

Univerzita Karlova

Filozofická fakulta

Ústav dálného východu / koreanistika

Bakalářská práce

Markéta Pašteková

Strojový překlad a korejština

Machine translation and Korean language

Praha 2019

Vedoucí práce: PhDr. Tomáš Horák, Ph.D.

Poděkování

Tímto bych ráda poděkovala PhDr. Tomáši Horákovi, Ph.D. za to, že se ujal vedení mé práce, doc. Miriam Löwensteinové za její podporu a rady a také Miloši Cholavovi, který po celou dobu poskytoval morální podporu při vypracovávání této práce.

Prohlášení:

Prohlašuji, že jsem bakalářskou/diplomovou/rigorózní/dizertační práci vypracoval/a samostatně, že jsem řádně citoval/a 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 1.7.2019

Markéta Paštéková

Klíčová slova:

Korejština, strojový překlad, jazykové teorie, aplikovaná lingvistika, modely strojového překladu

Klíčová slova (anglicky):

Korean language, machine translation, theories of language, applied linguistics, machine translation models

Abstrakt

Bakalářská práce se zabývá strojovým překladem v souvislosti s korejštinou. První část je teoretickou částí, kde jsou představeny hlavní směry strojového překladu v chronologickém pořadí a výhody a nevýhody těchto modelů strojového překladu včetně příkladů z korejského jazyka. Druhá, praktická část představuje moderní světové strojové překladače Naver Papago a Google Translate. Těmto překladačům jsou předloženy příklady textů v korejštině z publicistického a komerčního odvětví a také z odvětví mluvené řeči. Tato část také obsahuje správné anglické překlady pro porovnání s překlady vyprodukovanými zmiňovanými překladači. Cílem těchto překladů je poskytnout nahlédnutí na typické chyby v překladu a tyto chyby jsou analyzovány z pohledu teoretické části, což umožňuje vytvořit spojení mezi teorií a praxí. Oba překladače jsou pak také se sebou porovnány a slouží jako příklad jak stejné modely strojového překladu produkcí různé překlady a s nimi i chyby. Poslední část obsahuje krátké shrnutí současného stavu strojového překladu a jeho vyhlídky do budoucna.

Abstract:

This bachelor thesis deals with machine translation in relation to Korean language and as a . First part is a theoretical part in which are presented main directions of machine translation in chronological order and advantages and disadvantages of these machine translation models including examples from Korean language. Second, practical part introduces modern global machine translators Naver Papago and Google Translate. These translators are given text examples in Korean from publicistic and commerce field and also from spoken language. This part also includes proper English translation for comparison with translations produced by mentioned machine translators. The purpose of these translations is to provide an insight to typical errors in translations which are analyzed from the view of theoretical part which allows us to see a connection between theory and practice. Both translators are also compared with each other and act as an example of how same machine translation models produce different translations and mistakes. Last part contains a brief summary of the current state of machine translation and its expectations for the future.

Obsah

1. Úvod	7
2. Přístupy strojového překladu.....	9
2.1 Přímý překlad.....	9
2.2 Strojový překlad založený na pravidlech (Rule-Based MT)	10
2.2.1 Transferový překlad	16
2.2.2 Interlingva.....	16
2.2.3 Výhody a nevýhody překladu založeného na pravidlech	17
2.3 Statistický překlad	18
2.3.1 Překlad založený na příkladech (Example-Based MT)	21
2.3.2 Výhody a nevýhody překladače statistického překladače	22
2.4 Neurální strojový překlad	22
3. Srovnání překladačů Naver Papago a Google Translate	26
3.1 Naver Papago.....	27
3.2 Google Translate	28
3.3 Porovnání schopností překladačů.....	29
3.3.1 Publicistický text	30
3.3.2 Komerční text	36
3.3.3 Mluvená řeč	42
3.4 Shrnutí výsledků	48
4. Buoucnost strojového překladu	50
5. Závěr	52
Seznam použité literatury.....	54
Přílohy	63

1. Úvod

Strojový překlad se stal dnes běžnou součástí našich životů, která se stala v moderním globalizovaném světě nutností. Vzhledem k obrovskému množství dokumentů z oblasti vojenství, práva, zdravotnictví a dalších je cenným pomocníkem, který sice v současnosti stále nedokáže plně nahradit lidského překladatele, ale může překladatelskou práci minimálně velmi urychlit. V první řadě plní funkci automatického slovníku, který za prvé předčí klasické papírové slovníky, ve kterých musí uživatel zdoluhavě listovat, a to nejen svou rychlostí, ale také aktuálností, protože slovní zásoba se neustále přirozeně vyvíjí a do slovníku je potřeba nová slova dodávat a upozadit zastaralé termíny. Strojový překlad ovšem dnes už zdaleka nekončí jen obyčejných internetových slovníků. Už od dob studené války se snaží lidé o mnohem obtížnější úkol, a to naučit stroje nejen smysluplně překládat jednotlivá slova, ale i celé větné konstrukce.

