprávy. Po testovacím provozu v rámci Skype Translator Preview je nyní tento překladač integrován do Skypu na všech zařízeních. Na poli veřejně dostupných a bezplatných služeb zaměřených pouze na automatické tlumočení představuje jednu z nejstarších – poprvé byl Skype Translator představen v roce 2014 a je pravděpodobné, že se stal základem pro výše popsanou MS Live feature. V roce 2016 Microsoft oznámil, že MS Translator API, na kterém Skype Translator běží, budou firmy moci zakoupit a použít pro přidání funkce automatického tlumočení do svých vlastních programů či aplikací. Právě Skype Translator také použijeme v testu, který je součástí této práce, a v rámci kterého budeme hodnotit kvality a využitelnost tohoto systému.

### d) iTranslate

Služba iTranslate od stejnojmenné rakouské společnosti je rozdělena na čtyři samostatné aplikace – „základní“ aplikaci iTranslate, která nabízí takřka totožné funkce jako Google Translate, aplikace iTranslate Voice a iTranslate Converse, které se na první pohled takřka neliší, a aplikaci iTranslate Lingo pro učení se jazykům. Aplikace dokáží rozpoznávat a generovat řeč ve zhruba 40 jazycích, včetně češtiny. Jelikož z webu výrobce není zcela patrný rozdíl mezi aplikacemi Voice a Converse, rozhodli jsme se je v rámci této práce krátce

vyzkoušet na krátkém dialogu spočívajícím v domluvě návštěvy kina v páru angličtina-čeština. Z výstupu bylo patrné, že obě aplikace používají pro překlad/tlumočení stejnou technologii, jelikož byly výstupy prakticky totožné. Hlavním rozdílem mezi aplikacemi je pak to, že zatímco aplikace Converse v bezplatné verzi nenabízí transkripty konverzací, aplikace Voice ano a v její placené verzi je pak možné získat i konverzační příručky na několik základních každodenních témat. Výhodou aplikace Converse je však automatické rozpoznávání jazyků ve vybrané dvojici – není tedy nutné aplikaci pokaždé říci, kterým jazykem na ni uživatel hovoří. Co se týče kvality výstupu, trůfáme si říct, že s přirozeným jazykem mají tyto aplikace ještě značné problémy, a to jak s jeho pochopením, tak s vytvořením přirozeně znějícího výstupu. Není také zcela zřejmé, jestli aplikace dokáží pracovat s intonací. Je však pochopitelné, že kvalitu výstupu nelze objektivně hodnotit na základě desetiminutového vyzkoušení každé aplikace. Pro potenciálního uživatele však může tento první dojem hrát zásadní roli v tom, zda aplikaci bude dále využívat, či ne.



Obrázek 6: screenshot z vlastního krátkého testování aplikace iTranslate.

#### e) Fanyijun od Tencent

Překladač Fanyijun byl čínským technologickým gigantem představen v roce 2017 a v současnosti funguje jako bezplatný online překladač a také slouží k překladu zpráv uvnitř populární čínské chatovací aplikace WeChat. Nabízí stejné funkce, jako výše zmiňovaná konkurence, jeho velká chvíle však měla přijít na Asijském fóru v Boao v roce 2018, kde měl Fanyijun zajistit simultánní přepis a tlumočení vedlejších přednášek, které pak měly být takto tlumočené vysílány na monitorech na konferenci a také živě pro ty, kteří ji sledovali prostřednictvím WeChatu. Velká míra publicity před událostí se však výrobcí vymstila a

„nesmyslné brebentění“<sup>33</sup> překladače proto oběhlo novinové titulky po světě a vyvolalo debaty o tom, zda bude umělá inteligence skutečně někdy schopna nahradit tlumočníky.

### 3.4.2 Přenosná zařízení

Přenosná zařízení pro automatické tlumočení mají nejčastěji podobu sluchátek a k jejich fungování je často (ne však vždy) nutná speciální aplikace. Podstata toho, jak tlumočení funguje, je tedy stále stejná – převod z jednoho jazyka do druhého stále probíhá pomocí systému založeného na korpusovém překladači. Jako vstupní/výstupní zařízení však neslouží mobilní telefon, ale nejčastěji sluchátka – oba účastníci konverzace mají po jednom sluchátku a stisknutím tlačítka na svém sluchátku dají aplikaci najevo, že má poslouchat a uslyšené převést do druhého jazyka. Kromě toho jsou však na trhu i zařízení, která pracují bez aplikace a vzhledem připomínají diktafon.

#### a) **Sluchátková zařízení**

##### i. Pilot

Pilot má podobu dvou oddělených sluchátek. Výrobce uvádí podporu 15 jazyků a 42 dialektů, v aplikaci nabízí také slovník a konverzační příručku a slibuje přirozené hlasy. Pilot je možné používat ve dvou módech a také k poslechu hudby, jako klasická bezdrátová sluchátka. Konverzační mód funguje tak, že každý z účastníků má jedno sluchátko a každý také musí mít aplikaci Pilot na svém telefonu. Pak už se jen střídají v mluvení, zatímco na displeji svého telefonu může každý vidět přepis konverzace. Dále Pilot nabízí funkci tzv. Listen módu, který dokáže tlumočit mluvčího, který je poblíž řečníka, ale nemá sluchátko a mluví do mikrofону chytrého telefonu. Dostupné recenze tohoto zařízení (získat ho od výrobce pro testování se nám bohužel nepodařilo) hovoří o značných problémech v rušnějším prostředí, jelikož zařízení poslouchá na určité frekvenci, kterou ale můžou lidé v okolí přerušit. Zařízení je možné zakoupit na webu výrobce a jeho běžná prodejní cena činí 249 USD<sup>34</sup>.

---

<sup>33</sup> AI-powered translation still needs work after errors mar debut at Boao Forum | South China Morning Post. International Edition | South China Morning Post [online]. Copyright © 2019 South China Morning Post Publishers Ltd. All rights reserved. [cit. 09.04.2019]. Dostupné z: <https://www.scmp.com/tech/innovation/article/2141940/ai-powered-translation-still-needs-work-after-errors-mar-debut-boao>

<sup>34</sup> Více informací čtenář nalezne na webových stránkách výrobce: <https://www.waverlylabs.com/products-bf/>





Obrázek 7: Sluchátka Pilot a aplikace Pilot, převzato z webu výrobce.

ii. WT2 Plus

Velice podobně jako Pilot funguje zařízení WT2 Plus, kde však stačí mít přidruženou aplikaci pouze na jednom chytrém telefonu. Dále pak WT2 nabízí také plně automatický režim, kdy se uživatelé nemusí svých sluchátek vůbec dotýkat a rozpoznávání promluv, které mají být přetlumočeny, probíhá zcela automaticky. Tento režim se hodí především do tiššího prostředí bez rušivých okolních zvuků. WT2 nabízí také tzv. Ask mode, který je určen pro rychlou komunikaci, kdy protistrana dokáže na otázky odpovědět gesty či jednoduchými výrazy jako ano a ne. V tomto režimu stačí sluchátko pouze tazateli – dotazovanému pak přetlumočený dotaz přehraje chytrý telefon. Zařízení lze rovněž zakoupit, a to za 219 USD.

NATURAL POSTURE BRINGS NATURAL COMMUNICATION

- Real-time Automatic Audio Interaction**
- Voice Latency**  
1-3 seconds
- Free posture**  
Face-to-face communication as usual.
- Charging Case**  
Automatically wake up translating earphones  
Charging at any time.

- Smart Speech Translation Switch**
- Translation Accuracy**  
Meet daily communication
- Three Operation Mode**  
Auto Mode, Manual Mode, Ask Mode
- Ready to use**  
Just one App  
No need to download

Obrázek 8: Sluchátka WT2 Plus, převzato z webu cool-mania.cz.



### iii. Translate One2One

Toto zařízení je dílem australské společnosti Lingmo International a slibuje podporu 27 jazyků. Poprvé bylo představeno v roce 2017 na summitu „AI for Good“ pořádaném pod záštitou OSN a je založeno na technologii počítače Watson od společnosti IBM, který byl původně vyroben k tomu, aby se pokusil porazit lidské soutěžící v populární americké vědomostní soutěži Jeopardy! (obdoba českého pořadu Riskuj!). Na rozdíl od jiných sluchátkových zařízení Translate One2One<sup>35</sup> nevyžaduje připojení k žádným přidruženým aplikacím ani dalším zařízením a je tak v tomto směru zcela autonomní – místo toho používá speciální sim kartu Lingmo Travel SIM card. Zařízení se ovládá pomocí dotykového displeje, kterým je každé sluchátko vybaveno. Běžná cena zařízení na webu výrobce je 249 AUD.

#### **b) Kapesní zařízení**

Kromě nepřehledného množství strojových překladačů a automatických tlumočnicků, které fungují jako aplikace pro chytré telefony, najdeme na trhu i autonomní kapesní zařízení, jejichž hlavní výhodou bývají drobné rozměry, a především schopnost fungovat bez připojení k internetu (nicméně tuto možnost nabízí i řada mobilních aplikací).

#### i. ili

Ze všech zmíněných zařízení je ili<sup>36</sup> co do jazykové vybavenosti nejskromnější – tlumočí pouze z angličtiny do japonštiny, španělštiny a mandarínštiny. Jeho jednosměrnost je záměrem výrobce, který tvrdí, že při testování dvousměrného tlumočení narazil na potíže s tím, že si uživatelé nebyli jisti, jak se zařízením spolupracovat a používali například příliš dlouhé a košaté věty či slangové výrazy, nebo zařízení nedůvěřovali a přistupovali k němu se značnou obezřetností. Podle doporučení výrobce se tedy zařízení hodí především pro cestovatele, kteří pomocí něj mohou pokládat jednoduché dotazy z oblasti dopravy, nakupování, či objevování památek, neboť právě na tato témata se zařízení zaměřuje. Kromě toho výrobce upozorňuje také na to, že ili dokáže tlumočit pouze po jednotlivých větách a zařízení tak vyžaduje od uživatele značnou míru přizpůsobení. Prodejní cena zařízení je 199 USD.

### 3.4.3 Ostatní zařízení a systémy

Kromě zařízení, která míří na běžné spotřebitele, se objevují rovněž pokusy o vytvoření systémů, které by se dokázaly vyrovnat simultánním tlumočnickům. V této kapitole se budeme věnovat jednomu takovému systému, přestože není vybaven funkcí syntézy hlasu a jedná se tak spíše o jakýsi „simultánní překlad“.

#### STACL od Baidu

Na podzim roku 2018 oznámil čínský gigant Baidu svůj nejnovější výdobytek – systém STACL (Simultaneous Translation with Anticipation and Controllable Latency), první systém automatického tlumočení se schopností anticipace a možností regulace časového posuvu.

<sup>35</sup> Translate One2One Real-Time Translation Earpieces – Lingmo International. Lingmo International – Understand the World [online]. Copyright © [cit. 09.04.2019]. Dostupné z: <https://lingmointernational.com/translate-one2one/>

<sup>36</sup> ili – Instant offline translation device for travelers. [online]. Dostupné z: <https://iamili.com/us/>

V základu se stále jedná o systém založený na korpusech a neurálních sítích, který musí být „vycvičen“ pomocí ohromného množství dat. Díky poměrně jednoduchému řešení ale STACL přinesl průlom v tom, co je pro stroje na simultánním tlumočení nesmírně těžké – totiž to, že různé jazyky mají různý slovosled a různě řadí podmět, přísudek a předmět. Dostupné komerční systémy tento problém řeší tak, že si počkají na celou větu a teprve pak vyprodukují překlad, což ale znamená značné zpoždění. Baidu však představuje poměrně jednoduchý systém pro predikci, kdy se v systému nastaví požadovaný posuv pomocí modelu „wait-*k*“, kdy *k* představuje požadovaný posuv vyjádřený počtem slov, který si systém pak po celou dobu za řečníkem udržuje, což mu umožňuje získat kontext a na základě něj pak predikovat. STACL navíc také nepředpovídá slova ve výchozím jazyce, ale rovnou v jazyce cílovém, k čemuž používá jak dostupnou část vstupu ve výchozím jazyce, tak to, co již převedl do jazyka cílového – jinými slovy tedy dokáže pracovat s kontextem. Možnost nastavení délky posuvu dává systému STACL značnou flexibilitu – pro blízké jazyky, jako například španělština a italština, si tak uživatel může nastavit krátký posuv, zatímco pro jazyky, které jsou si vzdálené nebo mají jinou strukturu slovosledu lze pro zajištění optimální kvality výstupu nastavit posuv větší (ovšem za cenu delšího časového posuvu výstupu). Při hodnocení kvality pomocí systému BLEU, který podrobněji popíšeme v kapitole 4.3, dosahuje STACL skóre o 3.4 bodu nižšího, než běžný neurální (nesimultánní) převod, který pracuje s celými větami za cenu vyššího posuvu. Systém STACL měl navíc být otestován na Světové konferenci Baidu v Pekingu 1. listopadu, ale informace o tom, zda k tomu skutečně došlo a jaké byly výsledky zatím nejsou dostupné.

### 3.5 Automatické tlumočení versus lidské tlumočení

V jedné z předchozích kapitol jsme se již zamýšleli nad tím, jestli je možné automatické tlumočení vůbec za skutečné tlumočení považovat, a došli jsme k závěru, že současné technologie znaky tlumočení nenaplnují. V této kapitole se však zaměříme na jeden systém, který se cíleně pokouší napodobit některé strategie lidských simultánních tlumočnicků, a pokusíme se porovnat tento systém s modelem úsilí Daniela Gila a určit, zda je možné srovnat procesy, které probíhají v mozku tlumočnicka s tím, jak pracuje daný systém automatického tlumočení. Dále se pokusíme tento systém porovnat se strategiemi a postupy, které tlumočnick při výkonu vědomě používá, a popsat, zda je možné je u zvoleného systému automatického tlumočení pozorovat či ne. V této kapitole tedy vždy nejdříve stručně popíšeme každou strategii či postup a poté se pokusíme určit, zda je možné ji u posuzovaného programu pozorovat či ne.

Systémem, který budeme v této kapitole porovnávat s modelem úsilí a tlumočnickými strategiemi, je čínská novinka STACL od Baidu popsaná výše, neboť právě tento systém podle našeho názoru nejvíce naplňuje znaky skutečného simultánního tlumočení. Porovnání s novým systémem STACL uvádíme i přesto, že tento systém zatím nevytváří výstup v podobě řečové syntézy, neboť jsme toho názoru, že pokroky, které přináší, jsou pro srovnání s lidským tlumočením dostatečně zajímavé. Dalšími důvody, proč jsme pro srovnání zvolili právě tento systém je fakt, že je princip fungování tohoto systému popsán v odborném článku a především to, že se nám prostřednictvím sítě LinkedIn podařilo spojit s jedním z členů vývojového týmu

STACL, který ochotně odpovídal na naše dotazy. Vzhledem k tomu, že v současné době tým pracuje na rozšíření systému o další vlastnosti a schopnosti napodobovat lidské tlumočnické strategie si však tento člen týmu nepřál být v naší práci jmenován. V této práci ho tedy budeme označovat výrazem „odborný poradce“.

Jiné komerčně dostupné systémy jsme se rozhodli z porovnání vynechat, a to především proto, že dle našeho názoru nenaplnují znaky tlumočení, dále proto, že drtivá většina z nich představuje pouze textové překladače rozšířené o možnost hlasového vstupu a výstupu, a také vzhledem k jejich množství a (ne)dostupnosti podrobných a dostatečně odborných informací o principech jejich fungování.

### 3.5.1 Model úsilí Daniela Gila

Ve svém modelu úsilí, který Gile aplikuje na tlumočení simultánní, konsekutivní, z listu a s listem, se Daniel Gile pokusil popsat mentální procesy, ke kterým dochází v mozku tlumočnicka v průběhu tlumočení. Model je založen na existenci tzv. „úsilí“, která tlumočnick při svém výkonu vynakládá, a na tom, že pracuje s určitou pracovní kapacitou, kterou tato úsilí využívají. Úsilí Gile definuje takto:

- a) úsilí poslechu a analýzy, tedy pochopení smyslu výchozího textu a jeho interpretace, nikoliv pouhé rozpoznání jednotlivých slov,
- b) úsilí krátkodobé paměti, tedy uchovávání informací mezi poslechem a produkcí,
- c) úsilí produkce, které má mít za výsledek správně a srozumitelně podané sdělení,
- d) úsilí koordinace, které řídí rozložení pracovní kapacity mezi předchozí tři úsilí a jejich koordinaci.

Základní myšlenkou modelu tedy je, že tlumočení vyžaduje určitou mentální energii (pracovní kapacitu), které je omezené množství, a zároveň, že tlumočnick při výkonu využívá tuto energii téměř všechnu. Pokud nároky na kapacitu převyšují její dostupný objem (dojde tedy k zahlcení pracovní kapacity) či je tato kapacita rozdělena v daný okamžik mezi jednotlivá úsilí tak, že nestačí k adekvátnímu zpracování určitého úseku (dojde tedy k individuálnímu deficitu a selhání v konkrétním úsilí, na které nebyl vyčleněn dostatek kapacity), dojde k selhání. To se nemusí projevit vždy, jelikož je možné přesunout nevyužitou kapacitu z jednoho úsilí do druhého, pokud se však selhání projeví, záleží jeho podoba na tom, kde přesně k selhání došlo. Jeho projevem pak může být ztráta či posun informace, interference, horší jazyková kvalita projevu atd. Zdrojů problémů, které mohou způsobit selhání, může být při tlumočení celá řada: Gile jmenuje hustotu informací v textu, rychlost řeči, kvalitu zvuku, náročnou terminologii, silný přízvuk řečníka či nesprávnost gramatiky a lexika, neobvyklý styl projevu či argumentace, neznámá jména, čísla, názvy a zkratky. Gile také tvrdí, že se selhání zřídka projeví přímo v okamžiku, kdy k němu došlo a projeví se spíše později, často v úseku, který vůbec problematický není. Toto Gile vysvětluje existencí tzv. deficitního řetězce.

### 3.5.2 Model úsilí vs STACL

Úvodem je nutné říci, že srovnávat člověka a stroj z kapacitního hlediska je prakticky nemožné, jelikož kapacita stroje, byť má také svá výpočetní omezení, je mnohonásobně vyšší než kapacita lidského mozku. Hovořit na straně počítače o úsilí jako takovém, ve smyslu vynaložení určité části kapacity k určitému úkonu, tedy považujeme za bezpředmětné, jelikož bychom porovnávali neporovnatelné. Pokusíme se tedy alespoň popsat, zda počítač určitý úkon (úsilí) provádí či ne, bez ohledu na kapacitní omezení.

#### a) Úsilí poslechu a analýzy

Jedinečnou vlastností systému STACL je jeho schopnost predikce, kterou jsme již popsali v předchozí kapitole. Ta mu umožňuje nepracovat pouze s významem jednotlivých slov, ale s celkovým významem věty či sdělení, jak sám výrobce ukazuje na příkladu čínštiny a angličtiny, kdy STACL při posuvu o dvě slova správně odhadne, že by měl použít výraz „meets“, přestože v čínském originálu sloveso ještě nezaznělo (viz obr. 9). Samotný proces predikce jako takový je umožněn díky redundanci jazyka a probíhá na několika úrovních, přesněji na úrovni slabiky, slova, syntagmatu, výpovědi, sdělení a komunikativní výpovědi, většinou však probíhá na více úrovních najednou (Čeňková 2001). Budeme-li tedy predikci považovat za prostředek analýzy textu, je možné tvrdit, že systém STACL vstupní text skutečně analyzuje. Dalším prvkem analýzy, který můžeme u systému STACL pozorovat, je analýza kontextu. STACL totiž pracuje se dvěma modely neurální architektury pro pochopení jazyků, totiž s architekturou RNN, kterou jsme již popsali v předchozí části práce (viz kapitola 2.2), a s novou architekturou Transformer<sup>37</sup> od Googlu, která je vylepšenou verzí RNN a disponuje tzv. „self-attention“ mechanismem. Tento mechanismus Transformer aplikuje na vstupní text při každém kroku, čímž se vytváří vztah mezi jednotlivými slovy nezávisle na jejich pozici ve větě. Díky tomu je tak STACL schopen analyzovat sdělení na základě kontextu a s každým novým slovem tento kontext aktualizovat a upravovat.

Pokud se pak zaměříme na samotný poslech, stejně jako člověka mohou i výkon stroje a jeho schopnost dobře rozpoznat hlasový vstup řečníka ovlivnit faktory spojené s kvalitou zvukového vstupu, řečníkův přízvuk, či gramatická a lexikální adekvátnost a správnost jeho projevu, neboť počítač je „cvičen“ na korpusu projevů se standardními gramatickými a lexikálními vlastnostmi. Je tedy možné říci, že jak u člověka, tak i u stroje může dojít na úrovni poslechu k selhání, které je však u stroje způsobeno pouze externími faktory, nikoliv nedostatkem jeho pracovní kapacity.



Obrázek 9: ukázka fungování systému STACL, převzato z tiskové zprávy společnosti Baidu.

<sup>37</sup> Uszkoreit, Jakob. Transformer: A Novel Neural Network Architecture for Language Understanding. 2017. Dostupné z: <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

### **b) Úsilí krátkodobé paměti**

Pro přítomnost prvku krátkodobé paměti u STACL opět hovoří využití architektur RNN a Transformer, jelikož se zakládají na tom, že chápou jazyk prostřednictvím vektorových zobrazení o pevně dané nebo variabilní délce. Začínají tedy například na zobrazení slova či dokonce jeho části a postupně shromažďují informace z přibývajících okolních slov, což značí, že si tyto výrazy „pamatují“. Zda a na jak dlouho pak při procesu tlumočení tyto výrazy zůstávají v paměti STACLu se nám bohužel zjistit nepodařilo.