Ačkoliv v současnosti stále neexistuje univerzální strojový překladač, který by lidského překladatele zvládl plně nahradit, za 70 let jeho vývoje dosáhl obrovského pokroku. Během jeho vývoje bylo využito mnoho různých přístupů od čistě lingvistického přes hybridní řešení až po čistě matematický přístup. Pokud se podíváme na kvalitu a možnosti dnešních překladačů, je stále možné najít mnohé nedostatky a během jeho dosavadního vývoje se mnohokrát nabízela otázka, jestli je vůbec možné vytvořit takový strojový překladač, který by se dokázal vypořádat s překladem nabízených textů stejně dobře nebo i lépe než člověk, který má sice limitovanou paměť, ale na rozdíl od stroje dokáže v textu pochopit souvislosti a kontext, ve kterém je zamýšlen.

V této otázce záleží na úhlu pohledu. Nemůžeme očekávat, že strojový překladač dosáhne úrovně lidského překladatele v oblasti uměleckých textů, na druhou stranu poptávku po překladech tvoří spíše texty odborné, publicistické, patentové texty a další, u kterých není cílem zachování umělecké formy. Krom toho, pokud se podíváme na současné překladače od společností Google nebo Naver, běžní lidé po celém světě je používají k překladu textů všeho druhu; odborných textů, článků na internetu, konverzací apod. a hlavním požadavkem pro tyto překlady bývá především porozumění, což je cíl, který už je pro strojový překlad realistický.

Jak uvádí John Hutchins ve své knize *An introduction to machine translation*, hlavní překážky pro strojový překlad neleží v možnostech počítače, nýbrž jde o překážky lingvistického rázu, jako je lexikální nejednoznačnost, složitá syntaktická struktura, rozdíly ve slovní zásobě apod.

¹ V této práci se budu zabývat překladem dvojice korejštiny s angličtinou, která dnes zaujímá pozici světového jazyka, a která má i vedoucí pozici ve strojovém překladu. Pokud jde o tuto dvojici jazyků, jsou výše zmíněné překážky o to znatelnější, kvůli rozdílům mezi korejštinou s charakteristikou aglutinačního jazyka a angličtinou, která se typologicky řadí mezi jazyky izolační.

Tuto práci rozdělím do několika částí. V první teoretické části představím chronologicky hlavní přístupy technologie strojového překladu a pokusím se vypíchnout problémové části jednotlivých přístupů zejména v souvislosti s korejským jazykem.² V druhé části bakalářské práce se zaměřím na současné veřejně dostupné automatické strojové překladače jako je Google Translate a Naver Papago a na několika ukázkách různých druhů textů vyzdvihnu, jakých chyb se překladače dopouští a pokusím se chyby dát do souvislosti s přechozí teoretickou částí a nastínit, pro jaký typ překladače jsou typické. Pokusím se také poukázat na případné rozdíly mezi těmito překladači. V poslední části se pokusím rozvinout úvahu ohledně současného stavu a budoucnosti strojových překladačů.

¹ HUTCHINS, W. John, HAROLD L. S. An introduction to machine translation. 2. vydání. Londýn: Acad. Press, 1992. ISBN 012362830X. S. 2.

² Teoretické modely jsou často předpovězeny dříve, než se začnou uvádět do praxe. Přibližná data vzniku uvedená v teoretické části se proto mohou v různých zdrojích lišit. Jako určující hodnotu jsem zde určila dobu, kdy se začala teorie převádět do praxe.

2. Přístupy strojového překladu

Automatický překlad pomocí počítačové techniky je záležitost, která spadá jak do oboru IT a tak zároveň i do lingvistiky. Od jeho počátků v 50. letech bylo vyzkoušeno několik různých přístupů, a prvním z nich je tzv. přímý překlad, následovaný překladačem založeným na pravidlech, statistickým překladačem a momentálně nejmodernějšími překladači založenými na neurálních sítích. Vývoj strojového překladu ve světě má počátky již v polovině minulého století zejména ve Spojených státech amerických. Značnou motivací pro vývoj této technologie najdeme ve studené válce, během které se Američané potýkali s velkým množstvím ruských textů, přičemž byla důležitá především rychlost překladu.