### **c) Úsilí produkce**

STACL produkuje výstup v podobě „simultánního překladu“ a ve fázi poslechu zároveň dochází také k transkripci vstupu. Vzhledem k tomu, že je STACL založen na strojovém učení pomocí korpusů paralelních textů, kde lze předpokládat gramatickou a lexikální správnost tohoto tréninkového materiálu, lze očekávat i gramatickou a lexikální správnost a adekvátnost výstupu. Srozumitelnost výstupu (např. správnost terminologie) je pak ovlivněna tím, zda systém dostal dostatečné množství textů k danému tématu, což lze přirovnat k přípravě tlumočnicka před akcí. S ohledem na to, že cílem STACLu je učinit simultánní tlumočení (nebo simultánní překlad, jak uvádí Baidu<sup>38</sup>) dostupnějším tedy lze konstatovat, že systém skutečně vyvíjí snahu o zajištění co nejkvalitnějšího výstupu.

### **d) Úsilí koordinace**

Jak jsme již zmiňovali v úvodu tohoto porovnání, srovnávat člověka a stroj z kapacitního hlediska považujeme za bezpředmětné vzhledem k nepoměru mezi procesní kapacitou člověka a procesní kapacitou počítače, která je mnohonásobně větší. Z tohoto důvodu tedy nelze v praxi očekávat její vyčerpání a s tím spojené selhání – pokud k němu u automatického tlumočení dojde, je nutné hledat příčinu v nedokonalém provedení jednoho z předchozích úsilí. Bylo by samozřejmě možné zkoumat, jakou část operační paměti počítače RAM každé z těchto úsilí vytěžuje a zda se toto rozložení mění či ne, jednalo by se však už o práci z oblasti výpočetní techniky a umělé inteligence, nikoliv o práci z oblasti tlumočení, a musíme proto konstatovat, že v rámci této práce nedisponujeme dostatkem prostředků k tomu, abychom v této otázce mohli činit jakékoliv závěry či předpoklady.

### **3.5.3 Strategie a postupy v simultánním tlumočení**

Úvodem je nutné zmínit, že problematice tlumočnických strategií se věnuje řada významných badatelů jako Gile, Černov, Širjajev, Pöschhacker a další, což však vede k problému s nejednotností terminologie. Někteří totiž rozlišují mezi strategií a postupem, kdy první chápou jako celkový přístup k tlumočení projevu či jeho části a druhé jako konkrétní mechanismus, který tlumočnick používá k vypořádání se s problematickým prvkem. Jiní však mezi těmito pojmy nerozlišují a používají je jako synonyma, čehož se budeme v této práci držet rovněž a používat zde termín strategie. Vycházet budeme z Gilova popisu a dělení těchto strategií, které představil ve své publikaci *Basic Concepts and Models for Interpreter and*

---

<sup>38</sup> Baidu Research. Baidu Research [online]. Copyright © 2018 Baidu Research [cit. 09.04.2019]. Dostupné z: <http://research.baidu.com/Blog/index-view?id=107>

Translator Training (1995), kde dělí tlumočnické strategie na strategie ve fázi pochopení, preventivní strategie a reformulační strategie. Z důvodu povahy srovnání stroje a člověka z této kapitoly zcela vyloučíme extralingvistické strategie jako kontaktování organizátorů akce v případě problémů, psaní poznámek či pomoc od kolegy v kabině.

#### a) Strategie ve fázi pochopení

Tyto strategie tlumočnické používá ve fázi poslechu, ještě než začne sám hovořit. Základní strategií je zde práce s časovým posuvem, která umožňuje tlumočnickovi pružně reagovat na projev řečníka.

##### i. Časový posuv

Práce s časovým posuvem patří k jedné ze základních strategií tlumočnické, který si ho může upravovat dle aktuální potřeby. Při výčtu čísel tedy tlumočnick může zvolit posuv velmi krátký, zatímco v případě složitější argumentace může volit naopak posuv delší, aby získal více vstupních informací a měl více času na rozmyšlenou. Prodlužování posuvu s sebou však přináší riziko nakumulování příliš mnoha informací v krátkodobé paměti a následné selhání.

##### a. STACL a časový posuv

Právě práce s časovým posuvem patří k jedné z průlomových vlastností systému STACL. Ten v základu pracuje s pevným posuvem definovaným jako „wait- $k$ “, kdy  $k$  představuje počet slov. Autoři však navrhují také možnost práce s tímto posuvem, tzv. catchup:

*„...the decoder of proposed “wait- $k$ ” process is always  $k$  words behind the incoming source stream. In the ideal case which the source and target sentences are in the same length, the last  $k$  words on target side will be generated without waiting for the encoder since there is no new word expected on encoder side (for example, a period on source side indicates the source sentence finishes).“* (Ma et. al, 2018)

Vzhledem k textové formě výstupu STACLu však může tento catchup představovat problém pro uživatele, kteří budou nečekaně vystaveni nutnosti změnit rychlost čtení, což může být problematické v jazykových párech, kde je cílový jazyk delší než jazyk výchozí.

##### ii. Rekonstrukce segmentu z kontextu

Tuto strategii tlumočnick volí, pokud určitý výraz, číslo, jméno či jiný typ řečového segmentu neslyšel dostatečně dobře či vůbec. Pokusí se ho tedy rekonstruovat na základě kontextu – vlastní znalosti jazyka, tématu, situace a dalších informací od řečníka.

##### a. STACL a rekonstrukce segmentu z kontextu

Na základě informací, které jsou o algoritmech RNN a Transformer dostupné, a na základě pokusů s využitím těchto algoritmů k nejrůznějším predikcím<sup>39</sup> jsme předpokládali, že STACL schopností doplňovat chybějící výrazy na základě kontextu disponuje. Abychom se vyhnuli spekulacím, obrátili jsme se na jednoho z autorů systému STACL, který nám potvrdil, že v rámci práce s predikcí dokáže STACL rovněž na základě okolního kontextu učinit pokus o uhodnutí

---

<sup>39</sup> What Is a Transformer? — Inside Machine Learning – DZone AI. DZone [online]. Dostupné z: <https://dzone.com/articles/what-is-a-transformer-inside-machine-learning>

chybějícího výrazu, podobně, jako lidský tlumočník. Zvláště dobrých výsledků v tomto případě dosahuje, podle vyjádření našeho odborného poradce, algoritmus Transformer, a to díky již zmiňovanému „self-attention“ mechanismu.

### **b) Preventivní strategie**

Tyto strategie, jak název napovídá, pomáhají tlumočnickům v situacích, kdy cítí, že tlak na jejich procesní kapacitu je takový, že by mohlo dojít k selhání, a využívají proto těchto strategií, aby selhání zabránili.

#### *i. Práce se zpožděním mezi poslechem a řečí*

Pomocí této strategie může tlumočník do určité míry řídit, jaké množství procesní kapacity jednotlivá úsilí využívají. Zkrácením zpoždění řeči za poslechem může tlumočník ulevit krátkodobé paměti, což může být užitečné například při výčtech, ale zároveň si tím zhorší možnosti anticipace, což mu může způsobit potíže s dokončením věty v cílovém jazyce. Delším zpožděním naopak zlepší možnosti anticipace, ale hrozí přetížení krátkodobé paměti.

#### *a. STACL a práce se zpožděním mezi poslechem a řečí*

Jak už jsme zmínili, STACL zatím nevytváří hlasový výstup, pro účely srovnání tedy budeme pracovat se zpožděním mezi poslechem a textovým výstupem. Schopnost dynamicky pracovat se zpožděním mezi poslechem a řečí (produkcí textu) by vyžadovala schopnost systému dynamicky pracovat s modelem „wait- $k$ “ a hodnotu  $k$  si upravovat samostatně. Zatím tomu tak, podle vyjádření odborného konzultanta, není – před výkonem se v systému nastaví fixní hodnota  $k$ , které se systém drží po celou dobu.

#### *ii. Segmentace*

Tuto strategii tlumočník využívá v případě tlumočení mezi jazyky s velice odlišnou syntaxí, v případě komplikovaných větných struktur, nebo když hrozí přetížení krátkodobé paměti. Pomocí této strategie si tlumočník rozdělí projev na menší segmenty, než jaké by volil běžně, a reformuluje je dříve, někdy i natolik, že ještě nemá ucelenou představu o tom, co chce řečník říci. V takovém případě často volí neutrální a co nejméně svazující formulace.

#### *a. STACL a segmentace*

Při současném stavu systému STACL nelze přesvědčivě tvrdit, že systém dokáže pracovat se segmentací dokonale, což je dáno (zatím) rigidní povahou jeho wait- $k$  modelu. V případě jazyků se silně odlišnou syntaxí řeší STACL tento problém možností nastavení vyšší hodnoty  $k$  pro celý proces převodu. Stejným způsobem by bylo možné řešit i komplikované větné struktury, podmínky stroje a tlumočnicka se však zde liší. Zatímco tlumočník má omezené možnosti dopředu zjistit, jak bude řečník hovořit, hodnota  $k$  se v systému STACL nastavuje manuálně a jelikož se stále jedná o rozpracovaný systém, je na místě předpokládat, že při testování zadavatel dopředu ví, jaký druh projevu bude systém „tlumočit“ a může tak nastavit odpovídající hodnotu  $k$ . Jak jsme se však dozvěděli od odborného konzultanta, systém zatím nedokáže samostatně dynamicky tuto hodnotu  $k$  měnit.

### *iii. Změna pořadí prvků ve výčtu*

Tato strategie spočívá v tom, že v případě výčtů, které mají vysokou informační hustotu a představují tak značnou zátěž pro krátkodobou paměť tlumočnick převede poslední prvek výčtu do cílového jazyka jako první. Není zcela zřejmé, jakým způsobem tato strategie přispívá ke snížení zátěže krátkodobé paměti, Gile se však domnívá, že převedení posledního prvku do cílového jazyka ještě dříve, než dojde k jeho hlubokému zpracování a uložení do sémantické sítě umožňuje ušetřit procesní kapacitu.

#### *a. STACL a změna pořadí prvků ve výčtu*

Vzhledem k tomu, že tato strategie je zaměřena především na ušetření procesní kapacity a úlevu krátkodobé paměti si troufáme tvrdit, že pro STACL není tato strategie vůbec potřebná, jelikož počítač žádná taková kapacitní omezení nemá. Fixní podoba *wait-k* modelu navíc zaručuje, že po získání prvotního *k* počtu slov „přetlumočí“ stroj další výraz vždy, když se na straně zdroje přidá nové slovo, a tudíž systém uslyšený výraz zapomenout ani nemůže, na rozdíl od lidského tlumočnicka, u kterého může dojít k přetížení kapacity.

### **c) Reformulační strategie**

K tomu, aby tlumočnick věrně předal obsah řečnickova sdělení, využívá rovněž řadu specifických strategií. To však neznamená, že je nutné přesně kopírovat řečnickovy výrazy či slovosled, a právě proto přichází na řadu reformulační strategie.

#### *i. Časový posuv*

Tuto strategii Gile ve své publikaci zmiňuje jak v rámci strategií ve fázi pochopení, tak v rámci strategií reformulačních – v této fázi ji lidský tlumočnick využívá k tomu, aby získal čas pro hledání vhodného výrazu či vhodnou formulaci sdělení v cílovém jazyce. Jelikož jsme se této strategii v rámci systému STACL již věnovali na předchozích stránkách, odkážeme čtenáře na příslušný oddíl výše.

#### *ii. Explanace či parafrázování*

K této strategii se tlumočnick uchyluje v situaci, kdy výrazu ve výchozím jazyce rozumí, ale nezná jeho přesný ekvivalent v jazyce cílovém. V takovém případě tedy tento výraz vysvětlí či opíše jinými slovy. Pro lidského tlumočnicka však může mít tato strategie i nevýhody – je totiž náročná na procesní kapacitu a čas a také může zbytečně strhávat posluchačovu pozornost a snižovat tak důvěryhodnost tlumočnickova výkonu.

#### *a. STACL a explanace či parafrázování*

Vzhledem k tomu, že před používáním je nutné systém STACL „vycvičit“ pomocí rozsáhlého korpusu paralelních textů, je šance vzniku situace vyžadující použití této strategie při kvalitním obsahu a tematické relevanci tréninkového korpusu minimální. Dle vyjádření odborného konzultanta používá systém STACL k práci s neznámými slovy tzv. „Byte Pair Encoding“ (BPE), který neznámé výrazy rozdělí na segmenty a vytvoří z nich tzv. „podřlova“, na jejichž základě



se poté pokusí o překlad<sup>40</sup>. Tento postup však rozhodně nelze označit za parafrázování či vysvětlení.

iii. Reprodukce zvuku z originálního sdělení

Tuto strategii může tlumočnický využít v situaci, kdy narazí na neznámý výraz či například jméno a rozhodne se ho v cílovém jazyce reprodukovat stejně, jak zní v jazyce výchozím.

a. STACL a reprodukce zvuku z originálního sdělení

Systém STACL v současné době nepodporuje výstup v podobě řečové syntézy, aplikovat tuto strategii je tedy pro něj zatím z technických důvodů nemožné.

iv. Naturalizace

Strategie naturalizace spočívá v přizpůsobení výrazu z výchozího jazyka morfologickým či fonologickým pravidlům jazyka cílového a tlumočnický ji může použít v případě, že nezná přesný ekvivalent určitého termínu v cílovém jazyce. Tato strategie bývá úspěšná především v případě jazyků, které jsou si morfologicky blízké, u oborů, jejichž terminologie využívá velké množství výpůjček ze zdrojového jazyka, a pokud jsou posluchači seznámeni s danými termíny ve výchozím jazyce.

a. STACL a naturalizace

Vzhledem k tomu, že systém STACL zatím nepředstavuje finální produkt a na jeho funkcích a schopnostech se stále pracuje, neexistuje dostatek dostupných příkladů jeho využití k posouzení toho, zda je systém takto pokročilé operace schopen. Současné neurální korpusové překladače bývají v situacích, kdy narazí na slovo, které není v jejich lexikonu, bezmocné a většinou ho jednoduše nepřeloží, ale ponechají ve výchozím jazyce. Některé z nich, například Google Translate, se pokusí takové neznámé slovo rozložit na části a správný překlad pak složit z nich<sup>41</sup>. Vzhledem k tomu, že je systém STACL rovněž založen na korpusovém systému a využívá výše zmíněný algoritmus BPE, je možné přesvědčivě předpokládat, že by naturalizace schopen nebyl, jelikož by se při dobře zvoleném korpusu vůbec neměl dostat do situace, ve které by se setkal s neznámým termínem.

v. Překódování

Tuto strategii může tlumočnický využít v podobných situacích, jako výše zmíněnou naturalizaci. Podstatou překódování je doslovný převod termínu či slovního spojení do cílového jazyka. V oborech, jejichž terminologie vznikala kalkováním, může tato strategie dokonce vést k existujícímu termínu v cílovém jazyce.

a. STACL a překódování

Stejně jako v případě předchozí strategie, ani zde nedisponujeme dostatkem materiálu k posouzení přítomnosti této strategie v systému STACL. Vzhledem k její podobnosti se

---

<sup>40</sup> Sennrich, Rico, Haddow, Barry, Birch, Alexandra. (2015). Neural Machine Translation of Rare Words with Subword Units. Dostupné z: <https://arxiv.org/pdf/1508.07909.pdf>

<sup>41</sup> A Neural Network for Machine Translation, at Production Scale [online]. 27.9.2016 [cit. 2019-04-09]. Dostupné z: <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>

strategií naturalizace však můžeme předpokládat, že ani překódování systém STACL schopen není.

vi. Vynechání informace

Tato strategie představuje záměrné vynechání informace tlumočnickem, nikoliv vynechání v důsledku selhání krátkodobé paměti či nedostatku procesní kapacity ve fázi poslechu. Učinit rozhodnutí určitou informaci vynechat může tlumočnick z několika důvodů – může se jednat o redundantní informaci a tlumočnick tak uplatní princip ekonomie, nebo si tlumočnick není jistý, že informaci pochopil správně, a raději ji proto vynechá, než aby riskoval chybný převod.

a. STACL a vynechání informace

Jak vyplývá z popisu strategie výše, k vynechání informace dochází na základě vědomého rozhodnutí tlumočnicka v danou chvíli, což předpokládá schopnost dynamického myšlení a rozhodování. Systém STACL však v současné době žádnými dynamickými funkcemi nedisponuje.

vii. Paralelní reformulace

Jedná se o extrémní strategii, kterou se tlumočnick může rozhodnout využít když má pocit, že je nezbytné aby mluvil i přesto, že není za aktuálních podmínek schopen informaci adekvátním způsobem přijmout, zpracovat a přeformulovat. V takovém případě se pokusí produkovat vlastní výstup, který bude zapadat do zbytku řečnickova projevu, nebude ale věrně odrážet obsah problematické pasáže.

a. STACL a paralelní reformulace

Využití této extrémní strategie vyžaduje vlastní produkci výstupu bez jakéhokoliv zdroje. Přestože pokusy s neuronovými sítěmi vytvářejícími psané texty na základě zadaného tématu ukazují podivuhodné výsledky<sup>42</sup>, odborný konzultant z vývojového týmu STACL nám potvrdil, že taková funkce do systému integrována není.

---

<sup>42</sup>Feng, Xiaocheng, Liu, Ming, Liu, Jiahao, Qin, Bing, Sun, Yibo, Liu, Ting. Topic-to-Essay Generation with Neural Networks. IJCAI, 2018. Dostupné z: <https://www.ijcai.org/proceedings/2018/0567.pdf>

## 4 Využití, omezení a hodnocení strojového překladu a automatického tlumočení

Vzhledem k našemu předpokladu, že v současnosti dostupné systémy automatického tlumočení znaky lidského tlumočení nenaplňují (s částečnou výjimkou systému STACL), jsme se rozhodli kapitoly věnované hodnocení, využití a omezení strojového překladu a automatického tlumočení sloučit, neboť kvůli stejnému technickému základu automatického tlumočení a strojového překladu jsou i tyto charakteristiky z velké části totožné. V této kapitole budeme opět vycházet z dostupných publikací (např. The Routledge Encyclopedia of Translation Technology), z odborných statí a článků, a také z prací renomovaných autorů, které uvedeme v bibliografii v závěru práce.

### 4.1 Využití strojového překladu a automatického tlumočení

O strojovém překladu je po nástupu neurálních sítí slyšet stále víc, a to jak od překladatelů spekulujících o jeho kvalitách, tak od klientů i překladatelských agentur, které se snaží držet krok s dobou a zároveň ušetřit či zvýšit produktivitu svých překladatelů. S touto zvýšenou pozorností se pojí také polemika o využitelnosti strojového překladu (a potažmo i automatického tlumočení). Ať už se to překladatelům líbí nebo ne, strojový překlad si již vydobyl své místo a je proto potřeba s ním počítat. Samy společnosti, které strojová řešení nabízí, často také publikují na svých stránkách pro zákazníky jakési návody ohledně toho, kdy je strojový překlad vhodný, na základě jakých kritérií je potřeba se rozhodovat, a tak dále.

Pokud tedy budeme vycházet ze zkušeností samotných poskytovatelů strojového překladu a odborné literatury, lze uživatele strojového překladu rozdělit na několik základních skupin. Těmi jsou:

- firmy, které využívají na míru ušité a naučené systémy strojového překladu, například se specializací na konkrétní obor,
- poskytovatelé překladatelských služeb, kteří využívají strojový překlad v kombinaci s post-editací pro zefektivnění a zrychlení práce,
- vládní orgány, jejichž cílem je zvýšení produktivity lidských překladatelů či sběr informací,
- jednotlivci, kteří využívají bezplatné online překladače podle svých aktuálních potřeb a pro něž jsou kvalita překladu a bezpečnost dat druhořadé.

Specifickou skupinou, kterou je rovněž potřeba zmínit, je pak armáda. Zejména armáda Spojených států za desítky let vývoje strojového překladu financovala a podporovala celou řadu výzkumných programů v této oblasti s cílem lépe analyzovat cizojazyčné dokumenty a zlepšit komunikaci ve válečných konfliktech. Dnes je pak armáda hybatelem ve vývoji právě automatického tlumočení – o tomto tématu se podrobněji rozepíšeme v sekci věnované automatickému tlumočení.

Zajímavý pohled na účely využití strojového překladu nabízí John Hutchins<sup>43</sup> který uvádí, že pro posouzení vývoje a užitelnosti strojového překladu je nezbytné rozlišovat mezi čtyřmi základními potřebami v překladu. Prvním typem potřeby je tzv. **šíření** (dissemination) – tato potřeba znamená vysokou kvalitu překladu a v případě překladu strojového je tedy podle Hutchinse vždy nezbytná post-editace. Druhým typem je tzv. **asimilace** (assimilation). Kritéria na kvalitu jsou o něco nižší, hlavně v oblasti stylistiky, a hlavním cílem recipienta takového textu je pochopit hlavní myšlenku textu – špatný překlad je pro něj tedy hodnotnější než žádný překlad. Třetím typem potřeby je tzv. **sdílení informací** (interchange), kdy se dva uživatelé potřebují dorozumět a pochopit smysl vzájemných sdělení okamžitě (například v případě SMS, e-mailů, chatů). Posledním typem potřeby, kde strojový překlad hraje roli je tzv. **přístup k informacím** (information access). V tomto případě je strojový překlad integrován do systémů pro získávání informací.

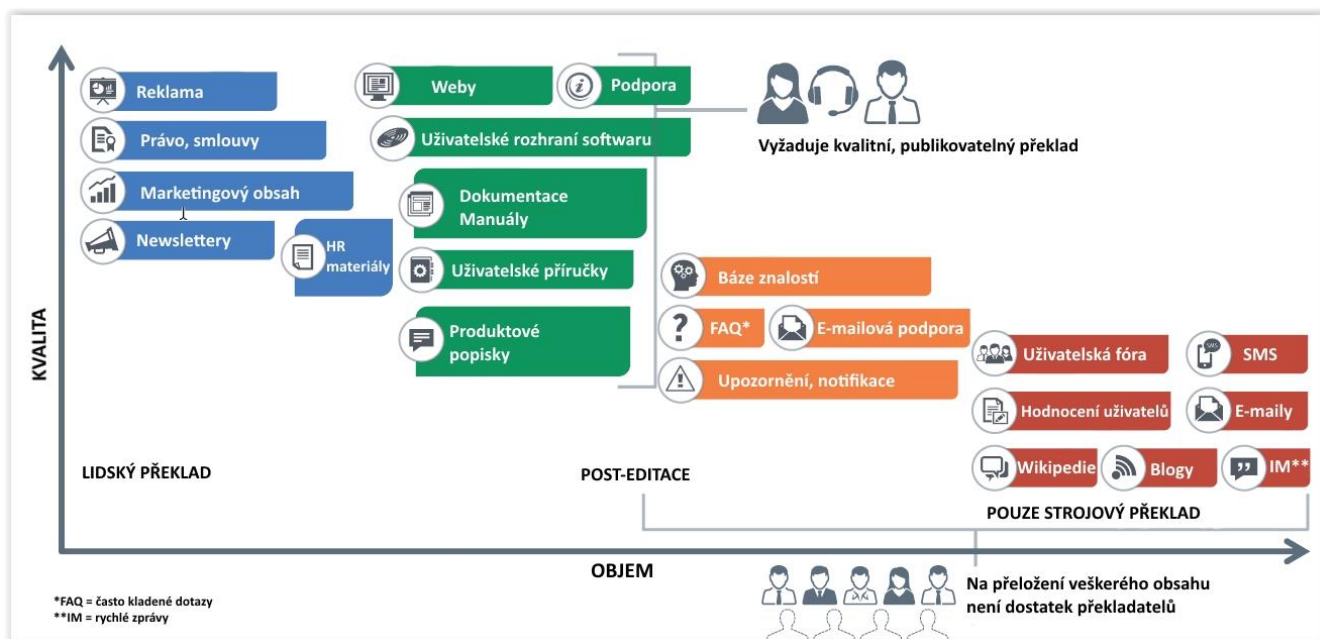
Vyvstávají také nejrůznější otázky ohledně kvality strojového překladu, například jak dobře si poradí s kreativními či marketingovými texty či zda dokáže podat stejně dobrý výkon ve všech jazycích. Proto je vždy nezbytné zvážit, zda je vhodné pro překlad daného textu použít stroj, nebo je lepší práci svěřit člověku. Za obecné pravidlo pak můžeme považovat, že čím „důležitější“ je obsah textu a čím kvalitnější výstup je vyžadován, tím lepší je pověřit překladem člověka. To se týká především textů propagačního či marketingového charakteru a textů, kde by chyby či nepřesnosti mohly vést k vážným a dalekosáhlým následkům (právním, zdravotním či jiným). Stejně principy lze pak aplikovat i v případě automatického tlumočení – zatímco například pro dorozumění se v situacích spojených s cestováním je hlavní, aby komunikace proběhla úspěšně, nehledě na její podobu, a uživatel se tak může spokojit s nedokonalým, ale přenosným a okamžitě dostupným systémem automatického tlumočení, na mezinárodní konferenci může hrát roli každé slovo a každá nuance, a přítomnost profesionálních lidských tlumočnicků je proto zcela klíčová. Budeme-li pak na využití automatického tlumočení nahlížet z pohledu potřeb uživatele definovaných Johnem Hutchinsem, které jsme již popsali výše, zjistíme, že nejčastější potřebou, kterou se uživatelé automatického tlumočení snaží uspokojit, je **sdílení informací** (interchange), tedy dorozumění se s jinou osobou.

Dalším důležitým aspektem při strojovém překladu je pak objem textu – strojový překlad je proto využíván tam, kde by objem textu znamenal v případě překladu člověkem týdny či měsíce práce, a kvalita je tak obětována na úkor rychlosti. Jakýmsi kompromisem je post-editace strojového překladu, kde však velmi záleží na povaze textu a kvalitě samotného softwaru a jeho korpusu a paměti. V případě kvalitního softwaru totiž může strojový překlad práci překladatele skutečně urychlit, v opačném případě ji ale naopak ještě ztíží a mnohdy se tak ukáže, že by člověk zvládl práci lépe a rychleji bez zásahu strojového překladu.

Vztah mezi kvalitou a kvantitou textu a „nástrojem“ překladu pak můžeme dobře ilustrovat pomocí grafického znázornění níže:

---

<sup>43</sup> HUTCHINS, John. The development and use of machine translation systems and computer-based translation tools [online]. 1999 [cit. 2019-04-09]. Dostupné z: <http://www.hutchinsweb.me.uk/Beijing-1999.pdf>



Obrázek 10: graf zobrazující vztah mezi kvalitou a kvantitou textu a prostředkem překlada, převzato z případové studie společnosti SDL<sup>44</sup>, překlad Yulia Skadchenko.

Z obrázku výše si lze vytvořit poměrně přesnou představu o tom, jaké druhy textů se pomocí strojového překlada zpracovávají. Obrázek zároveň ukazuje, že je stále dostatek oblastí, ve kterých je lidský překladatel nezastupitelný, což ve svých doporučeních klientům<sup>45</sup> potvrzují i sami poskytovatelé, kteří služby strojového překlada nabízejí.

Důkazem, že použít strojový překlad nelze ještě zdaleka na každý typ textu, je ostatně i sama společnost Google, tvůrce nejpopulárnějšího online překladače, která sama svůj výtvar pro překlad vlastních obchodních textů nepoužívá, jak v rozhovoru z roku 2011<sup>46</sup> uvedl překladatel a autor knihy o překlada „Is That a Fish in Your Ear?“ David Bellos: „*Google Translate se dá ve mnoha případech využívat smysluplně, a navíc se jedná o skutečně fantastický technologický a intelektuální výdobytek. Ale ani sám Google by nikdy nepoužil Google Translate k překlada svých veřejných, propagačních textů do jazyků, ve kterých prodává své služby. K tomu používá lidské překladače.*“

Zatím se tedy zdá, že se překladatelé o své živobytí obávat nemusí. Strojový překlad si sice uzurpoval část trhu s překlady, zároveň ale otevřel i nový trh post-editace. Ta navíc podle názoru některých překladatelů vyžaduje jiný soubor dovedností než klasický překlad, a práce post-editora je tak mnohdy srovnávána spíše s prací redaktora než s prací překladače.

Co do potenciálního využití automatického tlumočení lze dostupné systémy podle Seligmana a Zonga rozdělit na tři kategorie, podle toho, jak široké je jejich jazykové pokrytí, tematické

<sup>44</sup> Multilingual Chat and Machine Translation Use Case | SDL. SDL: Language Translation & Content Management Company [online]. Dostupné z: <https://www.sdl.com/download/customer-support-multilingual-chat-and-machine-translation-use-case/71534/>

<sup>45</sup> When Using Machine Translation Makes Sense (and when it doesn't) | Acclaro. Professional Translation Services | Acclaro [online]. Copyright © 2002 [cit. 09.04.2019]. Dostupné z: <https://www.acclaro.com/blog/when-using-machine-translation-makes-sense/>

<sup>46</sup> Celý rozhovor dostupný na: <https://www.npr.org/2011/11/14/142309214/meaning-of-everything-often-lost-in-translation?t=1550500359963&t=1551877319768>, překlad citátu do českého jazyka: Yulia Skadchenko

pokrytí, a kolik spolupráce vyžadují od uživatele. Níže uvádíme tabulku vytvořenou na základě jejich dělení.

Kategorie	Hlasové frázové překladače	Objemné úzce zaměřené systémy	Vysoce interaktivní systémy s širokým jazykovým a tematickým pokrytím
Jazykové pokrytí	Úzké	Široké	Široké
Tematické pokrytí	Úzké	Úzké	Široké
Míra spolupráce	Nízká	Střední	Vysoká

Tabulka 1: kategorizace systémů automatického tlumočení.

### Hlasové frázové překladače

Tyto systémy, jako například zmiňované kapesní zařízení ili, mohou sloužit jako databáze užitečných frází z oblasti cestování, stravování, dopravy a tak dále. Svému uživateli však neposkytují prakticky žádnou volnost co do vyjádření, jelikož si uživatel musí vybírat z uložených frází či šablon. Z uživatelského pohledu se tedy jedná o zařízení, které může pomoci s komunikací v oblasti, na kterou se zaměřuje, a zároveň přináší několik dalších výhod jako přenosnost, skladnost, možnost využití bez připojení k internetu a díky úzkému tematickému zaměření například také lepší kvalitu nahraných frází v cizím jazyce.

### Objemné úzce zaměřené systémy

V porovnání s výše uvedenou kategorií může takový systém nabídnout uživateli mnohem větší svobodu ve vyjadřování, avšak striktně v mezích daného tématu, na které se specializuje. Na rozdíl od nutnosti držet se při vyjadřování svého požadavku předem daných formulací by však uživatel mohl formulovat věty přirozeně a volně, bez nutnosti svůj hlasový vstup pozorně kontrolovat a přizpůsobovat. Aby byl systém schopen přirozeně formulované požadavky rozpoznat, vyžadoval by „výcvik“ s využitím objemného korpusu na dané téma, čehož jsou však dnešní neurální systémy schopny bez větších potíží.

### Vysoce interaktivní systémy s širokým jazykovým a tematickým pokrytím

Takový systém umožňuje uživateli velmi vysokou volnost ve vyjadřování i v přecházení mezi tématy. Aby však systém za těchto podmínek dosáhl dostatečně kvalitního výstupu je nutné, aby uživatel hlídal a, v případě potřeby, proces rozpoznání řeči a převodu i upravoval. Uživatelé tedy musí očekávat, že budou nuceni do výsledků rozpoznání řeči často zasahovat a ručně je upravovat, než tento hlasový vstup „pošlou“ do fáze překladu. Rovněž se musí vyhnout hezitačním zvukům, falešným začátkům a dalším prvkům přirozené řeči, jelikož by je systém nesprávně považoval za slova.

Zajímavé je, že Seligman a Zong u této kategorie předpokládají nutnost toho, aby si systém na řečníka a jeho individuální projev zvykl:

*„Thus it appears that each user will need to train an individual acoustic model, a process which would take several minutes.“ (Seligman, Zong, 2006)*

Pokud však porovnáme tuto tezi se současnými nejpůlárnějšími veřejně dostupnými systémy, které nepochybně mají ambice být právě tímto typem automatického tlumočení (Google Translate, Microsoft Translator), zjistíme, že nutností „naučit“ daný systém na konkrétního řečníka nedisponuje žádný z nich.

Budeme-li pak na využití automatického tlumočení nahlížet z pohledu potřeb uživatele definovaných Johnem Hutchinsem, které jsme již popsali výše, zjistíme, že nejčastější potřebou, kterou se uživatelé automatického tlumočení snaží uspokojit je **sdílení informací**, tedy dorozumění se s jinou osobou. To navíc koresponduje i se současnou nabídkou a s ambicemi technologií, které jsou v současné době dostupné jak zadarmo, tak za poplatek.

## 4.2 Omezení automatického tlumočení a strojového překladu

Omezením strojového překladu je věnována řada odborných publikací, kde se můžeme podrobně dočíst o omezeních jednotlivých druhů strojového překladu či najít hloubkové rozборы jednotlivých omezení z pohledu jazykového i softwarového. Z důvodu úspornosti se proto v této kapitole zaměříme pouze na omezení týkající se současných překladačů založených na statistických a neurálních metodách která mají vliv na použitelnost a kvalitu výstupu strojového překladu, přičemž kvalitou rozumíme srozumitelnost, věrnost, a styl.

Základním problémem strojového překladu je fakt, že počítače jednoduše nejsou (a dle některých ani nikdy nebudou) schopny vnímat a chápat drobné nuance a detaily, které lidský jazyk dokáže vyjadřovat. Přesněji řečeno, počítače se nedokáží adekvátně vypořádat s takovými prvky jako mnohoznačnost, idiomy, kolokace a další, a navíc nedokáží zachytit věci jako kulturní a sociální kontext či vzít v potaz vliv různých jazyků na lidské vnímání reality, které jsou pro kvalitní překlad nezbytné.

Výše zmíněné jevy můžeme pro zjednodušení souhrnně označit jako **sociálně-kulturní prvky**. Ty pak můžeme rozdělit na mnohoznačné výrazy, asymetrické výrazy, idiomy, a kolokace.

V případě **mnohoznačných a asymetrických výrazů** (kdy jednomu výrazu ve výchozím jazyce odpovídá víc výrazů v cílovém jazyce či naopak) chápe člověk správný význam z kontextu, který však nemusí být dán pouze v textové rovině – může se jednat o kontext situační, kontext vycházející z předchozí komunikace, či dokonce o kontext prezentovaný ve formě neverbální komunikace. Počítač však tyto možnosti nemá a pokud na podobnou situaci narazí, může pracovat pouze se samotným textem a svou databází a hledat tak nápovědu například v gramatické struktuře či ve slovosledu dané věty. Pokročilejší programy pak mohou využít maximálně kontext z okolního textu, například ze zbytku odstavce.

**Idiomy** můžou působit potíže i lidským překladatelům, pro počítač ale jejich problémovost spočívá v tom, že jejich význam nelze pochopit pouze z významu jednotlivých slov<sup>47</sup>. Kolokace pak představují problém proto, že se jedná o jednotky skládající se z více slov, které společně tvoří jednu lexikální jednotku, a význam jejich jednotlivých prvků se tak může lišit od slovníkového (srovnejme například „silný kuřák“ a „silný vzpěrač“). Dnes už však lze předpokládat, že díky strojovému učení a neurálním sítím dokáží překladače alespoň část idiomů a kolokací rozpoznat a, pokud byly trénovány s použitím kvalitních textů, je i dobře přeložit.

Další omezení strojového překladu se pak pojí s jeho **technickou stránkou**. Jak jsme již popsali v předchozích kapitolách, současné překladače fungují na statistické bázi a spoléhají tak na kvalitu korpusu, který byl použit při jejich tréninku. Některé online překladače, například Google Translate, dokonce využívají tzv. crowdsourcingu, tedy umožňují laickým uživatelům vyprodukované překlady hodnotit nebo je dokonce vytvářet. To představuje zřejmý problém – nekvalitní korpus logicky vede k méně kvalitním překladům a neprofesionální zásahy laiků do překladových pamětí či korpusu tedy mohou vést ke zhoršení jejich kvality a jedna chyba se tak může rychle rozšířit do všech textů. To pak vede na jedné straně k zábavným chybám ve výstupu, na straně druhé to ale může znamenat i nenápadné, závažné chyby, jak ilustruje ve svém článku „Can't I just use Google Translate?“<sup>48</sup> Richard Brooks, ředitel britské firmy poskytující jazykové služby K International. Nebezpečí překladu, který se na první pohled zdá být v pořádku, ukazuje na případu překladu potravinových obalů, kdy překladač narazil při překladu slova „nuts“ do francouzštiny na asymetrický výraz – „noix“ totiž může znamenat jak obecně ořechy, tak konkrétní druh, vlašské ořechy. Zatímco v obecném textu by taková chyba nejspíš nevadila, v případě potravinového obalu může představovat zdravotní riziko pro alergiky.

Ještě jednou oblastí, kde strojový překlad často naráží, jsou pak jména, názvy či zkratky. Jejich úspěšný překlad opět závisí na kvalitně použitého tréninkového korpusu, případně na pečlivé post-editaci.

Z podstaty fungování strojových překladačů a z výše uvedených problémů při překladu sociálně-kulturních prvků jazyka pak vyplývá, že pro překlad literárních textů jsou (zatím) nepoužitelné. Přestože v této oblasti probíhá řada pokusů<sup>49</sup>, i neurální překladače zatím dosahují při překladu známých literárních děl spíše slabších výsledků, přestože si pochopitelně vedou lépe, než starší metody<sup>50</sup>.

Na omezení týkající se výlučně automatického tlumočení můžeme nahlížet pouze optikou omezení ve fázi recepce – tedy omezení systémů pro rozpoznávání řeči – a ve fázi produkce, tedy omezení systémů pro syntézu řeči, nicméně vzhledem k tomu, že cílem této práce není

---

<sup>47</sup> GUERRA, Ana Fernández. Machine Translation. Capabilities and limitations. Universitat de València, 2000. ISBN 8437042666.

<sup>48</sup> Can't I Just Use Google Translate? Language Translation Services from K International [online]. Copyright © Copyright 2018 [cit. 09.04.2019]. Dostupné z: <https://www.k-international.com/blog/google-translate/>

<sup>49</sup> TORAL, Antonio a Andy WAY. What Level of Quality can Neural Machine Translation Attain on Literary Text? [online]. 15. ledna 2018 [cit. 2019-04-09]. Dostupné z: <https://arxiv.org/pdf/1801.04962.pdf>

<sup>50</sup> Machine Translates Literature and About 25% Was Flawless, Research Claims | Slator | Language Industry Intelligence [online]. Copyright © 2019 [cit. 09.04.2019]. Dostupné z: <https://slator.com/technology/machine-translates-literature-and-about-25-was-flawless-research-claims/>



zevrubný popis dílčích technologií zapojených do procesu automatického tlumočení, odkážeme zde čtenáře na odbornou literaturu věnovanou těmto technologiím (např. Springer Handbook of Speech Processing). Omezení věnovaná fázi převodu z jednoho přirozeného jazyka do druhého jsme pak popsali výše, v oddílu věnovanému omezením strojovému překladu – tato jsou totiž u obou druhů technologií, vzhledem k podstatě jejich fungování, totožná.

Krátce se však ještě pozastavíme u obecného pohledu na omezení automatického tlumočení. Hlavním problémem automatického tlumočení i překladu je fakt, že i u nejvyspělejších systémů, jako je STACL od Baidu, se stále často pracuje především se slovy, s jejich počtem, s tím, o kolik slov je výstup pozadu za originálem, zda má každé slovo svůj ekvivalent atd. Hlavním cílem jak v překladu, tak v tlumočení, je však předat smysl sdělení, nikoliv zachovat přesný počet slov ve výchozím i cílovém jazyce. Problematika smyslu je však velice těžce uchopitelná pomocí matematických vzorců či rigidních rámců, jelikož smysl výrazu se může měnit v závislosti na kontextu, řečnickových intencích, řečnickově vnímání světa, tématu, a dalších faktorech, a to i v rámci jedné věty či promluvy.

Výzkum v oblasti tzv. **označování smyslu** (sense tagging), tedy snahy tento jev zachytit pomocí exaktních metod, je stejně starý, jako samotný výzkum v oblasti strojového překladu a automatického tlumočení a mnozí badatelé se domnívali, že vyřešení tohoto úkolu přímo povede k vysoce kvalitnímu strojovému překladu:

*„It was assumed by many that if only wordsense ambiguity could be tamed, by the process we are calling sense-tagging, then MT of high quality would be relatively straightforward.“* (Wilks, 2009)

Do dnešního dne však nebyl tento problém uspokojivě vyřešen. Wilks uvádí mnoho důvodů, proč tomu tak je, mezi nimi například to, že přestože mnohoznačný výraz bude v rámci jednoho textu s velkou pravděpodobností vždy využit pouze s jedním smyslem, zdaleka to nezaručuje stejný smysl téhož slova i v jiných textech, i pokud jde o texty stejného typu nebo na stejné téma<sup>51</sup>.

Pokud pak budeme na omezení automatického tlumočení nahlížet z pohledu uživatele, Zong a Seligman ve své publikaci „Toward practical spoken language translation“ jako dva zásadní problémy pro uživatele automatického tlumočení uvádí nutnost spolupráce ze strany uživatele a nutnost přistupovat k určitým kompromisům. Mezi tyto řadí například:

- nutnost mluvit nahlas a srozumitelně,
- nutnost přizpůsobení systému danému mluvčímu,
- nezbytnost tichého prostředí,
- nutnost přijímat hrubé či neúplné překlady,
- nutnost ručně opravovat chyby, které systém učiní při rozpoznání hlasu,
- nutnost vyslovovat prvky jako závorky či dělení slov,
- hláskování či psaní slov, která má systém problém rozpoznat,

---

<sup>51</sup> WILKS, Yorick. Machine translation: its scope and limits. New York: Springer, 2008, s. 164. ISBN 978-0-387-72773-8.

- nutnost držet se při zadávání hlasového vstupu určitých konvencí,

a další<sup>52</sup>. Čím větší má pak systém ambice co do množství pokrytých témat a lingvistických jevů, tím více kompromisů a spolupráce takový systém od uživatele vyžaduje. Systém automatického tlumočení, který by od uživatele nevyžadoval žádné kompromisy, spolupráci, a ani jeho pozornost, pak považují za prakticky utopický:

*„Clearly, if effortless automatic operation cannot presently be achieved in text-only translation, its achievement in speech-to-speech translation is even less likely. In fact, even human interpreters rarely perform as transparently as the Star Trek gadget: they often interrupt dialogs to clarify the intentions of the speech partners.“* (Seligman, Zong, 2005)

*„Equipped with such a communicator, you could talk just as you habitually do when addressing fellow native speakers of your own language.“* (Seligman, Zong, 2005)

Seligman a Zong si tak všímají problému, jehož podstatu lze spatřovat v tom, že lidé, a tím pádem i jejich jazyk, jsou nedokonalí. Velice výstižně tento problém shrnula profesionální tlumočnice Barbora Tite:

*„Člověk mluví z cesty, plete si jména, mluví ironicky, neví, co chce říct, nesrozumitelně drmolí, má emoce, dělá nepřeložitelné vtipy a odbočuje do úplně jiného tématu. Někdy také odbočuje do úplně jiného jazyka. Chudák robottlumočnick, jak k tomu přijde?“* (Tite, 2018)<sup>53</sup>

Na základě výše uvedeného tedy lze za hlavní omezení automatického tlumočení, pokud odhlédneme od omezení způsobených technologiemi pro zpracování a syntézu jazyka, s určitou nadsázkou označit právě nutnost pracovat s lidským vstupem, a tedy i určitými lidskými vlastnostmi, které nelze jednoduše popsat pomocí matematických vzorců.

### 4.3 Hodnocení strojového překladu a automatického tlumočení

Vzhledem k principu fungování současně dostupných systémů strojového překladu a automatického tlumočení jsme se rozhodli věnovat jim společně i v této kapitole, neboť pokud se budeme zabývat pouze hodnocením výstupu a opomeneme u automatického tlumočení hodnocení dílčích systémů (tedy rozpoznávání a produkci řeči), což není primárním cílem této práce, budeme využívat stejné metody, jak ostatně dokazují i nejnovější odborné články představující novinky v oblasti automatického tlumočení – v drtivé většině z nich najdeme jednu z metod popsanych níže, nejčastěji pak metodu BLEU.

Na úvod je také nutné zmínit, že na kvalitu překladu lze nahlížet i z mnoha jiných pohledů než jen z pohledu kvality výsledného produktu, kterému se budeme v této práci věnovat. Pro

<sup>52</sup> ZONG, Chengqing a Mark SELIGMAN. Toward Practical Spoken Language Translation. Machine Translation [online]. 2005, 19(2), 113-137 [cit. 2019-04-09]. DOI: 10.1007/s10590-006-9000-z. ISSN 0922-6567. Dostupné z: <http://link.springer.com/10.1007/s10590-006-9000-z>

<sup>53</sup> TITE, Barbora. Jáchyme, hod' tlumočnicka do stroje [online]. 1.11.2018 [cit. 2019-04-09]. Dostupné z: <https://barboratite.cz/jachyme-hod-tlumocnika-stroje/>

úplnost však dodáváme, že kvalitu je možné posuzovat i z pohledu uživatele, procesu, účelu, či z pohledu souladu s normami.

Díky rozmachu strojového překladu se hodnocení jeho kvality postupně vyvinulo do samostatného a značně různorodého oboru v rámci překladu. Základem hodnocení je vždy kvalita výsledného překladu. Hlavním problémem je však absence obecně uznávaného objektivního měřítka oné kvality, jelikož v překladu často neexistuje jedno jediné „správné“ řešení, což má za následek existenci různých druhů kvalitativních a kvantitativních přístupů k hodnocení.

Mezi první příklady systematického hodnocení strojového překladu lze počítat zprávu ALPAC z roku 1966, kterou jsme již zmiňovali v kapitole pojednávající o historii strojového překladu, a která se kromě samotné kvality výstupu zajímala i o náklady či americký trh s překlady. V roce 1967 představili David B. Orr a Victor H. Small studii<sup>54</sup>, která hodnotila srozumitelnost strojem podporovaného překladu vědeckých dokumentů z ruštiny. Studie srovnávala srozumitelnost překladů vyprodukovaných pouze strojově, strojově s post-editací, a člověkem, a přestože strojový překlad v tomto srovnání skončil na posledním místě, konstatuje studie překvapivě vysokou kvalitu strojových výstupů a doporučuje další výzkum v této oblasti. Prvním skutečně komplexním hodnocením strojového překladu se stala evaluace Systranu, kterou provedla v 70. letech Evropská společenství a která pokrývala veškeré soudobé metody hodnocení strojového překladu.

V průběhu posledních desetiletí se hodnocení strojového překladu přesunulo z rukou lidí do rukou automatizovaných systémů, což umožňuje hodnotit velké objemy textu a velké množství jazykových párů při zachování příznivé finanční a časové náročnosti.

Při hodnocení strojového překladu je třeba mít na paměti, že kritéria a metodologie hodnocení musí být přizpůsobeny tomu, k jakému využití je daný systém strojového překladu vytvořen a co je cílem samotného hodnocení. Jak už jsme zmínili v kapitole 4.1, způsobů využití strojového překladu je více, a tyto pak vyžadují různou míru kvality výstupu. Použitelnost strojového překladu tak vždy záleží na konkrétním cíli, se kterým bude výstup využíván. Cíle hodnocení strojového překladu se rovněž liší na základě toho, kdo hodnocení provádí, a jaké vlastnosti systému jsou pro hodnotitele primární. Jinak proto bude k hodnocení přistupovat vědec, vývojář, laický uživatel nebo překladatel, jelikož každého z nich bude zajímat něco jiného, a každý z nich proto musí zvolit odpovídající typ hodnocení.

John S. White popisuje šest typů hodnocení strojového překladu<sup>55</sup>, které lze aplikovat i na automatické tlumočení:

- Test proveditelnosti (feasibility test) zjišťuje, zda je teorie či metoda realizovatelná a jaký má potenciál.
- Interní hodnocení (internal evaluation) zkoumá, zda součásti experimentálního systému fungují a pokud ne, proč.

---

<sup>54</sup> ORR, David a Victor SMALL. Comprehensibility of Machine-aided Translations of Russian Scientific Documents. *Mechanical Translation and Computational Linguistics* [online]. 1967(10) [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/MT-1967-Orr.pdf>

<sup>55</sup> WHITE, John. How to evaluate machine translation. In: SUMMERS, Harold. *Computers and Translation: A Translator's Guide*. John Benjamins Publishing, 2003, s. 211-244. ISBN 9027216401.

- Hodnocení použitelnosti (usability evaluation) zjišťuje užitečnost pro koncové uživatele, tedy zda jim systém pomůže dosáhnout jejich cíle.
- Hodnocení provozu (operation evaluation) analyzuje finanční náklady a výnosy systému.
- Hodnocení deklarativnosti (declarative evaluation) hodnotí schopnost systému vytvářet použitelné, publikovatelné texty.
- Srovnávací hodnocení (comparison evaluation) srovnává určitý aspekt napříč několika systémy.

Po určení cíle a typu hodnocení je pak možné nastavit odpovídající kritéria. Obecně jsou programy pro strojový překlad hodnoceny jednak z hlediska kvality výstupu, a jednak jako software, dle kritérií jako snadnost obsluhy, rychlost atd. V této kapitole se však z důvodu úspornosti a vzhledem k povaze práce budeme věnovat pouze hodnocení kvality výstupu strojového překladu a automatického tlumočení.

Při snaze určit kritéria pro hodnocení kvality výstupu strojového překladu narazili vědci hned na počátku na zásadní problém, totiž jak identifikovat „kvalitní překlad“ – neexistoval totiž žádný univerzální standard. Klíčovým je tak pojem „kvalita textu“, kterou lze charakterizovat z jednojazyčného a dvojjazyčného pohledu. Až na jednu věc však kladou oba tyto pohledy na text stejná kritéria, jako jsou gramatická správnost či koherence. Překlad však ještě navíc musí korespondovat se zamýšleným smyslem výchozího textu, a právě toto kritérium ekvivalence je středobodem mnoha teorií ohledně kvality překladu a předmětem debat.

V praxi je k hodnocení kvality výsledného překladu často používán systém založený na chybách, kdy jsou tyto odhaleny a poté posuzovány z hlediska jejich dopadu na překlad.

Kit Chunyu a Billy Wong Tak-ming ve svém příspěvku v odborné publikaci Routledge Encyclopedia of Translation Technology uvádí, že hodnocení výstupu strojového překladu je na pomezí čtení textu s účelem pochopení a s účelem evaluace, a podle účelu hodnocení se tak hodnotitel může více zaměřit na srozumitelnost výstupu, nebo naopak více analyzovat přítomné chyby. Setkávají se tak zde dva aspekty – na jedné straně srozumitelnost výstupu a na druhé straně přesnost zachování významu zdrojového textu. Tyto pak často nebývají v rovnováze a ve většině případů bývá strojový výstup dobrý z hlediska dodržování pravidel cílového jazyka, ale selhává co do zachování významu zdrojového textu. Dalším problémem je pak fakt, že neupravený výstup strojového překladu je zřídka kdy publikovatelný, s výjimkou silně specializovaných a na míru vytvořených systémů. Je proto těžké na výstup plný chyb ve slovosledu, volbě výrazů a jiných aspektech aplikovat přísnější kritéria hodnocení kvality, jako jsou například stylistické požadavky. Ačkoliv je tedy vrcholným cílem strojového překladu dosáhnout stejné (nebo i lepší) kvality, jako překlad lidský, je nutné používat k jeho hodnocení jiné metody a jiná kritéria.

Hodnocení strojového překladu může probíhat manuálně, kdy jsou hodnotiteli lidé, a automaticky, kdy se jedná o využití kvantitativních ukazatelů bez účasti člověka. V případě manuálního hodnocení tedy použitelnost a kvalitu hodnotí koncoví uživatelé, a přestože je tento typ hodnocení označován za subjektivní a nekonzistentní, jedná se stále o standard, kterému se automatické hodnocení nemůže rovnat. Toto lidské hodnocení pak lze rozdělit na dvě kategorie, kdy první hodnotí kvalitu jazyka (tzv. vnitřní evaluace) a druhá použitelnost

strojového překladu v určité situaci (tzv. vnější evaluace). Cílem automatického hodnocení je pak nedostatky manuálního hodnocení překonat. Jeho výhodou je také mnohem menší časová a finanční náročnost a jedná se tak o rychlý a důvěryhodný způsob, jak ocenit kvalitu výstupu strojového překladu.

Chunyu a Wong popisují následující druhy manuální a automatické evaluace, které dále popíšeme podrobněji:

- a) Manuální hodnocení – vnitřní evaluace
  - i. Hodnocení kvality
  - ii. Sestavení žebříčku překladových variant
  - iii. Analýza chyb
- b) Manuální hodnocení – vnější evaluace
  - i. Získání informací
  - ii. Test porozumění
  - iii. Post-editace
- c) Automatické hodnocení
  - i. Ukazatele podobnosti textu
    - a. BLEU a NIST
    - b. METEOR
    - c. TEC
    - d. ATEC
  - ii. Odhad kvality
  - iii. Meta hodnocení

- a) Manuální hodnocení – vnitřní evaluace
  - i. Hodnocení kvality

V tomto případě je úkolem hodnotitelů na základě vlastního úsudku ohodnotit, jak „dobrý“ překlad je. Dvěma základními kritérii jsou zde *věrnost* a *srozumitelnost*. John Hutchins uvádí ve své publikaci<sup>56</sup> ještě třetí kritérium, kterým je *styl*. V případě *věrnosti* je hodnoceno zachování sdělení původního textu. Hodnotitelé mohou být bilingvní, ale i monolingvní, pokud je pro srovnání k dispozici lidský překlad. *Srozumitelnost* pak hodnotí, nakolik snadné je překladu porozumět, bez ohledu na přesnost zachování původního sdělení a hodnotitelům tak postačí znalost jednoho jazyka. K hodnocení samotnému se nejčastěji používá pětibodová škála (5 – nejlepší, 1 – nejhorší, viz tabulka 2).

---

<sup>56</sup>HUTCHINS, John. Evaluation of MT systems. In: HUTCHINS, John a Harold SOMERS.: An introduction to machine translation. Academic Press, 1992, s. 161-174. ISBN 0-12-362830-X. příslušná kapitola dostupná z <http://www.hutchinsweb.me.uk/IntroMT-9.pdf>

body	Věrnost	Srozumitelnost
5	Kompletní	Bezchybné
4	Většinová	Dobré
3	Velká	Úroveň nerodilého mluvčího
2	Malá	Neplynulé
1	Žádná	Nesrozumitelné

Tabulka 2: pětibodová škála pro hodnocení věrnosti a srozumitelnosti. Převzato z Routledge Encyclopedia of Translation Technology, překlad Yulia Skadchenko.

## ii. Sestavení žebříčku překladových variant

Zde je úkolem hodnotitelů seřadit varianty překladu od nejlepší po nejhorší, případně vybrat z dvou navrhovaných variant tu, kterou preferují (nebo ani jednu, pokud v jejich kvalitě není rozdíl). Čím častěji jsou pak výstupy jednoho systému pro strojový překlad v průměru hodnoceny lépe než ostatní, tím lepší daný systém strojového překladu je. Tento způsob hodnocení je od roku 2008 považován za oficiální manuální metodu hodnocení statistického strojového překladu.

## iii. Analýza chyb

Tato metoda přistupuje k hodnocení z opačné strany a hodnotí „špatnost“ překladu tím, že se nejprve identifikují chyby v překladu, a na konci procesu se vytvoří odhad toho, jak pracné by bylo upravit výstup tak, aby odpovídal pro překlad přijatelnému standardu. Hodnotitel počítá každý případ, kdy je nutné přidat, odstranit, přemístit, nebo nahradit slovo, a z výsledku pak spočítá procento korekcí, tedy chyb, v celém textu. Kromě počítání chyb je však nezbytné je i klasifikovat, jelikož zatímco některé je možné odstranit poměrně snadno, jiné mohou mít složitější původ. Navzdory tomu, že i tento přístup ponechává prostor pro subjektivitu (různí hodnotitelé mohou mít různá kritéria pro to, co je a co není chybou, a různě je klasifikovat), je tato metoda objektivnější než hodnocení kvality a výsledky jsou navíc snáze interpretovatelné.

## b) Manuální hodnocení – vnější evaluace

### i. Získání informací

Tato metoda využívá jako hlavní ukazatel kvality strojového překladu to, do jaké míry dokáže uživatel získat z výstupu strojového překladu požadované informace. Tuto hodnotící metodu představili Chi-kiu Lo a Dekai Wu na 49. výroční schůzi Asociace počítačové lingvistiky pod zkratkou MEANT. Zakládá se na premise, že dobrý překlad je takový, ze kterého recipient pochopí základní strukturu sdělení: *kdo udělal co, komu, kdy, kde, a proč*. Metoda pracuje se zdrojovým textem, strojovým překladem, a lidským překladem. V překladech jsou pak označeny sémantické role a hodnotitelé porovnávají obsah sémantických rolí v obou překladech a hodnotí, zda jsou tyto ve strojovém překladu správné, částečně správné, nebo nesprávné.

### ii. Test porozumění

Tento způsob hodnocení funguje podobně jako například test porozumění při jazykových zkouškách. Hodnotitelé obdrží přeložené úryvky jednoho textu, kdy některé části jsou

přeloženy člověkem a některé pomocí strojového překladu, a na základě jejich přečtení pak odpovídají na otázky ohledně textu. To, jak si hodnotitelé v testu vedli, je pak odrazem kvality strojového překladu a toho, nakolik přesný a srozumitelný je jeho výstup. Součástí testu může být také hodnocení srozumitelnosti jednotlivých vět ze strany hodnotících.<sup>57</sup> Zajímavou odnoží testu porozumění je tzv. „cloze procedure“, metoda poprvé popsaná v roce 1953 W.L. Taylorem, kterou v roce 2000 v oblasti testování strojového překladu oživil H. Somers a E. Wildová<sup>58</sup>. Ti tuto metodu využili ke srovnání výstupů z několika různých strojových překladačů. Tento test probíhá tak, že jsou z textu odstraněna některá slova a úkolem hodnotitele je správně je doplnit na základě kontextu. Kvalita výstupu strojového překladu je pak hodnocena na základě toho, nakolik pomohl tento výstup hodnotiteli správně doplnit chybějící slova.

### iii. Post-editace

Při hodnocení pomocí post-editace je za ukazatel kvality strojového výstupu považováno úsilí vynaložené na uvedení výstupu na přijatelnou úroveň kvality – tedy čím menší úsilí je nezbytné vynaložit, tím lepší je výstup. Toto úsilí je pak měřeno buď jako čas vynaložený na post-editaci, nebo jako počet změn provedených při post-editaci (přidávání, odstraňování, nahrazování a přemísťování slov). Pokud je tedy potřeba vynaložit na post-editaci malé úsilí, může strojový výstup překladateli ušetřit čas a poskytnout mu jakýsi koncept překladu, v opačném případě se ale může stát, že post-editace zabere víc času, než kolik by překladatel potřeboval na vyhotovení překladu „od nuly“.

### c) Automatické hodnocení

#### i. Ukazatele podobnosti textu

V automatickém hodnocení je velice často využíváno lidských překladů, které slouží jako poměřovací ukazatel, se kterým je výstup strojového překladu porovnáván. Nejzákladnějším způsobem porovnání je srovnání slovní zásoby a slovosledu. Slabým místem této metody je fakt, že existence jediného správného překladu určitého textu je velice vzácná, metoda ale spoléhá na to, že různé verze překladu budou sdílet společné rysy jako slova či fráze.

#### a. BLEU a NIST

Metoda BLEU je považována za jednu z nejlivnějších v oblasti automatického hodnocení strojového překladu. Základní myšlenkou BLEU je, že „čím víc se strojový překlad blíží tomu lidskému, tím je lepší“ (Papineni et. al., 2001). Metoda funguje na základě počítání tzv. n-gramů, tedy sledu po sobě jdoucích položek o různé délce, které se vyskytnou jak ve strojovém výstupu, tak v alespoň jednom referenčním překladu. Gramatická správnost či srozumitelnost přitom nejsou brány v potaz. Metoda NIST je pak upravenou verzí metody BLEU, která bere v potaz i to, jak „informativní“ každý n-gram je. V současnosti jsou tyto metody považovány za oborový standard.

---

<sup>57</sup> PFAFFLIN, Sheila. Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments. Mechanical Translation [online]. 1965, 1965(8) [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/MT-1965-Pfafflin.pdf>

<sup>58</sup> SOMERS, Harold a Elizabeth WILD. Evaluating Machine Translation: the Cloze Procedure Revisited. In: Translating and the Computer 22: Proceedings of the Twenty-second international conference [online]. Londýn: Aslib, 2000 [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/00/Aslib-2000-Somers.pdf>

#### b. METEOR

I tato metoda vychází z BLEU, pracuje však s unigramy, tedy na základě propojení jednotlivých slov ve strojovém výstupu a referenčním překladu. Pro maximální šanci na shodu pak tato metoda pracuje také s kritérii jako kmeny slov, přesné pořadí písmen a synonyma.

#### c. TER

Metoda TER (Translation Edit Rate) je založena na měření počtu zásahů, které by člověk musel provést ve strojovém výstupu, aby tento přesně odpovídal referenčnímu textu, jinými slovy tedy hledá minimální počet zásahů nutný k transformaci jednoho prvku do jiného. Za zásahy považuje opět přidávání, odstraňování, nahrazování a přemísťování slov. Výsledkem je skóre v rozmezí 0–1, tedy od případu, kdy není nutná jediná změna až po situace, kdy je nutné změnit každé slovo.

#### d. ATEC

Metoda ATEC hodnotí kvalitu strojového výstupu na základě volby slov a slovosledu ve srovnání s referenčními překlady. Metoda tedy bere v potaz to, že každé slovo ve větě nese jiné množství informací.

##### ii. Odhad kvality

Na rozdíl od výše popsaných metod automatického hodnocení, při odhadu kvality se nepoužívají referenční překlady vytvořené lidmi – úkolem této metody je totiž „předpovědět“ kvalitu strojového výstupu. Tato metoda předpokládá existenci vztahu mezi vlastnostmi a znaky výchozího textu/jazyka a kvalitou strojového překladu, tedy například čím komplexnější je větná struktura zdroje, tím horší výsledek je možné očekávat. Tato metoda je založena na strojovém učení – prediktor je možné „vycvičit“ k tomu, aby rozpoznával vztah těchto znaků a kvality strojového výstupu.

##### iii. Meta hodnocení

Za meta-hodnocení je označován proces, během kterého jsou výsledky výše uvedených automatických způsobů hodnocení dodatečně posuzovány tak, aby bylo možné určit jejich spolehlivost a silné či slabé stránky. Toto posuzování je založeno na korelaci výsledků automatického hodnocení s výsledky manuálního hodnocení a využívá tedy speciálních korelačních koeficientů.

Důvěryhodnost a spolehlivost automatického hodnocení je však stále předmětem debat, jelikož velkou roli hraje typ hodnoceného textu. Vzhledem k tomu, že se automatické systémy spoléhají především na přiřazování slov k referenčnímu překladu, mohou poskytnout spolehlivé výsledky u doslovných překladů, ale naopak označit za katastrofu například volný literární překlad. Automatické metody hodnocení strojového překladu se tak používají především pro hodnocení samotných systémů – ke spolehlivé hloubkové analýze a ohodnocení kvality přeloženého textu je stále nezbytná evaluace člověkem.



## 5 Empirická část práce – praktický test programu Skype Translator

Součástí práce je praktický test s využitím jednoho ze zmiňovaných nástrojů automatického tlumočení, konkrétně softwaru Skype Translator. Test byl proveden ve formě hlasových hovorů s pěti dobrovolníky. Veškeré testovací rozhovory jsme rovněž nahrávali pomocí bezplatné verze programu FlashBack Pro 5. Nahrávky obsahují jak veškerý zvukový materiál, tak i videonahrávku obrazovky v průběhu experimentu, aby bylo možné vidět i transkripci řeči a převod Skypu z jednoho jazyka do druhého v písemné podobě. Nahrávky jsou dostupné prostřednictvím cloudové služby OneDrive a jsou chráněny heslem (to je možné získat na požádání od autorky práce). Veškeré materiály použité k hodnocení testu pak čtenář najde v příloze k této práci.

### 5.1 Cíl testu a metodologie

Cílem testu je především vyhodnotit schopnost tohoto softwaru zajistit úspěšnou komunikaci mezi dvěma lidmi, jež nehovoří stejným jazykem. Pro účely testu jsme využili (obousměrně) jazykový pár angličtina-ruština, jelikož oba tyto jazyky jsou v programu k dispozici v režimu „speech to speech“. Experimentu se zúčastnilo 5 dobrovolníků-laiků (tedy nepřekladatelů), kteří hovořili anglicky, a autorka experimentu, která hovořila rusky a pro účely experimentu reagovala pouze na výstup z programu a nebrala ohled na to, co skutečně vyřkli mluvčí angličtiny, přestože angličtinu ovládá.

Testovací dialogy byly vždy zhruba 25 minut dlouhé (přičemž program tlumočil oběma směry) a probíhaly podle předem daného obecného scénáře, který byl sestaven tak, aby alespoň přibližně pokryl úroveň jazyka dle Společného evropského referenčního rámce (CEFR). Před zahájením experimentu účastníci obdrželi krátké instrukce ohledně scénáře a programu, které čtenář najde v příloze k této práci. Celkem se jednalo o čtyři úrovně konverzace – základní seznámení a představení, popis cesty a zopakování získaných instrukcí, konverzaci na jednodušší, lineární téma, a konverzaci na náročnější téma.

Vzhledem k tomu, že testovaný program nepracuje pouze s textovým překladem a textovými soubory, ale také s rozpoznáváním hlasu a řečovou syntézou, byl výběr vhodných metod hodnocení poměrně omezený. Využití automatických metod, zmiňovaných v předchozí části práce, není kvůli povaze programu možné, a bylo proto nutné využít manuálních metod. Jako nejvhodnější se, vzhledem k cíli experimentu (tedy provedení testu použitelnosti) a také k našim možnostem jevíly metody vnitřní evaluace – přesněji řečeno hodnocení kvality a analýza chyb.

Hodnocení testu je rozděleno do dvou nezávislých částí. V první části je program hodnocen z hlediska procesu a v roli hodnotitelů jsou sami účastníci testu, kteří hodnotí prostřednictvím dotazníku (v angličtině) – vyplněné dotazníky čtenář nalezne v příloze k této práci. Z pohledu hodnocení účastníků se jedná o test použitelnosti neboli usability evaluation (John S. White, viz kapitola 4.3. této práce), tedy zda program dokáže pomoci uživatelům dosáhnout jejich cíle – úspěšné komunikace. Jako metodu hodnocení jsme zvolili jednu z metod vnitřní evaluace,

konkrétně hodnocení kvality, kde hlavním kritériem pro hodnocení byla srozumitelnost – na věrnost a styl nebyl brán ohled. Kromě toho účastníci experimentu rovněž hodnotili i technickou stránku softwaru a pohodlnost jeho používání. Ve zmiňovaném dotazníku hodnotili vybrané aspekty pomocí pětibodové škály a prostřednictvím doplňujících komentářů. Po ukončení hovoru také s každým účastníkem proběhla krátká neformální diskuze, ve které sdělovali své okamžité dojmy autorce práce.

Druhou částí hodnocení je pak hodnocení autorkou práce, která program hodnotí z pohledu výsledného produktu. K hodnocení byla opět zvolena jedna z metod vnitřní evaluace, a to analýza chyb, kterou provedeme se všemi nahrávkami. Pro potřeby této práce a s ohledem na to, že testujeme systém pro automatické tlumočení, jsme postup analýzy chyb v porovnání s tím, jak je tato metoda popsána v kapitole 4.3, lehce upravili. Chyby sledujeme v obou jazycích, tedy jak v anglickém, tak v ruském výstupu programu. Pro účely hodnocení za chybu považujeme:

- špatně rozpoznané slovo při poslechu,
- nevhodně zvolený překlad správně rozpoznávaného výrazu,
- záměnu tázacích vět za věty jiného druhu a naopak.

Zde je nutné poznamenat, že kategorií chyb by se v hodnocení dalo využít více – hodnotit by se daly rovněž například chyby v segmentaci či chyby ve skloňování nebo časování – v rámci časové úspory však volíme pouze tyto tři. V hodnocení správnosti rozpoznání budeme hodnotit správné rozpoznání jednotlivých slov a v případě, že program kvůli delší pauze v řeči rozdělí jednu větu na dvě či více částí, budeme hodnotit, zda jsou správně přeložena slova v těchto jednotlivých částech, bez ohledu na to, že se v originálu jednalo o jednu celou větu a její význam tedy mohl být ztracen. Záměnu tázacích vět za jiné jsme jako kritérium zvolili především kvůli nápadnosti této chyby. Dále autorka práce rovněž připojí své vlastní krátké „uživatelské“ hodnocení softwaru, bude se však více soustředit na výkon ruský „mluvící“ části programu. Především však, že subjektivní hodnocení autorky práce může být do jisté míry zkreslené, jelikož ovládá oba zmiňované jazyky a nelze se tak zcela vyhnout srovnávání.

Celkem se testu zúčastnilo pět dobrovolníků (tři muži a dvě ženy). Hlavním kritériem výběru dobrovolníků byla jejich nulová znalost ruštiny a angličtiny na úrovni rodilého mluvčího, čtyři z pěti dobrovolníků pak rodilými mluvčími skutečně jsou. Tito pochází z Nového Zélandu, Kanady, USA (stát Kalifornie) a Spojeného království. Nerodilá mluvčí angličtiny se již 10 let zabývá výukou angličtiny, je držitelkou certifikátu CPE a žije v bilingvní domácnosti. Britský mluvčí pak pochází z oblasti Manchesteru a v jeho projevu je tak přítomen výrazný místní přízvuk (tzv. Mancunian dialect), bylo proto zajímavé pozorovat, zda si s tímto přízvukem program poradí.

## 5.2 Hodnocení softwaru Skype Translator z pohledu účastníků testu

V této části uvedeme všechny otázky z dotazníku<sup>59</sup>, bodové hodnocení účastníků a průměrný výsledek u každé položky. Rovněž zde zmíníme případné zajímavé komentáře účastníků či shrneme jejich dojmy na základě komentářů a neformálního ústního hodnocení po experimentu.

### **Otázka 1:**

Jak snadné bylo spuštění a nastavení tlumočeného Skype hovoru? (How easy was the setup of Skype's interpreted call?)

1: velmi složité (extremely difficult)

2: složité (difficult)

3: ani snadné, ani složité (neither easy nor difficult)

4: snadné (easy)

5: velmi snadné (extremely easy)

V této otázce měli uživatelé ohodnotit, jak snadné bylo tlumočený hovor spustit a nastavit. Zde je nutné poznamenat, že spuštění vždy prováděl autor experimentu, účastníci si však museli v novém překládaném (tlumočeném) chatu zvolit svůj jazyk a také hlas (mužský či ženský). Teprve poté bylo možné zahájit hlasový rozhovor.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
5	4	3	4	4	<b>4</b>

V psaných komentářích k tomuto aspektu někteří účastníci zmiňují zbytečně zahlcující a přehnané množství textu s počátečními instrukcemi, ve kterém je složité se zorientovat, a místo kterého by stačilo jednoduché rozhraní nabízející volbu svého jazyka a hlasu (mužský či ženský). Dále se ale všichni shodují, že po počátečních zmatcích se nejedná o nic složitého, přestože by toto prvotní nastavení mohlo být v rozhraní provedeno v jednodušší podobě.

### **Otázka 2:**

Jak hodnotíte rozvržení uživatelského rozhraní? (How well was the user interface arranged?)

1: velmi špatně (extremely poorly)

2: špatně (poorly)

3: ani dobře, ani špatně (neither well nor poorly)

---

<sup>59</sup> Celé, účastníky vyplněné dotazníky jsou součástí přílohy k této práci.

4: dobře (well)

5: velmi dobře (extremely well)

Cílem této otázky bylo zjistit, jak dobře se se uživatelé bez předchozích zkušeností s tímto programem dokáží zorientovat v uživatelském rozhraní a jeho ovládacích prvcích.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
4	2	4	4	3	<b>3,4</b>

Hlavním důvodem nižšího hodnocení byl v tomto bodě fakt, že transkripce překladu a programem rozpoznané řeči se objevovala na poměrně krátkou dobu, což působilo potíže především když mluvčí hovořili v delších úsecích a nestíhali tak číst, zda program jejich řeč rozpoznal správně, aby problémový úsek mohli případně zopakovat. Uživatelům rovněž vadilo, že se tento text nedal zpětně zobrazit v chatu ani nalézt v jeho historii.

### **Otázka 3:**

Jaká byla kvalita zvuku? (*What was the sound quality like?*)

1: velmi špatná (extremely poor)

2: špatná (poor)

3: ani dobrá, ani špatná (neither good nor poor)

4: dobrá (good)

5: velmi dobrá (extremely good)

V této otázce se kvalitou zvuku myslí technické aspekty jako ozvěna, zpoždování zvuku, výpadky atd.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
3	4	5	4	5	<b>4,2</b>

S tímto bodem neměli účastníci problémy, pouze jeden člověk zmínil ozvěnu, která ale byla pravděpodobně způsobena kvalitou jeho připojení, nikoliv programem – autorka práce v rozhovoru s tímto účastníkem žádné potíže nezaznamenala.

### **Otázka 4:**

Jak dobrá byla umělá řeč? (*How good was the synthesized speech?*)

1: velmi špatná (extremely poor)

2: špatná (poor)

3: ani dobrá, ani špatná (neither good nor poor)

4: dobrá (good)

5: velmi dobrá (extremely good)

Touto otázkou jsme zjišťovali, jak působil umělý hlas na posluchače z hlediska vlastností jako přirozenost, monotónnost, tón hlasu, emoce, výslovnost atd.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
4	5	4	3	4	<b>4</b>

Zde se účastníci experimentu shodují v tom, že přestože je syntetický hlas „robotický“, „monotónní“ a „bez emocí, kontextu, či tónu“, neubírají tyto „kosmetické“ nedostatky na srozumitelnosti konverzace, která je zachována díky dobré výslovnosti. Účastníci v komentářích také pozitivně hodnotí to, že vždy slyšeli i původní vstup v cizím jazyce, což jim umožňovalo získat z něj právě informace o emocích a intencích jejich konverzačního partnera, které program nedokázal předat dostatečně věrně či správně.

### **Otázka 5:**

Jak často bylo nutné opakovat vstup, aby Skype správně rozpoznal to, co říkáte? (How often did you have to repeat the input for Skype to recognize what you were saying correctly?)

1: velmi často (extremely often)

2: často (often)

3: občas (sometimes)

4: téměř vůbec (hardly ever)

5: ani jednou (not even once)

V této otázce nás zajímal subjektivní pocit uživatelů z toho, jak často se museli opakovat, aby program správně rozpoznal to, co říkají – počítat tyto případy však nebylo jejich úkolem.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
3	3	3	3	3	<b>3</b>

V této otázce získal program v průměru nejhorší hodnocení. Jako největší problém účastníci zmiňují nepříliš dobrou schopnost programu rozpoznat jména, názvy, či výrazy z cizích jazyků (např. „Red Square“, „Old Town Square“, „Peter Jackson“, „Prague“). V některých případech ale program nedokázal ani po několika pokusech správně rozpoznat běžná slova jako „zoo“ nebo „cat“. Nejhuře si program vedl, dle našeho očekávání, v případě řečníka s výrazným

Manchesterským přízvukem, kdy dokonce několikrát místo běžných výrazů rozpoznal prostá slova. Účastníci rovněž komentovali fakt, že tyto „přeslechy“ programu většinou vedly

k humorným situacím, někdy ale také způsobily zcela nesrozumitelný překlad a bylo tak nutné vstup opakovat znovu, pomaleji a s větší artikulací.

### **Otázka 6:**

Jak často jste museli pokládat doplňující otázky, aby se vám podařilo porozumět tomu, co chtěl váš partner říci? (How often did you have to ask additional question in order to understand what your partner meant?)

1: velmi často (extremely often)

2: často (often)

3: občas (sometimes)

4: téměř vůbec (hardly ever)

5: ani jednou (not even once)

Touto otázkou jsme zjišťovali subjektivní pocit uživatelů z toho, jak často se museli doptávat, aby správně pochopili sdělení svého protějšku (například v případě nesrozumitelného překladu). Počítat tyto výskyty však opět neměli.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
4	3	4	4	4	<b>3,8</b>

I přes poměrně kritické hodnocení v předchozí otázce se účastníci v tomto bodě převážně shodli na tom, že i přes nepřesnosti a chyby v překladu byli schopni bez větších obtíží pochopit smysl sdělení a vést konverzaci a doptávat se museli především v situacích, kdy se řečníci a software začali „překřikovat“ (pokud řečník mluvil déle, začal software po určité době překládat, přestože řečník pokračoval v řeči, a mluvili tak přes sebe) či když oba řečníci začali hovořit najednou.

### **Otázka 7:**

Jak dobře jste dokázali následovat instrukce, které vám Skype podával? (How well were you able to follow instructions based on Skype's output?)

1: velmi špatně (extremely poorly)

2: špatně (poorly)

3: ani dobře, ani špatně (neither well nor poorly)

4: dobře (well)

5: velmi dobře (extremely well)

Tato otázka se týkala části experimentu, ve které si účastníci a autor práce navzájem popisovali cestu z bodu A do bodu B a tuto poté měli zopakovat. Uživatelé tedy měli představit svůj subjektivní pocit z toho, zda by se jim na základě těchto instrukcí podařilo najít správnou cestu do cíle.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
5	4	5	3	4	<b>4,2</b>

Většina účastníků ponechala tuto otázku bez komentáře, pocity komentujících účastníků se pak zajímavě rozcházejí. Zatímco jeden zmiňuje jednoduchost pochopení instrukcí a považuje tuto část za jeden z nejlepších výkonů programu, druhý naopak zmiňuje potíže související s rychle mizejícím textem a nutnost se v případě delších popisů opakovat. Je tak zřejmé, že úspěch této části značně závisí na míře spolupráce ze strany řečníka – program si totiž mnohem lépe vedl v případech, kdy řečníci podávali instrukce po kratších segmentech a větách, než když byly instrukce prezentovány ve formě dlouhých souvětí a promluv.

### **Otázka 8:**

Ohodnoňte, prosím, časovou prodlevu mezi tím, než Skype přijal mluvený vstup a začátkem hlasového překladu. (Please rate how you felt about the amount of time between Skype's speech comprehension and speech production.)

1: velmi dlouhá (extremely long)

2: dlouhá (long)

3: ani krátká, ani dlouhá (neither long nor short)

4: krátká (short)

5: velmi krátká (extremely short)

V této otázce jsme zjišťovali subjektivní pocit uživatelů z prodlevy mezi fází poslechu a fází produkce programu Skype Translator a zda si této prodlevy všimli, či zda je nerušila.

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
4	4	3	3	4	<b>3,6</b>

V komentářích k této otázce účastníci spíše než samotnou délku prodlevy, která je většinou nijak nerušila, zmiňovali to, že si nebyli jisti, kdy začne program mluvit a kdy pak mohou znovu začít mluvit oni. Je také nutné podotknout, že čtyři účastníci používali tento program poprvé a jeden z účastníků měl pouze jednu předchozí zkušenost. Většina však vyjádřila přesvědčení, že by si na „spolupráci“ s programem při delším či pravidelném používání zvykli tak, aby vše probíhalo plynule.

### **Otázka 9:**

Jak byste ohodnotili celkovou srozumitelnost výstupu programu Skype Translator? (Overall, how comprehensible did you find Skype's output?)

- 1: velmi špatná (extremely poor)
- 2: špatná (poor)
- 3: ani dobrá, ani špatná (neither good nor poor)
- 4: dobrá (good)
- 5: velmi dobrá (extremely good)

V této otázce měli účastníci experimentu ohodnotit celkovou srozumitelnost výstupu programu (tedy jak faktické informace, tak způsob jejich podání).

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
5	4	4	4	5	<b>4,4</b>

Tuto otázku účastníci téměř nekomentovali, v následné diskuzi se však shodovali, že i přes nedostatky rozuměli výstupu dobře a bez větších obtíží.

### **Otázka 10:**

Jakou celkovou známku si podle vás Skype zaslouží, na základě výše zmíněných faktorů? (Considering all the factors mentioned above, what is the overall grade you think Skype deserves?)

- 1: naprostá katastrofa (total disaster)
- 2: nepříliš dobrý výkon (not very good performance)
- 3: průměrný výkon (average performance)
- 4: dobrý výkon (good performance)
- 5: vynikající výkon (extremely good performance)

Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5	Průměr
4	4	4	3-4 <sup>60</sup>	4	<b>3,9</b>

V závěrečných komentářích se většina účastníků shodla, že je program, navzdory jejich nízkým očekáváním, pozitivně překvapil – někteří do té doby ani netušili, že Skype tuto funkci nabízí a pouze jeden z nich měl s programem jednu předchozí zkušenost. Uživatelé si také všimli, že pro úspěšný převod je důležitá určitá spolupráce z jejich strany, především dobrá artikulace a

---

<sup>60</sup> V průměru jsme toho hodnocení počítali jako 3,5 bodu



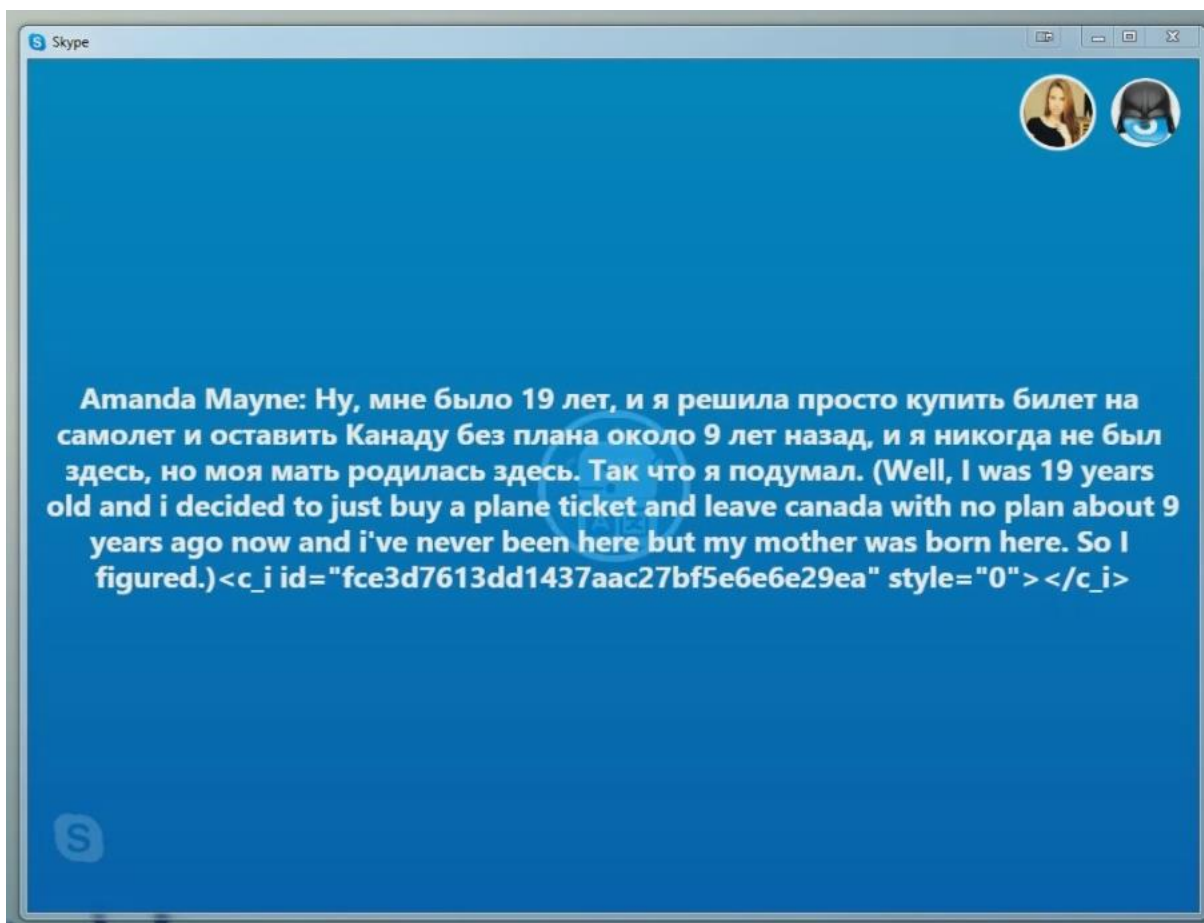
přizpůsobení rychlosti řeči, tedy aspekty, které mají velký vliv i na kvalitu tlumočení u lidských tlumočnicků.

V následné neformální diskuzi pak účastníci komentovali také nepříliš dobrý výkon programu v oblasti rozpoznání dotazů či správné segmentace – když se například řečník v projevu na chvíli zamyslel, program si toto často vyložil jako konec věty či promluvy a začal překládat, přestože řečník ještě nedokončil myšlenku, což vedlo ke zmatkům. Účastníci si také všímali potíží, které programu činily některé prvky přirozené řeči jako hezitační zvuky, zmíněné delší „přemýšleč“ pauzy, falešné začátky, opakování, parazitní výrazy jako „well“, „uhm“, „actually“, atd.

Pokud si spočítáme celkový bodový průměr ze všech deseti otázek (sečtením průměrů u každé z nich), získáme celkový výsledek **3,85 bodu**, po vynechání poslední otázky (celkové hodnocení) se dostaneme na průměr 3,84 bodu. Obě tato čísla jsou tedy velice blízko průměru celkového hodnocení programu účastníky experimentu v poslední otázce (**3,9 bodu**), což znamená průměrný až dobrý výkon. V celém dotazníku bylo špatné hodnocení uděleno pouze jednou, a to v oblasti uživatelského rozhraní, které nemá přímý vliv na hlavní funkci programu, tedy zajištění komunikace. Nejhorší průměrné hodnocení získal program za schopnost správně rozpoznávat řeč (3 body), zároveň však získal program nejvyšší počet bodů právě za celkovou srozumitelnost (4,4 bodu). Z toho lze usuzovat, že i přes jisté nedostatky na straně rozpoznávání řeči dokáže program svou funkci, tedy zajištění komunikace a dorozumění se, plnit dobře. Ze subjektivního hodnocení účastníků, jejich komentářů i bodového hodnocení jasně vyplývá, že program testem použitelnosti prošel a pomohl uživatelům dosáhnout jejich cíle.

### 5.3 Hodnocení programu autorkou práce

Je otázkou, zda by stejně dobře byla hodnocena i ruský „mluvící“ část programu. Na základě naší vlastní účasti v experimentu jsme získali dojem, že tato část podala výkon o něco horší, a to především v oblasti syntetizované řeči. Zatímco umělá řeč v angličtině zněla skutečně poměrně přirozeně a byla v ní přítomna živá intonace, umělá ruština byla o poznání více robotická a monotónní, méně plynulá a několikrát se program dopustil i nesprávné výslovnosti. Zajímavou chybou bylo také to, že program při převodu do ruštiny nesprávně určoval pohlaví anglického mluvčího. Ne všichni mluvčí za sebe nechali mluvit robotický hlas stejného pohlaví, jako jsou oni sami, program však nezávisle na této volbě u mužských mluvčích nejednou předával jejich sdělení v ženském rodě a naopak, u žen přepínal do rodu mužského. Rod řečníka navíc program náhodně měnil i v rámci jedné promluvy či věty, nezávisle na pohlaví umělého hlasu. Je tedy možné usuzovat, že výběr hlasu, který reprezentuje mluvčího, nemá u jazyků, které rozlišují rodové koncovky, na jejich jednotné užití vliv, což by mohlo být potenciálním zdrojem problémů.



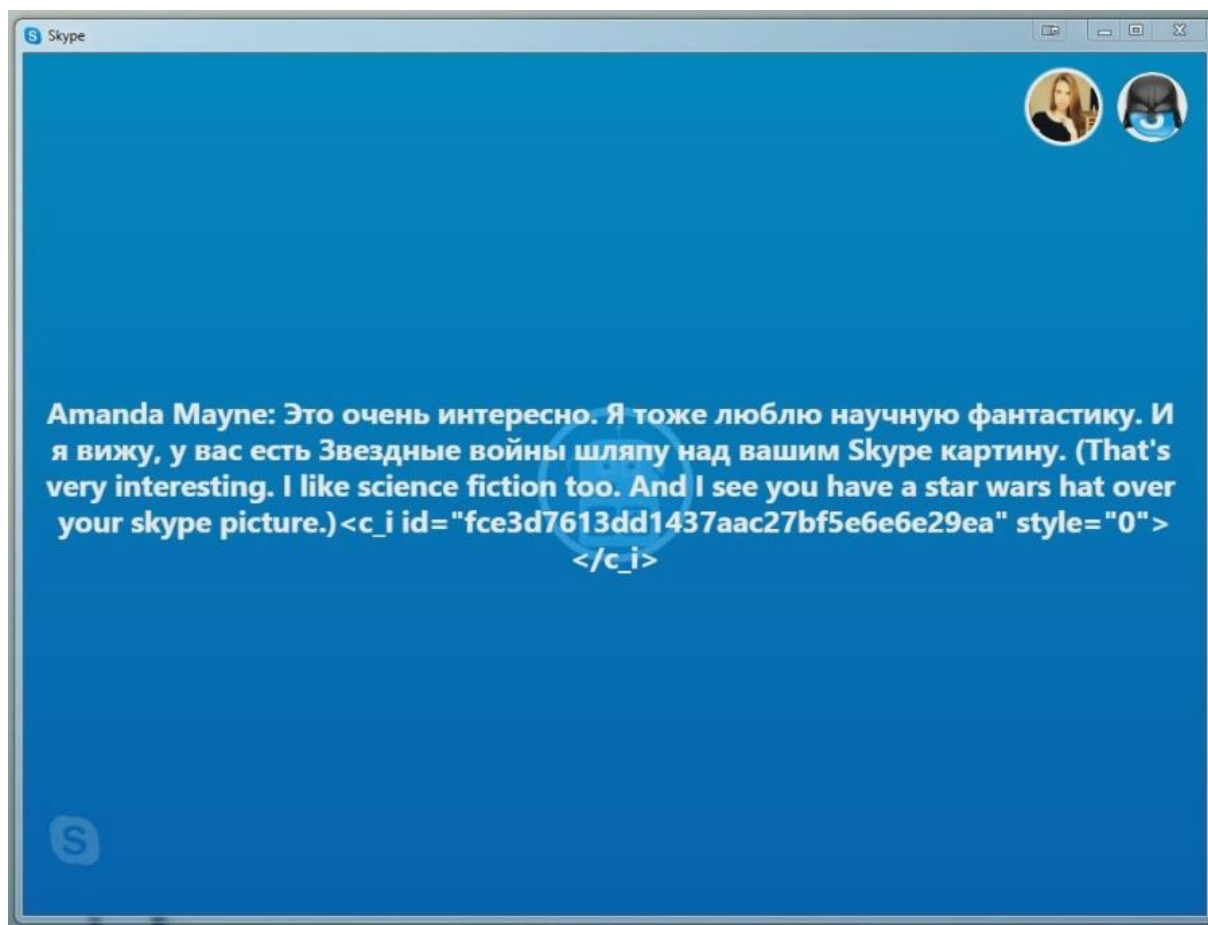
Obrázek 11: Screenshot z jednoho z rozhovorů ilustrující přeskakování mezi mužským a ženským rodem při převodu z angličtiny do ruštiny. Chybu v rodu je zřetelná na rozdílu «решила», «подумал» („rozhodla jsem se“, „řekl jsem si“).

Za povšimnutí stojí rovněž výkon programu v poslední části experimentu, kde měli uživatelé za úkol mluvit co nejpřirozeněji, neomezovat se ve slovní zásobě ani délce vět či promluv a vyjadřovat komplexní, nelineární myšlenky. Naproti našim očekáváním zde totiž program, při správném rozpoznání řeči, poměrně často podával až překvapivě dobré výkony, a to nejen co do srozumitelnosti myšlenek, ale i do formy a stylu výstupu. Překvapivě měl pak program spíše problémy s krátkými, velice jednoduchými větami, například snadnou úvodní otázkou „Привет, меня слышно?“<sup>61</sup> překládal jako „Hi, I hear you.“ Dobrý výkon v poslední části je možné částečně odůvodnit tím, že uživatelé měli v průběhu rozhovoru čas si na spolupráci s programem zvyknout a vědomě či podvědomě mu tak svou řeč do určité míry mohli přizpůsobovat. Spolehlivé informace o tom, zda je Skype Translator vybaven schopností přizpůsobení se mluvčímu (tak, jak to dělal například Verbmobil) se nám bohužel dohledat nepodařilo.

Velké nedostatky má program ve skloňování a časování, jelikož ruština, jako flektivní jazyk, je především v případě skloňování výrazně komplexnější než angličtina. I v situacích, kdy Skype dokonale rozpoznal všechna řečnickova slova a zvolené výrazy byly v cílovém jazyce zcela správné přesto docházelo ke špatnému skloňování a časování – obě chyby byly spíše pravidlem

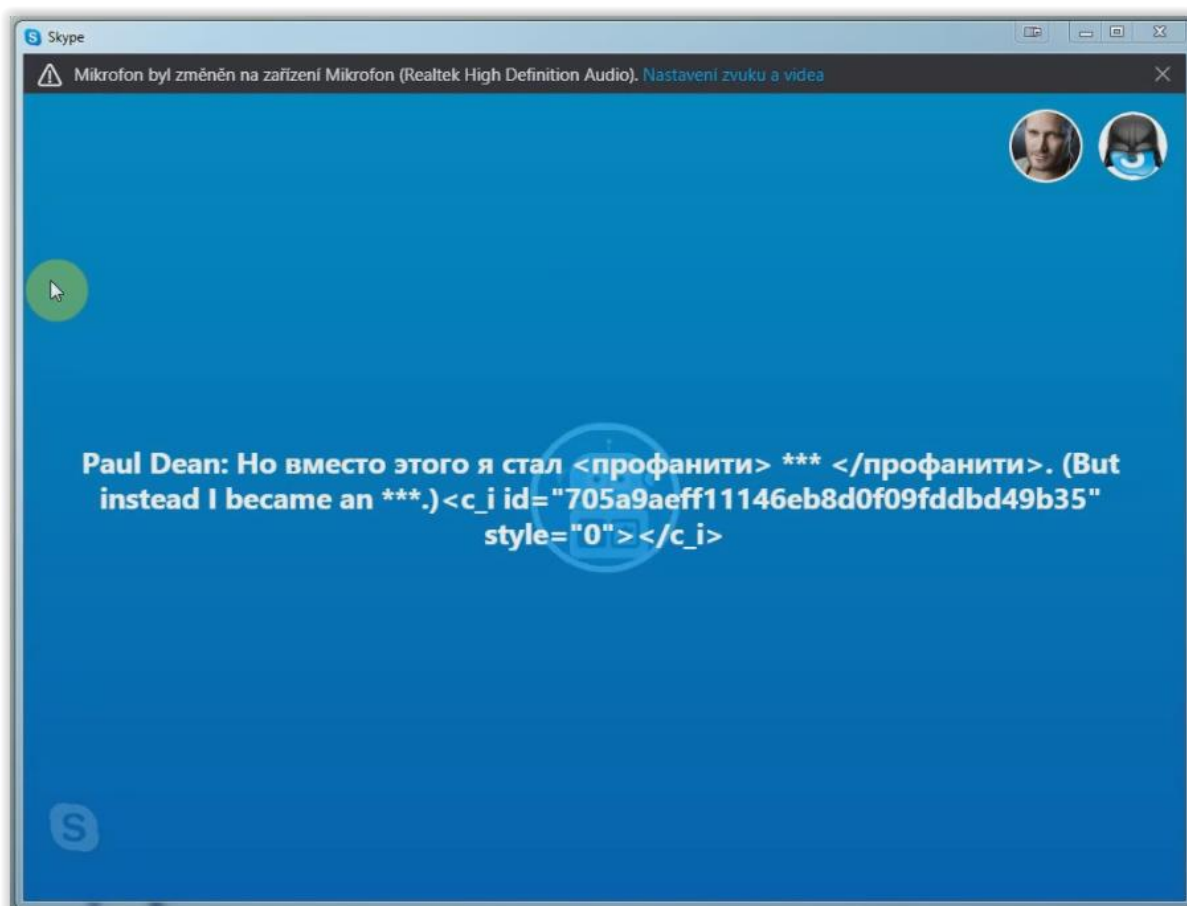
<sup>61</sup> Ahoj, je mě slyšet?

než výjimkou. V případě předkládaného experimentu toto nemělo na srozumitelnost příliš velký vliv, jde ale o potenciální zdroj problémů, především při chybách v časování.



Obrázek 12: Screenshot ilustrující problémy se skloňováním, viz «У вас есть Звездные войны шляпу над вашим Skype картину» („máte Hvězdné války čepice nad vaším Skype obrázku“).

Při vyhodnocování dotazníku jsme již zmínili, že program někdy chybně rozpoznával v řeči sprostá slova. Zajímavý je však také způsob, jakým Skype s těmito výrazy zachází, ať už je rozpozná správně či ne. Při rozpoznání nevhodného výrazu se spustí filtr nevhodného obsahu a program v transkripci nahradí takový výraz třemi hvězdičkami, v překladu se však zobrazí text připomínající počítačový kód (viz obrázek 13), a umělý hlas navíc celý tento kousek kódu doslova přečte. V ruštině je tak inkriminovaná věta narušena a uživatel, místo například v médiích běžně užívaného „vyprávění“, slyší: „меньше чем профанити больше чем меньше чем слеш профанити больше чем“, tedy „méně než profanity více než méně než lomeno profanity více než“. To značně naruší plynulost konverzace a vytrhne z ní posluchače, který si tak silně uvědomí přítomnost překladače. Vypnout tento filtr dospělého obsahu navíc není možné, jelikož společnost Microsoft 1. května 2018 upravila své smluvní podmínky a přidala do nich mimo jiné plošný zákaz užívání sprostých slov napříč všemi svými službami – filtr je tak přítomen ve všech službách společnosti.



Obrázek 13: Screenshot zobrazující práci programu se sprostými výrazy.

## 5.4 Analýza chyb

Manuální analýzu chyb jsme se rozhodli provést abychom mohli její výsledky porovnat s hodnocením účastníků experimentu a zjistit, zda bude jejich subjektivní hodnocení odpovídat faktické analýze, přesněji řečeno, zda aspekt, který uživatelé ohodnotili v průměru nejméně body (tedy schopnost programu rozpoznávat řeč, kterou hodnotili v otázce číslo 5), bude ze sledovaných skutečně tím nejvíce problematickým. Rovněž nás zajímala korelace počtu chyb a celkového průměrného hodnocení, které každý účastník programu udělil.

Pro účely analýzy za chybu považujeme:

- Špatné rozpoznání vstupu, tedy nesprávné či vynechané slovo (v obou jazycích)
- Špatně zvolený překlad dobře rozpoznávaného vstupu (v obou jazycích)
- Záměna tázacích vět za jiný typ věty a naopak (v obou jazycích)

V tabulce níže uvádíme přehled chyb v rozhovorech s jednotlivými účastníky, celkový počet chyb a průměr hodnocení každého účastníka. Jeden vzorek jsme bohužel byli nuceni z této části experimentu vyřadit kvůli technické závadě – došlo k poškození souboru s nahrávkou a analýzu chyb tak nebylo možné provést.

Druh chyby / účastník	Účastník 1	Účastník 2	Účastník 3	Účastník 4	Účastník 5
Chybné rozpoznání	55	vyřazen	35	79	77
Chybný překlad	27	vyřazen	26	23	24
Tázací věty	20	vyřazen	11	7	12
Celkem chyb	102	vyřazen	72	109	113
Průměrné hodnocení	4,1	3,7	3,9	3,55	4

Tabulka 3: výsledky manuální analýzy chyb.

Navzdory našemu očekávání nebyla prokázána korelace mezi celkovým počtem chyb a průměrným hodnocením – účastník s největším počtem chyb naopak dokonce udělil programu druhé nejlepší průměrné hodnocení. Důvodem mohla být zvýšená míra tolerance a jakási shovívavost či přívětivá očekávání účastníků vůči technologii, či naopak to, že výkon programu jejich očekávání předčil. Je také třeba připomenout, že se jednalo o subjektivní hodnocení každého z účastníků a každý tak mohl být ve svém hodnocení jinak přísný. Z tabulky je však zjevná shoda v otázce, která z hodnocených schopností programu podala nejhorší výkon. V dotazníku totiž účastníci experimentu za nejhorší označili schopnost programu rozpoznávat řeč a z naší analýzy vyplývá, že největšího počtu chyb se program dopouštěl právě na této rovině – téměř ve všech případech je počet chyb na programem rozpoznávaném vstupu minimálně dvojnásobný oproti počtu případů chybného překladu.

Nejvíce chyb v rozpoznání řeči dělal program, dle našeho očekávání, v případě britského účastníka, kdy si program často nedokázal poradit s jeho přízvukem a některá slova tak nerozpoznal vůbec či až po několika pokusech, kdy řečník musel artikulovat až přehnaně („cat“, „Prague“, „aeroexpress“). Nejlépe si pak program vedl v případě účastníka z Nového Zélandu – toto lze anekdoticky přisuzovat povolání účastníka, který je divadelním hercem a dobrý projev a především artikulace jsou pro něj nezbytné.

Výskyt chyb v překladu byl napříč všemi experimentálními rozhovory konzistentní. Tyto chyby nejčastěji spočívaly v tom, že program při převodu mnohoznačných výrazů zvolil nesprávnou variantu, např. „get“ v otázkách jako „How do I get to the station“ překládal jako „получить“, tedy „získat“, nikoliv „попасть“ – „dostat se“, slovo „square“ pak program i v případech, kdy bylo z kontextu zcela zřejmé, že se jedná o význam „náměstí“ (Red Square, Old Town square) vytrvale převáděl jako „квadrat“, tedy „čtverec“.

Nejméně často se z pozorovaných chyb vyskytovalo chybné rozpoznávání otázek, kdy program často tázací větu převedl do druhého jazyka jako větu oznamovací a v několika případech naopak z oznamovací věty udělal větu tázací. Nicméně díky tomu, že je neustále možné slyšet originální vstup konverzačního partnera toto nepředstavovalo příliš velký problém, jelikož každý z účastníků takto tlumočeného dialogu má možnost sledovat intenci řečníka právě z jeho intonace, případně pochopit, že se jedná o dotaz, z kontextu celé konverzace.

## 5.5 Závěrečné shrnutí hodnocení

Jak jsme již zmiňovali dříve, cílem testu bylo provést test použitelnosti programu Skype Translator. Na základě hodnocení účastníků testu pak můžeme konstatovat, že měl program

v tomto testu poměrně dobré výsledky a přes všechny přítomné chyby, nedostatky a problémy především v oblasti správného rozpoznání řeči dokázal úspěšně zprostředkovat komunikaci. Uživatelé si ve svém hodnocení zcela správně všimli, že pro maximální úspěch při pochopení jejich řeči je nezbytné se programu do určité míry přizpůsobit a rovněž věnovat pozornost své artikulaci a rychlosti řeči, obzvláště u výrazných dialektů. Zajímavé je, že délka promluv neměla na schopnost programu správně rozpoznávat řeč příliš velký vliv – ten byl patrný pouze v sekci experimentu, ve které bylo nutné popsat partnerovi cestu z jednoho místa na druhé, kde si program lépe vedl, pokud byly tyto instrukce podávány v kratších segmentech. V závěrečné části program naopak dokázal velice dobře zpracovat i delší promluvy, ve kterých řečníci vyjadřovali své subjektivní názory a nijak se ve svém projevu neomezovali. Dobrého výkonu programu v této části dialogu si účastníci testu také všimli a v ústním hodnocení toto hodnotili velice pozitivně. Navzdory očekávání nebyla prokázána korelace mezi počtem chyb, kterých se program dopustil v rozhovorech s jednotlivými účastníky, a průměrným bodováním, které mu udělili. Zatímco náš předpoklad byl takový, že čím více chyb, tím horší bude hodnocení daného účastníka a naopak, čím méně chyb, tím lepší bude účastníkově celkové hodnocení, celkové vyhodnocení výsledků testu tuto hypotézu vyvrátilo.

## 6 Omezení výzkumu a výhledy do budoucna

Jak jsme již zmínili v předchozích kapitolách i jak ukázal náš experiment, technologie pro automatické tlumočení (i strojový překlad) mají na jednu stranu řadu omezení, na druhou stranu však v určitých situacích komunikaci či získání potřebných informací zajistit dokážou (s různou mírou kvality). Co do samotného převodu, tedy části, kterou mají strojový překlad i drtivá většina systémů automatického tlumočení společnou, najdeme tato omezení především na poli mnohoznačnosti, idiomatických obrátů, ale i v oblasti korpusů – kvalita převodu je totiž přímo závislá na kvalitě korpusu, se kterým systém pracuje. Cestu ke zlepšení je tedy možné hledat například v zaměření se právě na to, jak kvalitní korpusy jsou a kdo je vytváří (např. prostřednictvím profesionálně sestavených korpusů, bez zásahů laiků, tedy omezením tzv. crowd-sourcingu). Potenciálně zajímavou by mohla být i možnost zvolit si v rozhraní strojového překladače konkrétní oborové korpusy a jiné naopak vyloučit, aby se snížila pravděpodobnost zvolení nesprávné překladové varianty u mnohoznačných výrazů a nedocházelo ke „kontaminaci“ výrazů mezi obory – srovnajme například výraz „mine“, který v oboru těžby znamená důl, ve vojenství minu, a může mít i význam přivlastňovacího zájmena. Jednalo by se tedy o jakousi obdobu překladových pamětí, které překladatelé využívají v nástrojích CAT.

V oblasti automatického tlumočení náš test, prováděný obousměrně v páru ruština-angličtina, ukázal, že největší potíže mají tyto programy s rozpoznáním přirozeného jazyka, ale také s ruchem z prostředí a s přirozenou formulací výstupu – tento problém byl obzvláště patrný u ruštiny, a to kvůli její flektivní povaze. Lze tedy očekávat, že největší pozornost bude věnována právě práci v těchto oblastech. V budoucnu proto předpokládáme zlepšení technologií pro rozpoznávání řeči a odlišení řeči od šumu, případně lze předpokládat také implementaci technologií rozpoznávání hlasu, které najdeme v pokročilých hlasových asistentech jako Siri, Alexa či Cortana, které již dovedou hlas rozpoznávat velice dobře. S tím se může pojit i vstup nových velkých hráčů na trh automatického tlumočení. Má-li se automatické tlumočení přiblížit naplnění ambicí, které spočívají v přiblížení se lidské řeči, lze rovněž očekávat snahy o zlepšení hlasového výstupu automatického tlumočení – zejména u flektivních jazyků náš test ukázal velký prostor pro zlepšení.

Se zájmem budeme rovněž sledovat další vývoj systému STACL od Baidu, který se zatím nejvíce přibližuje skutečnému tlumočení a některé jeho znaky již naplňuje – lze proto předpokládat, že se jeho vývojáři budou ubírat stejným směrem a i nadále se inspirovat procesy a postupy, které najdeme u lidských tlumočnicků, a snažit se je napodobit. Logickým pokračováním je i zapojení technologií pro syntézu řeči, kterými STACL zatím vybaven není – bude jistě zajímavé pozorovat, jak se do systému podaří zapracovat strategie, které tlumočníci využívají právě ve fázi produkce řeči. STACL však v oblasti „skutečného“ automatického tlumočení představuje první vlaštovku, a proto se zdá, že se profesionální tlumočníci o budoucnost své profese, alespoň zatím, obávat nemusí. Dodnes totiž nebyla naplněna ani ambice vytvořit strojový překlad stejně dobrý (či dokonce lepší) než ten lidský, a i mnozí autoři odborných publikací se shodují na tom, že vznik „babylonské rybky“ je opravdu otázkou žánru sci-fi, nikoliv dohledné budoucnosti.

S ohledem na výše uvedené skutečnosti bude také zajímavé sledovat, jakým směrem se bude ubírat akademický výzkum. Na základě výsledků našeho testu a poznatků popsaných v této práci lze budoucím autorům pro bádání doporučit podrobnější prozkoumání jednotlivých dílčích součástí technologií pro automatické tlumočení (tedy korpusy, technologie pro rozpoznávání řeči a pro syntézu řeči, algoritmy napodobující strategie lidských tlumočnicků) s popisem jejich omezení, důvodů těchto omezení a možností, jak tato omezení překonat. Takový výzkum však bude nutné pojmut mezioborově, jelikož nastíněná problematika zasahuje nejen do translologie, ale i do obecné lingvistiky, počítačnické lingvistiky či programování, a přesah lze najít i do oborů jako jsou psychologie či výzkum osvojování řeči.

I na poli translologie však existuje dostatek prostoru pro další a podrobnější výzkum problematiky strojového překladu a automatického tlumočení, jelikož jsou tyto oblasti v akademické sféře v podstatě neprobádané. Pozornost je tak možné věnovat například hlubší analýze chyb, kterých se zkoumané systémy dopouští (například zmiňovaným chybám v segmentaci či ve skloňování a časování), testování těchto technologií s větším počtem účastníků, srovnání chybovosti v různých jazykových párech a směrech (například v párech ze stejných či různých skupin dle morfologické typologie jazyků). Do budoucna je rovněž možné zamyslet se nad stanovením kritérií hodnocení výhradně pro automatické tlumočení a také nad jeho jasnějším zařazením či nad definováním jednoznačné a ustálené terminologie.



## 7 Závěr

V rámci této práce jsme popsali řadu aspektů strojového překladu a automatického tlumočení a rovněž provedli test jednoho z volně dostupných programů pro automatické tlumočení. Konkrétně jsme testovali program Skype Translator, se kterým jsme provedli test použitelnosti a také následnou analýzu chyb.

Teoretická část práce poskytuje čtenáři přehled o historii a současném stavu zmiňovaných technologií a rovněž informace o jejich omezeních, způsobech hodnocení a potenciálních způsobech využití. Na základě těchto informací si tak čtenář dokáže vytvořit ucelený přehled o zkoumané oblasti a aktuálně dostupných řešeních a najít zde i inspiraci pro další bádání. Z teoretické části práce je také zřejmé, že i přesto, že strojový překlad si již nesmazatelně našel cestu jak do oboru překladatelství, tak i do běžného života lidí na celém světě, kteří denně využívají nejrůznější překladače pro své soukromé účely, profesionální překladatelé ho stále nemusí vnímat nutně jako konkurenci, ale naopak ho v určitých situacích mohou využít jako nástroj k zefektivnění práce. Automatické tlumočení je pak pro široké publikum novinkou posledních zhruba 10 let, ve svém jádru se však tyto programy neliší od strojového překladu a uživatelům jen usnadňují užívání pomocí možnosti hlasového zadávání vstupu a „předčítáním“ následného výstupu. I toto však lze považovat za velký krok dopředu, především z pohledu zpřístupnění těchto technologií širšímu okruhu uživatelů. Je však nutné zmínit, že tyto programy nejsou zdaleka dokonalé a od uživatelů vyžadují značnou míru spolupráce a přizpůsobení – v závislosti na konkrétním programu či zařízení může jít například o nutnost používat kratší věty, mluvit jasně a nahlas, hlídat přechytlivky, parazitní zvuky či falešné začátky atd. – aby fungovaly správně a komunikace tak proběhla úspěšně. Totéž ostatně prokázal i náš test, ve kterém si nutnosti přizpůsobovat se programu všimli i samotní uživatelé. Automatická řešení pro tlumočení, která by skutečně naplňovala znaky lidského tlumočení jsou pak zatím skutečně v začátcích a za pozornost stojí především program STACL od Baidu. Nicméně i přesto, že ve srovnání s lidskými tlumočníky má nesrovnatelně větší pracovní kapacitu a nemusí se tak vůbec potýkat s jejím správným využitím a rozdělením mezi jednotlivá úsilí, nedokáže zatím napodobit zdaleka všechny strategie a postupy lidských tlumočnicků a nepředstavuje proto pro profesionální tlumočníky konkurenci.

Empirická část práce popisuje test programu Skype Translator, který jsme provedli v režimu „speech to speech“ v jazykovém páru ruština-angličtina (obousměrně), s účastí pěti mluvčích angličtiny, kteří mají zároveň nulovou znalost ruštiny. Naším cílem bylo provedení testu použitelnosti programu. Vyhodnocování probíhalo na dvou úrovních, a to z hlediska procesu (zde hodnotili výkon programu samotní uživatelé, a to pomocí dotazníku – viz příloha), a rovněž jsme program hodnotili z hlediska výsledného produktu, prostřednictvím analýzy chyb. Z hodnocení účastníků testu vyplynulo, že i přes určité nedostatky dokáže program poměrně úspěšně zprostředkovat komunikaci mezi dvěma lidmi, kteří nehovoří stejným jazykem. Analýza chyb pak ukázala, že nejvíce program chyboval na poli správného rozpoznání toho, co vyřkli uživatelé, čehož si ostatně všimli i oni sami, jelikož právě v otázce věnované tomuto aspektu získal program od účastníků testu nejhorší hodnocení. Navzdory tomu ale nebyla prokázána korelace mezi hodnocením, které účastník udělil, a počtem chyb, kterých se program při testu s tímto účastníkem dopustil. Autorka práce rovněž věnovala jednu

podkapitolu vlastnímu uživatelskému hodnocení programu, které může sloužit jako inspirace pro další a podrobnější výzkum.

V poslední kapitole práce jsme se rovněž krátce zamysleli nad tím, co omezuje současný výzkum v oblasti strojového překladu a automatického tlumočení a jakým směrem se mohou ve svém bádání a snažení vydat vývojáři těchto technologií a akademici jak z oboru translatologie, tak v rámci mezioborového výzkumu.

Závěrem také považujeme za důležité zmínit, že i přes svůj rozsah práce nepokrývá zdaleka všechny aspekty strojového překladu a automatického tlumočení či přidružených oblastí (např. nástroje počítačem podporovaného překladu, speciální programy pro konferenční tlumočníky a jiné) a je pouze rozsáhlým úvodem do problematiky. Dané obory se však především v posledním desetiletí výrazně posunuly vpřed a představují tak dle našeho názoru veliké pole pro bádání jak z pohledu translatologického, tak z pohledu technologického, ale především z pohledu mezioborového – toho se totiž, i dle slov odborného konzultanta ze společnosti Baidu, nedostává, a spojení a blízká spolupráce translatologů, počítačnických lingvistů a počítačových profesionálů by tak mohlo být právě tím, co pomůže posouvat strojový překlad a automatické tlumočení ještě dál.

## 8 Bibliografie

A Strategy For American Innovation: Driving Towards Sustainable Growth And Quality Jobs. 2009. s. 22. Dostupné z: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a524201.pdf>

BAKER, Mona a Gabriela SALDANHA. Routledge encyclopedia of translation studies. 2nd ed. New York: Routledge, 2009. ISBN 978-0-415-36930-5.

BAUR, Wolfram et al. Man vs. Machine: Proceedings of the XXth FIT World Congress. Berlin: BDÜ Weiterbildungs – und Fachverlagsgesellschaft mbH, 2014, p. 93-99. ISBN 978-3-938430-63-7.

CONNEAU, Alexis et al. Word Translation Without Parallel Data. In: ICLR 2018 [online]. 2018 [cit. 2019-04-11]. Dostupné z: <https://arxiv.org/pdf/1710.04087.pdf>

ČEŇKOVÁ, Ivana. Úvod do teorie tlumočení. Praha: Česká komora tlumočnicků znakového jazyka, c2008. ISBN 978-80-87153-74-1.

ČEŇKOVÁ, Ivana. Teorie a didaktika tlumočení I. Univerzita Karlova, 2001. ISBN 80-7308-019-2.

ČERNOV, Gelij. Teorija i praktika sinchronnogo perevoda. Moskva: LIBROKOM, 2014. ISBN 978-5-397-04244-4.

DE PEDRO RICOY, Raquel a Jemina NAPIER. Introduction: Innovations in interpreting research methods. The International Journal for Translation & Interpreting Research [online]. **2017**(9), 1-3 [cit. 2017-06-02]. DOI: 10.12807/ti.109201.2017.a01. Dostupné z: <http://trans-int.org/index.php/transint/article/view/678/270>

DECAMP, Jennifer a Jost ZETZSCHE. A History Of Translation Technology In The United States. SIN-WAI, Chan. Routledge encyclopedia of translation technology. New York, 2015, s. 375-392. ISBN 978-0-415-52484-1.

DIABOVÁ, Amalaine, Petr KAUTSKÝ a Miroslav POŠTA. Tlumočení a jak na to, aneb, Chcete, aby vám rozuměli posluchači (a spolu s nimi i tlumočníci)? Praha: JTP, 2006. 23 s. ISBN 8086711986.

FENG, Xiaocheng et al. Topic-to-Essay Generation with Neural Networks. IJCAI, 2018. Dostupné z: <https://www.ijcai.org/proceedings/2018/0567.pdf>

GILE, Daniel. Basic Concepts and Models for Interpreter and Translator Training. John Benjamins Publishing Company, 2009. ISBN 9789027224323.

GUERRA, Ana Fernández. Machine Translation. Capabilities and limitations. Universitat de València, 2000. ISBN 8437042666.

HORVÁTH, Ildikó. The Modern Translator and Interpreter. Budapest, 2016, s. 183-191. ISBN 978 963 284 750 4. Dostupné z: <http://www.eltereader.hu/media/2016/04/HorvathTheModernTranslator.pdf>

HUTCHINS, John. Evaluation of MT systems. In: HUTCHINS, John a Harold SOMERS.: An introduction to machine translation. Academic Press, 1992, s. 161-174. ISBN 0-12-362830-X. příslušná kapitola dostupná z <http://www.hutchinsweb.me.uk/IntroMT-9.pdf>

HUTCHINS, John. Machine translation: a concise history. In: John Hutchins. Publications on machine translation, computer-based translation technologies, linguistics and other topics [online]. [cit. 2019-04-11]. Dostupné z: <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>

HUTCHINS, John. Machine Translation: History. In: Keith Brown, (Editor-in-Chief) Encyclopedia of Language & Linguistics, Second Edition, volume 7, s. 375-383. Oxford: Elsevier. 2006. Dostupné z: <http://www.hutchinsweb.me.uk/EncLangLing-2006.pdf>

HUTCHINS, John. Retrospect and Prospect in Computer-based Translation. In: Machine Translation Summit VII '99: MT in the Great Translation Era [online]. Kent Ridge Digital Labs, Singapore, 1999 [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/MTS-1999-Hutchins.pdf>

HUTCHINS, John. The development and use of machine translation systems and computer-based translation tools [online]. 1999 [cit. 2019-04-09]. Dostupné z: <http://www.hutchinsweb.me.uk/Beijing-1999.pdf>

CHO, Kyunghyun et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation [online]. 2014 [cit. 2019-04-11]. Dostupné z: <https://arxiv.org/pdf/1406.1078.pdf>

CHUNYU, Kit a Billy WONG TAK-MING. Evaluation in machine translation and computer-aided translation. In: SIN-WAI, Chan. Routledge encyclopedia of translation technology. New York, 2015, s. 213-236. ISBN 978-0-415-52484-1.

LEMMETTY, Sami. Review of Speech Synthesis Technology [online]. 1999 [cit. 2019-04-09]. Dostupné z: [http://research.spa.aalto.fi/publications/theses/lemmetty\\_mst/thesis.pdf](http://research.spa.aalto.fi/publications/theses/lemmetty_mst/thesis.pdf). Diplomová práce. Helsinki University of Technology. Vedoucí práce Matti Karjalainen.

LO, Chi-kiu a Dekai WU. MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies [online]. Portland: Association for Computational Linguistics, 2011, s. 220-229 [cit. 2019-04-10]. Dostupné z: <https://www.aclweb.org/anthology/P11-1023>

MA, Mingbo et al. STACL: Simultaneous Translation with Integrated Anticipation and Controllable Latency [online]. 2018. Dostupné z: <https://arxiv.org/pdf/1810.08398.pdf>

NIRENBURG, Sergei. Progress in machine translation. Tokyo: Ohmsha, 1992. ISBN 978-90-5199-074-4.

ORR, David a Victor SMALL. Comprehensibility of Machine-aided Translations of Russian Scientific Documents. Mechanical Translation and Computational Linguistics [online]. 1967(10) [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/MT-1967-Orr.pdf>

PAPINENI, Kishore, Salim ROUKOS, Todd WARD a Wei-Jing ZHU. Bleu: a Method for Automatic Evaluation of Machine Translation [online]. In: . 17.9.2001 [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/00/IBM-2001-Papineni.pdf>

PFÄFFLIN, Sheila. Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments. Mechanical Translation [online]. 1965, 1965(8) [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/MT-1965-Pfafflin.pdf>

PÖCHHACKER, Franz, Nadja GRBIĆ, Peter MEAD a Robin SETTON, ed. Routledge encyclopedia of interpreting studies. London, 2015. ISBN 978-0-415-63432-8.

SELIGMAN, Mark. Nine Issues in Speech Translation. [cit. 2019-04-09]. DOI: 10.1023/A:1011180928513. ISBN 10.1023/A:1011180928513.

SENNRICH, Rico, Barry HADDOW a Alexandra BIRCH. Neural Machine Translation of Rare Words with Subword Units. 2015. Dostupné z: <https://arxiv.org/pdf/1508.07909.pdf>

SKERRY-RYAN, R. J. et al. "Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron." 2018, ICML. Dostupné z: <https://arxiv.org/pdf/1803.09047.pdf>

SNOVER, Matthew, et al. A Study of Translation Edit Rate with Targeted Human Annotation. In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas [online]. 2006, s. 223-231 [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/AMTA-2006-Snover.pdf>

SOLTAU, Hagen, Tanja SCHULTZ, Thomas SHAAF a Florian METZE. Multilingual Speech Recognition. In: WAHLSTER, Wolfgang. VerbMobil: Foundations of Speech-to-Speech Translation. Springer, 2000. ISBN 3540677836.

SOMERS, Harold a Elizabeth WILD. Evaluating Machine Translation: the Cloze Procedure Revisited. In: Translating and the Computer 22: Proceedings of the Twenty-second international conference [online]. Londýn: Aslib, 2000 [cit. 2019-04-10]. Dostupné z: <http://www.mt-archive.info/00/Aslib-2000-Somers.pdf>

SVOBODA, Tomáš. Hodnocení kvality strojového překladu. In: ZEHNALOVÁ, Jitka. Kvalita a hodnocení překladu: Modely a aplikace. Univerzita Palackého v Olomouci, 2015, s. 245-272. ISBN 978-80-244-4795-7.

SVOBODA, Tomáš. Man and Machine: Translation in the Era of Augmented Reality.

TAKEZAWA, Toshiyuki et al. (1998). A Japanese-to-English speech translation system: ATR-MATRIX. Dostupné z: [https://www.researchgate.net/publication/221492004\\_A\\_Japanese-to-English\\_speech\\_translation\\_system\\_ATR-MATRIX](https://www.researchgate.net/publication/221492004_A_Japanese-to-English_speech_translation_system_ATR-MATRIX)

TAN, Lee. Speech Translation. In: SIN-WAI, Chan. The Routledge Encyclopedia of Translation Technology. 2015, s. 619-631. ISBN 978-0-415-52484-1.

TORAL, Antonio a Andy WAY. What Level of Quality can Neural Machine Translation Attain on Literary Text? [online]. 15. ledna 2018 [cit. 2019-04-09]. Dostupné z: <https://arxiv.org/pdf/1801.04962.pdf>

USZKOREIT, Jakob. Transformer: A Novel Neural Network Architecture for Language Understanding. 2017. Dostupné z: <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

VASCONCELLOS, Muriel a Marjorie LEÓN. Spanam And Engspan: Machine Translation At The Pan American Health Organization. Computational Linguistics [online]. 1985(Volume 11) [cit. 2019-04-11]. Dostupné z: <https://www.aclweb.org/anthology/J85-2003>

WAHLSTER, Wolfgang, ed. Verbmobil: Foundations of Speech-to-Speech Translation. Springer, 2000. ISBN 3540677836.

WAIBEL, Alex, Alon LAVIE a Lori LEVIN. (1999). JANUS – A System for Translation of Conversational Speech. Dostupné z: [https://www.researchgate.net/publication/2287818\\_JANUS\\_-\\_A\\_System\\_for\\_Translation\\_of\\_Conversational\\_Speech](https://www.researchgate.net/publication/2287818_JANUS_-_A_System_for_Translation_of_Conversational_Speech)

WHITE, John. How to evaluate machine translation. In: SUMMERS, Harold. Computers and Translation: A Translator's Guide. John Benjamins Publishing, 2003, s. 211-244. ISBN 9027216401.

WILKS, Yorick. Machine translation: its scope and limits. New York: Springer, 2008. ISBN 978-0-387-72773-8.

WU, Yonghui et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation [online]. In: . 2016 [cit. 2019-04-11]. Dostupné z: <https://arxiv.org/pdf/1609.08144>

ZONG, Chengqing a Mark SELIGMAN. Toward Practical Spoken Language Translation. Machine Translation [online]. 2005, 19(2), 113-137 [cit. 2019-04-09]. DOI: 10.1007/s10590-006-9000-z. ISSN 0922-6567. Dostupné z: <http://www.nlpr.ia.ac.cn/cip/ZongPublications/2006.06.23%20Machine%20Translation%20Journal%20OnlinePDF.pdf>

### **Výhradně webové zdroje: blogy, články, webové stránky**

'A Fish In Your Ear': What Gets Lost In Translation. In: National Public Radio [online]. 2011 [cit. 2019-04-11]. Dostupné z: <https://www.npr.org/2011/11/14/142309214/meaning-of-everything-often-lost-in-translation?t=1550500359963&t=1551877319768>

A Neural Network for Machine Translation, at Production Scale. Google AI Blog [online]. 27.09.2016 [cit. 2019-04-11]. Dostupné z: <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>

AI-powered translation still needs work after errors mar debut at Boao Forum | South China Morning Post. International Edition | South China Morning Post [online]. Copyright © 2019 South China Morning Post Publishers Ltd. All rights reserved. [cit. 09.04.2019]. Dostupné z:

<https://www.scmp.com/tech/innovation/article/2141940/ai-powered-translation-still-needs-work-after-errors-mar-debut-boao>

Baidu Research Announces Breakthrough in Simultaneous Translation [online]. 24.10.2018 [cit. 2019-04-11]. Dostupné z: <http://research.baidu.com/Blog/index-view?id=107>

Can't I Just Use Google Translate?. Language Translation Services from K International [online]. Copyright © Copyright 2018 [cit. 09.04.2019]. Dostupné z: <https://www.k-international.com/blog/google-translate/>

Facebook Acquires “Mobile Technologies”, Developer Of Speech Translation App Jibbig – TechCrunch. TechCrunch – Startup and Technology News [online]. Dostupné z: <https://techcrunch.com/2013/08/12/facebook-acquires-mobile-technologies-speech-recognition-and-jibbig-app-developer/?guccounter=1>

Google CEO Sundar Pichai revealed a jaw-dropping fact about its translation app that shows how much money is still sitting on the table. Business Insider [online]. [cit. 2019-04-11]. Dostupné z: <https://www.businessinsider.com/sundar-pichai-google-translate-143-billion-words-daily-2018-7>

Microsoft Translator in the Classroom – YouTube. YouTube [online]. Dostupné z: [https://www.youtube.com/watch?v=5tZn\\_oslNXw](https://www.youtube.com/watch?v=5tZn_oslNXw)

ili – Instant offline translation device for travelers. [online]. Dostupné z: <https://iamili.com/us/>

Integration with Machine Translation systems. Memoq [online]. [cit. 2019-04-11]. Dostupné z: <https://www.memoq.com/en/integration-with-machine-translation>

Internet Trends Report 2017 | Kleiner Perkins. Kleiner Perkins | Make History [online]. Dostupné z: <https://www.kleinerperkins.com/perspectives/internet-trends-report-2017>

Jibbig Translator 2.0 3.2 Free Download [online] [cit. 09.04.2019]. Dostupné z: <https://jibbig-translator-2-0.soft112.com/>

Machine Translates Literature and About 25% Was Flawless, Research Claims | Slator | Language Industry Intelligence [online]. Copyright © 2019 [cit. 09.04.2019]. Dostupné z: <https://slator.com/technology/machine-translates-literature-and-about-25-was-flawless-research-claims/>

Microsoft Translator Languages – Microsoft Translator. Microsoft Corporation [online]. Dostupné z: <https://www.microsoft.com/en-us/translator/languages/>

Multilingual Chat and Machine Translation Use Case | SDL. SDL: Language Translation & Content Management Company [online]. Dostupné z: <https://www.sdl.com/download/customer-support-multilingual-chat-and-machine-translation-use-case/71534/>

Nova Smartest Machine on Earth PBS Documentary. YouTube [online]. Dostupné z: <https://www.youtube.com/watch?v=yywl7ugBubA>

Prague Dependency Treebank 3.0 [online]. [cit. 2019-04-11]. Dostupné z: <https://ufal.mff.cuni.cz/pdt3.0>

Pražský závislostní korpus 2.0. Home page | ÚFAL [online]. Copyright © 2006 ÚFAL [cit. 11.03.2019]. Dostupné z: <https://ufal.mff.cuni.cz/pdt2.0/index-cz.html>

TC-STAR Project: Technology and Corpora for Speech to Speech Translation. TC-STAR Project: Technology and Corpora for Speech to Speech Translation [online]. Copyright © 2004 ITC [cit. 09.04.2019]. Dostupné z: <http://tcstar.org/>

TITE, Barbora. Jáchyme, hod' tlumočnicka do stroje [online]. 1.11.2018 [cit. 2019-04-09]. Dostupné z: <https://barboratite.cz/jachyme-hod-tlumocnika-stroje/>

Translate One2One Real-Time Translation Earpieces – Lingmo International. Lingmo International – Understand the World [online]. Copyright © [cit. 09.04.2019]. Dostupné z: <https://lingmointernational.com/translate-one2one/>

Waverly Labs. Waverly Labs: No More Language Barriers [online]. Dostupné z: <https://www.waverlylabs.com/products-bf/>

Vocabulary size and auditory word recognition in preschool children. National Center for Biotechnology Information [online]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5400288/>

What Is a Transformer? — Inside Machine Learning – DZone AI. DZone [online]. Dostupné z: <https://dzone.com/articles/what-is-a-transformer-inside-machine-learning>

When Using Machine Translation Makes Sense (and when it doesn't) | Acclaro. Professional Translation Services | Acclaro [online]. Copyright © 2002 [cit. 09.04.2019]. Dostupné z: <https://www.acclaro.com/blog/when-using-machine-translation-makes-sense/>

Yandex — Company blog — One model is better than two. Yandex.Translate launches a hybrid machine translation system. Yandex [online]. Copyright ©. Dostupné z: <https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/>

Yandex.TranslateFAQ [online]. [cit. 2019-04-11]. Dostupné z: <https://yandex.com/support/translate/index.html>



## 9 Přílohy

Seznam příloh:

Příloha 1: Odkaz pro přístup k nahrávkám testu (heslo lze vyžádat na e-mailové adrese ska.yulia@gmail.com)

Příloha 2: Pokyny k testu (v angličtině)

Příloha 3: Dotazník účastníka č.1

Příloha 4: Dotazník účastníka č.2

Příloha 5: Dotazník účastníka č.3

Příloha 6: Dotazník účastníka č.4

Příloha 7: Dotazník účastníka č.5

Příloha 1: <https://1drv.ms/f/s!AqgfZE91YHTzqBsHSWOb9rgliunE>

## Příloha 2: Pokyny k testu (v angličtině)

### Introduction

Please note that I will be recording the call and Skype's screen during the call. The recordings will only be used for the purpose of my own evaluation of the experiment and will not be made available to any third parties without your prior written consent.

The guidelines are divided into levels based on the CEFR (Common European Framework of Reference for Languages) to help identify the Skype Translator's capability of dealing with various complexity levels of conversation and vocabulary. Feel free to make visible transitions between these levels during your conversation (i.e. by saying "let's move on to level 2"). The times indicated for each level are approximate – you don't need a stopwatch



#### **Level 1**

Introduce yourselves to each other. **Present your partner with basic information about yourself**, like your name, age, family, job, where are you from, or where do you live. Say as much as you feel comfortable, or even make stuff up – the factual accuracy of the information is not important here, but please keep it realistic. Please refrain from asking questions during this introductory stage. **This stage should take about 1-2 minutes for each of you.**

#### **Level 2**

**Ask your partner to give you directions** (walking or involving public transport) from a bus station to a shopping center, write down what you hear, **then repeat them to see if you got them right** and thank your partner as if they were a stranger in the street. **Then, switch roles.** Please keep the instructions between 3 to 6 steps. The directions can be made up and don't need to represent the reality. Use landmarks, such as "turn left at the roundabout", "walk past the green building", etc. Keep it natural. **This stage should take about 2 minutes for each of you.**

#### **Level 3**

**Remember who you wanted to be when you were little?** An astronaut, an actor, a doctor? **Tell your partner about it, and about what made you dream of that career as a child.** Please keep it conversational at this level – **ask each other follow up questions but try to keep them simple.** Here's few examples to get you started: "Did you want to be a doctor because you had a doctor in your family?" "Did you want to be an astronaut because of a book you read or a movie you saw?" "Why did you not choose to pursue this dream?" "What made you want to become a programmer instead of an actor?" **This stage should take about 2-4 minutes for each of you.**

#### **Level 4**

**Talk about your favorite books or films.** If your partner doesn't know it, narrate the story and the main topic shortly. Try to use complex terminology, reported speech, and describe abstract ideas. **Do not try to purposefully simplify your speech.** Describe you like about your

chosen book/movie. **Try to express subjective opinions, ambiguity, and complex thoughts** (e.g. “I liked how the actor portrayed character A, but I think his relationship with character B was marked by their joint experience from their childhood when they...”). **Keep the conversation fluid, don’t stick to just one movie or book if the conversation takes you elsewhere, don’t be shy from interrupting each other, asking questions, and expressing opinions** – imagine you are chatting with your friend over a cup of coffee. **This stage should take about 5-7 minutes in total, as you don’t switch roles at this point.**

Příloha 4: Dotazník účastníka č.1

**Please rate the following aspects on a scale of 1-5 (1: worst, 5: best):**

### **USER EXPERIENCE**

**How easy was the setup of Skype's interpreted call?**

1: extremely difficult 2: difficult 3: neither easy nor difficult 4: easy 5: extremely easy

Your grade: 5

Comments: Second time I have used the feature, the first time was to call someone in Czech Republic.

**How well was the user interface arranged?** (*Was everything easy to find? Was the functionality of UI elements clear?*)

1: extremely poorly 2: poorly 3: neither well nor poorly 4: well 5: extremely well

Your grade: 4

Comments: No comments.

**What was the sound quality like?** (*Consider aspects like echo, sound lag, sound dropping off, etc.*)

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 3

Comments: The quality of the call was not the best due to the echo, but I think it could have possibly been due to connection.

**How good was the synthesized speech?** (*Consider aspects like naturalness, monotony, emotion, tone of voice, pauses, pronunciation, how pleasant the voice was*)

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 4

Comments: The voice was surprisingly synthesized quite well. The voice did not sound natural, but it sounded like clear speech from a pre-recorded voice. The emotions conveyed by the voice sounded quite natural in comparison with native English speakers. Overall, I would have to say that the voiceover was quite spectacular, and it was only in some cases that it sounded robotic.

### **QUALITY**

**How often did you have to repeat the input for Skype to recognize what you were saying correctly?**

1: extremely often 2: often 3: sometimes 4: hardly ever 5: not even once

Your grade: 3

Comments: It was only at times with certain topics that Skype was not able to recognize my speech well enough to be translate to Russian. In addition, words from other languages were not picked up by Skype.

**How often did you have to ask additional question in order to understand what your partner meant?**

1: extremely often    2: often    3: sometimes    4: hardly ever    5: not even once

Your grade: 4

Comments: I would rarely have to ask questions about what my partner meant, in most cases it was because I had begun speaking at the same time as my speaking partner and the software didn't fully recognize what to do.

**How well were you able to follow instructions based on Skype's output?** *(This question is specific to the exercise that involved asking for directions.)*

1: extremely poorly    2: poorly    3: neither well nor poorly    4: well    5: extremely well

Your grade: 5

Comments: No comments.

**Please rate how you felt about the amount of time between Skype's speech comprehension and speech production.** *(Did you feel like the lag was too long, normal, or very short?)*

1: extremely long    2: long    3: neither long nor short    4: short    5: extremely short

Your grade: 4

Comments: No comments.

**Overall, how comprehensible did you find Skype's output?** *(Think on a scale of complete gibberish to human-like performance)*

1: extremely poor    2: poor    3: normal    4: good    5: extremely good

Your grade: 5

Comments: No comments.

**Considering all the factors mentioned above, what is the overall grade you think Skype deserves?**

1: total disaster    2: not very good performance    3: average performance  
4: good performance    5: extremely good performance

Your grade: 4

Final comments: Paying attention to one's intonation and clarity of speech was found to be extremely important as well as speed of speech.

Příloha 5: Dotazník účastníka č.2

**Please rate the following aspects on a scale of 1-5 (1: worst, 5: best):**

### **USER EXPERIENCE**

**How easy was the setup of Skype's interpreted call?**

1: extremely difficult 2: difficult 3: neither easy nor difficult 4: easy 5: extremely easy

Your grade: 4

Comments: I didn't understand how it would work and was surprised to hear both Russian and then English – I liked that.

**How well was the user interface arranged?** (*Was everything easy to find? Was the functionality of UI elements clear?*)

1: extremely poorly 2: poorly 3: neither well nor poorly 4: well 5: extremely well

Your grade: 2

Comments: Sometimes the caption disappeared before I finished reading them.

**What was the sound quality like?** (*Consider aspects like echo, sound lag, sound dropping off, etc.*)

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 4

Comments: no problems

**How good was the synthesized speech?** (*Consider aspects like naturalness, monotony, emotion, tone of voice, pauses, pronunciation, how pleasant the voice was*)

1: extremely poor 2: very poor 3: normal 4: very good 5: extremely good

Your grade: 5

Comments: a very pleasant surprise

### **QUALITY**

**How often did you have to repeat the input for Skype to recognize what you were saying correctly?**

1: extremely often 2: often 3: sometimes 4: hardly ever 5: not even once

Your grade: 3

Comments:

**How often did you have to ask additional question in order to understand what your partner meant?**

1: extremely often    2: often    3: sometimes    4: hardly ever    5: not even once

Your grade: 3

Comments:

**How well were you able to follow instructions based on Skype's output?** *(This question is specific to the exercise that involved asking for directions.)*

1: extremely poorly    2: poorly    3: neither well nor poorly    4: well    5: extremely well

Your grade: 4

Comments:

**Please rate how you felt about the amount of time between Skype's speech comprehension and speech production.** *(Did you feel like the lag was too long, normal, or very short?)*

1: extremely long    2: long    3: neither long nor short    4: short    5: extremely short

Your grade: 4

Comments: It was perfect, not distracting but I had troubles not being sure if I can talk or should wait for the other person to say something, but I'm not sure there is a way to solve this.

**Overall, how comprehensible did you find Skype's output?** *(Think on a scale of complete gibberish to human-like performance)*

1: extremely poor    2: poor    3: normal    4: good    5: extremely good

Your grade: 4

Comments:

**Considering all the factors mentioned above, what is the overall grade you think Skype deserves?**

1: total disaster    2: not very good performance    3: average performance  
4: good performance    5: extremely good performance

Your grade: 4

Final comments: I was blown away that this exists, I thought it would be really bad.

Příloha 6: Dotazník účastníka č.3

**Please rate the following aspects on a scale of 1-5 (1: worst, 5: best):**

### **USER EXPERIENCE**

**How easy was the setup of Skype's interpreted call?**

1: extremely difficult 2: difficult 3: neither easy nor difficult 4: easy 5: extremely easy

Your grade: 3

Comments: It was very simple to set up once I figured how it worked but was a bit confusing to begin with

**How well was the user interface arranged?** (*Was everything easy to find? Was the functionality of UI elements clear?*)

1: extremely poorly 2: poorly 3: neither well nor poorly 4: well 5: extremely well

Your grade: 4

Comments: Not sure exactly which UI elements, but it was very easy to find and read the translated text as it appeared on the screen

**What was the sound quality like?** (*Consider aspects like echo, sound lag, sound dropping off, etc.*)

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 5

Comments: There wasn't any issue at all with the sound quality, i could understand perfectly. Even the computer speaking was clear.

**How good was the synthesized speech?** (*Consider aspects like naturalness, monotony, emotion, tone of voice, pauses, pronunciation, how pleasant the voice was*)

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 4

Comments: Its pronunciation was very good; the voice was pleasant, and it was very clear and understandable to me. However, it was extremely robotic and monotonous like listening to text to speech

### **QUALITY**

**How often did you have to repeat the input for Skype to recognize what you were saying correctly?**

1: extremely often 2: often 3: sometimes 4: hardly ever 5: not even once

Your grade: 3



Comments:

**How often did you have to ask additional question in order to understand what your partner meant?**

1: extremely often    2: often    3: sometimes    4: hardly ever    5: not even once

Your grade: 4

Comments:

**How well were you able to follow instructions based on Skype's output?** *(This question is specific to the exercise that involved asking for directions.)*

1: extremely poorly    2: poorly    3: neither well nor poorly    4: well    5: extremely well

Your grade: 5

Comments:

**Please rate how you felt about the amount of time between Skype's speech comprehension and speech production.** *(Did you feel like the lag was too long, normal, or very short?)*

1: extremely long    2: long    3: neither long nor short    4: short    5: extremely short

Your grade: 3

Comments: It felt a bit long at times, if I was speaking in short concise sentences like I was waiting for it to start so I could keep speaking.

**Overall, how comprehensible did you find Skype's output?** *(Think on a scale of complete gibberish to human-like performance)*

1: extremely poor    2: poor    3: normal    4: good    5: extremely good

Your grade: 4

Comments:

**Considering all the factors mentioned above, what is the overall grade you think Skype deserves?**

1: total disaster    2: not very good performance    3: average performance  
4: good performance    5: extremely good performance

Your grade: 4

Final comments:

I was very surprised how good it was, as a layman I didn't even know this kind of thing was possible and so expected it to be completely unintelligible, but it was surprisingly good and completely understandable in English.

Příloha 7: Dotazník účastníka č.4

**Please rate the following aspects on a scale of 1-5 (1: worst, 5: best):**

### **USER EXPERIENCE**

**How easy was the setup of Skype's interpreted call?**

1: extremely difficult 2: difficult 3: neither easy nor difficult 4: easy 5: extremely easy

Your grade: 4

Comments: It was straight forward.

**How well was the user interface arranged?** *(Was everything easy to find? Was the functionality of UI elements clear?)*

1: extremely poorly 2: poorly 3: neither well nor poorly 4: well 5: extremely well

Your grade: 4

Comments:

**What was the sound quality like?** *(Consider aspects like echo, sound lag, sound dropping off, etc.)*

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 4

Comments:

**How good was the synthesized speech?** *(Consider aspects like naturalness, monotony, emotion, tone of voice, pauses, pronunciation, how pleasant the voice was)*

1: extremely poor 2: poor 3: normal 4: good 5: extremely good

Your grade: 3

Comments: The interpreting voice sounded a little retro. I found the pause to be longer than usual and once the interpretation started it was hard to speak over it. The conversation didn't seem natural because once an error is spoken, you must wait to correct yourself.

### **QUALITY**

**How often did you have to repeat the input for Skype to recognize what you were saying correctly?**

1: extremely often 2: often 3: sometimes 4: hardly ever 5: not even once

Your grade: 3

Comments: I had to repeat the names of places and people. Overall, I felt it was okay though.

**How often did you have to ask additional question in order to understand what your partner meant?**

1: extremely often    2: often    3: sometimes    4: hardly ever    5: not even once

Your grade: 4

Comments: I believe only 1 question or statement had to be repeated for me to understand. Some words were off, but I could still follow what my partner meant.

**How well were you able to follow instructions based on Skype's output? (This question is specific to the exercise that involved asking for directions.)**

1: extremely poorly    2: poorly    3: neither well nor poorly    4: well    5: extremely well

Your grade: 3

Comments: I found it difficult to follow the instructions because it doesn't save the conversation and with longer explanations the text disappeared faster. It was difficult to give long detailed directions without speaking very slowly or repeating several times.

**Please rate how you felt about the amount of time between Skype's speech comprehension and speech production. (Did you feel like the lag was too long, normal, or very short?)**

1: extremely long    2: long    3: neither long nor short    4: short    5: extremely short

Your grade: 3

Comments: The pause didn't feel natural and sometimes I wasn't sure when the interpretation would start. However, I think after using it several times it would become easier.

**Overall, how comprehensible did you find Skype's output? (Think on a scale of complete gibberish to human-like performance)**

1: extremely poor    2: poor    3: normal    4: good    5: extremely good

Your grade: 4

Comments: I could understand everything very well, however, it was slightly robotic.

**Considering all the factors mentioned above, what is the overall grade you think Skype deserves?**

1: total disaster    2: not very good performance    3: average performance  
4: good performance    5: extremely good performance

Your grade: 3-4

Final comments: I feel compared to google translate this could be very good on a professional level. Having a prepared document before using the interpreter makes using it a lot easier. I believe on the personal level it needs improvement in sounding more human and the pauses could be less confusing.

## Příloha 8: Dotazník účastníka č.5

**Please rate the following aspects on a scale of 1-5 (1: worst, 5: best):**

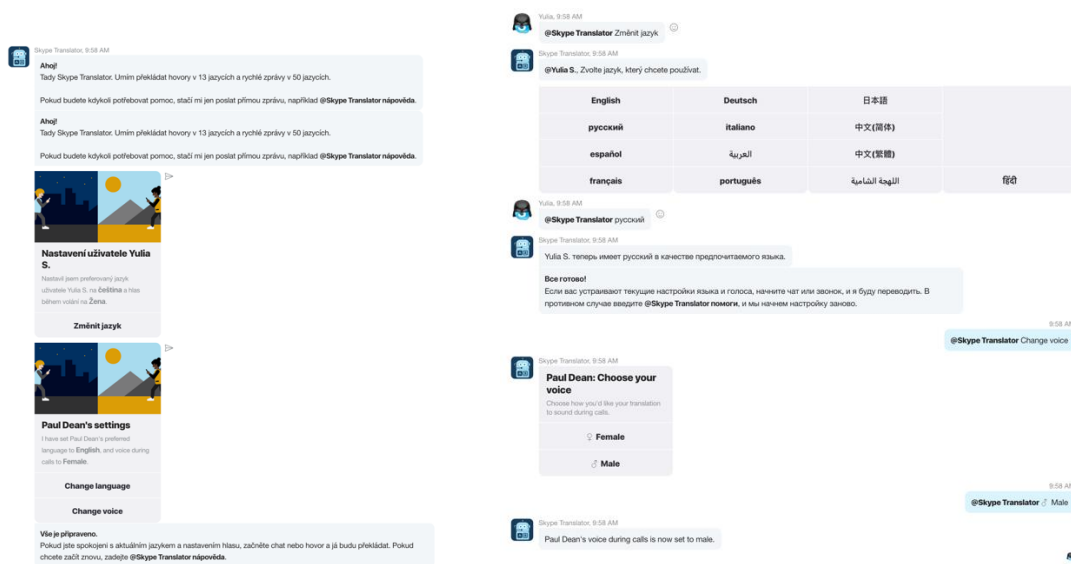
### USER EXPERIENCE

#### **How easy was the setup of Skype's interpreted call?**

1: extremely difficult 2: difficult 3: neither easy nor difficult 4: easy 5: extremely easy

Your grade: 4

Comments: There was a lot of text in a language I didn't understand, but I realized I was just supposed to select my language (English) and sex (Male). I think it would have been a lot easier and less intimidating with just a simple UI – first a list of languages to click on. Then, once selected, 'Male' or 'Female' to click on in that language – without the surrounding instructions: Here is what I saw, 2 whole pages on startup:



**How well was the user interface arranged? (Was everything easy to find? Was the functionality of UI elements clear?)**

1: extremely poorly 2: poorly 3: neither well nor poorly 4: well 5: extremely well

Your grade: 3

Comments: see above for starting experience. There was a lot of unnecessary stuff, probably as it was showing Julia's Russian stuff as well as my English parts? Once we started it was all very easy to use. The biggest problem was that when it shows what it thinks I said, the text appears for a very short period of time before disappearing entirely. This means that if I speak for any length of time it's impossible to see if the program understood me correctly. It would help to be able to see a record of what the program thinks I have said. Similarly, if the program has clearly misunderstood me, it would be useful to have an option to stop the translation and start again before it is read out completely to the other person. At certain points it completely misinterpreted me and could potentially make me look or sound stupid. Having the ability to prevent that would be useful in many situations.

**What was the sound quality like?** (Consider aspects like echo, sound lag, sound dropping off, etc.)

1: extremely poor      2: poor                                      3: normal      4: good                                      5: extremely good

Your grade: 5

Comments: No problems at all.

**How good was the synthesized speech?** (Consider aspects like naturalness, monotony, emotion, tone of voice, pauses, pronunciation, how pleasant the voice was)

1: extremely poor      2: poor                                      3: normal      4: good                                      5: extremely good

Your grade: 4

Comments: In all honesty, I tended to listen to the voice of the original speaker, from which I would get tone and clues as to emotion and intent. I would then often READ what was being said before ever listening to the 'bot'. I would only listen to the bot towards the end if the text had disappeared before reading it. I think this was just my natural preference, but I can understand why. I love that you can hear the speakers original voice, the 'bot' was excellent, but lacks emotion, context, tone, intent. All of these can be inferred from the original speaker then applied to translation. It's easier to apply these things to text then to spoken words, so I tended to ignore parts of the 'bot' where I could.

Similarly, the translator often struggled with recognizing the difference between a statement and a question. It struggled with other basic grammar choices, such as placing a comma – which can often be essential. One annoying thing was that if you paused too long, then the translator would stop and start translating. At one point it translated the first word of a sentence as I was completing the rest of the sentence. This interrupted my flow, which caused the translator to stop again and start translating. This can be a disorientating experience and on a couple of occasions I just stopped talking until it was finished then started again.

All this said. I think it's absolutely excellent. I was blown away by how natural the whole experience became after a couple of minutes and how I quickly felt like I was talking fluidly with the other person in my own language.

## **QUALITY**

**How often did you have to repeat the input for Skype to recognize what you were saying correctly?**

1: extremely often      2: often                                      3: sometimes      4: hardly ever                                      5: not even once

Your grade: 3

Comments: As a percentage of how much the translator got right, the mistakes were very small indeed. However, they did happen with enough frequency and at strange points, to be very noticeable and to cause significant breaks in the flow of conversation. At times these were funny, at other times they could be embarrassing. For instance, I was 'accused' of

swearing by the translator on a number of occasions when I hadn't sworn at all. Similarly, at one point, Julia told me she loved me, when she really meant she likes the films of a director we were talking about. On other occasions they just didn't make sense – as when it could not understand I was trying to say 'cat'.

**How often did you have to ask additional question in order to understand what your partner meant?**

1: extremely often    2: often    3: sometimes    4: hardly ever    5: not even once

Your grade: 4

Comments: This happened on only a couple of occasions. Often the mistranslations still left the sentences completely understandable. I can't really recall ever having to ask to repeat a question, and I think I was only asked to do so a couple of times. Often the intent is obvious even when the translation is not perfect.

**How well were you able to follow instructions based on Skype's output? (This question is specific to the exercise that involved asking for directions.)**

1: extremely poorly    2: poorly    3: neither well nor poorly    4: well    5: extremely well

Your grade: 4

Comments: I felt the whole directions section was very good. It was clear and simple to understand and where there was any uncertainty it was simple enough to ask for clarification. I felt this was one of the strongest parts of the experience and an obvious use.

**Please rate how you felt about the amount of time between Skype's speech comprehension and speech production. (Did you feel like the lag was too long, normal, or very short?)**

1: extremely long    2: long    3: neither long nor short    4: short    5: extremely short

Your grade: 4

Comments: I hardly notice lag at all.

**Overall, how comprehensible did you find Skype's output? (Think on a scale of complete gibberish to human-like performance)**

1: extremely poor    2: poor    3: normal    4: good    5: extremely good

Your grade: 5

Comments: As described above. I thought the whole experience was extremely good, even where there were mistakes these were mostly understandable.

**Considering all the factors mentioned above, what is the overall grade you think Skype deserves?**

1: total disaster    2: not very good performance    3: average performance  
4: good performance    5: extremely good performance

Your grade: 4 (Only a couple of minor niggles away from being a 5)

Final comments:

*(Use this section to add any general thoughts, describe your expectations (if you had any), how well were they met, etc.)*

I have had experience with translators. I regularly use google translate and even use the camera function to look at foreign words and translate them live on screen. So, I have seen some pretty cool translation experiences. I also follow tech news and have seen in ear monitors that use google translate to provide a similar 'Universal translator' experience. However, most of these have varying levels of success and are not quite there yet. So far, this was one of the most seamless experiences I have seen or heard of. I was very impressed with the whole thing, and how, in particular, it began to feel natural within a very short period of time.