Jižní Korea však začala s vlastním vývojem strojového překladu se značným zpožděním. Teprve začátkem 80. let 20. století se začaly v této zemi řešit základní technické problémy se zpracováním korejštiny, jako zpracování písma, fonty apod. Následně v polovině 80. let investovala Korea do vývoje japonsko-korejských a anglicko-korejských překladačů a tomuto odvětví se začalo věnovat několik korejských institucí, mezi které patří KAIST, SERI, ETRI a Söulská národní univerzita.³

V následující kapitole se budu věnovat jednotlivým přístupům a jejich výhodám a nedostatkům při překladu dvojice jazyků korejština a angličtina.

2.1 Přímý překlad

Přímý překlad je nejstarší a zároveň nejjednodušší a nejlevnější metodou vyvinutou pro překlad textu počítačem. Tento překladač byl poprvé použit v roce 1954 pro překlad 60 ruských vět do angličtiny a ve své době vyvolal velké nadšení.⁴ Jde o překladač, který bývá obvykle jen jednosměrný a je založený na bilinguálním slovníku a základní morfologické analýze, podle které nahrazuje ve větě zdrojového jazyka jednotlivá slova za slova cílového jazyka a za použití jednoduchých gramatických pravidel přeskupuje pořadí slov ve větě, aby sedělo větnému pořádku cílového jazyka. Jeho součástí však obvykle není hloubková analýza zdrojového jazyka, nepracuje s lingvistickými informacemi a nedokáže tedy správně rozlišovat nejednoznačné výrazy a

³ PARK, Se-Young; OH, Gil-Rok. Machine translation in Korea.7. *MT Summit*, 1999, 100-104.

⁴HUTCHINS, John. The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954. Dostupné z: [http://sts.bdtf.hu/btk/flli/romanisztika/OKTATSARS%20DOCENDI/TANANYAGOK%20\(OKTATÓ%20SZERINT\)/ANTONIO%20SCIACOVELLI/TRADUTTORI%20INTERPRETI/corrispondenza%20commerciale/traduzione%20elettronica_articolo_english.pdf](http://sts.bdtf.hu/btk/flli/romanisztika/OKTATSARS%20DOCENDI/TANANYAGOK%20(OKTATÓ%20SZERINT)/ANTONIO%20SCIACOVELLI/TRADUTTORI%20INTERPRETI/corrispondenza%20commerciale/traduzione%20elettronica_articolo_english.pdf)

složitějšími konstrukcemi. Tento jednoduchý systém může být částečně úspěšný pro jazykovou dvojici, která si je velmi blízká a existuje mezi nimi málo gramatických rozdílů, zatímco u dvojice jazyků, která si je vzdálená, je tato metoda naprosto nedostatečná. Co se týče této korejštiny a angličtiny, musíme brát tedy v potaz jejich odlišnosti. Zatímco angličtina patří do indoevropské jazykové rodiny, korejštinu někteří řadí mezi altajské jazyky, někteří ji považují za jazyk izolovaný, a je tedy zřejmé, že mezi tímto párem existuje mnoho odlišností týkajících se slovní zásoby i struktury věty.⁵ O něco úspěšnější může být vzhledem k slovní zásobě z čínštiny a aglutinačnímu charakteru jazyka například v případě korejsko-japonského překladu. Zároveň nelze u takového překladače z kapacitních důvodů předpokládat, že by si poradil s texty z více odvětví, kde by narážel na nezvyklé větné konstrukce a slovní zásobu.

Překlady provedené tímto stylem pochopitelně nedosahovaly nijak vysoké kvality a kromě demonstrací schopností počítačů nenašly mnoho uplatnění, a proto se od 60. let 20. století začalo pracovat na složitějším systému zahrnujícím práci lingvistů.

2.2 Strojový překlad založený na pravidlech (Rule-Based MT)

Tento typ překladu, jak už název napovídá, využívá na rozdíl od přímého překladu morfologická, syntaktická, případně sémantická pravidla pro analýzu zdrojového jazyka i syntézu textu cílového jazyka. Pravidlový překladač můžeme dělit na dva hlavní směry, a to transferový překlad a interlingvu. Vývojáři začali brát v potaz lingvistické přístupy, jako je Hallidayova funkční gramatika nebo zejména pak formální gramatika, která je vhodná pro zpracování počítačem. Co se týče formální gramatiky, jde o soubor lingvistických objektů, tvrzení, pravidel a principů, která se dají zapsat matematicky tak, aby s nimi mohl pracovat počítač⁶. Mezi formální gramatiky, které měly vliv na vývoj strojového překladu, patří zejména transformační-generativní gramatika, se kterou přišel Noam Chomský v 60. letech. V souvislosti s teorií Chomského se objevují výrazy jako hloubková a povrchová reprezentace a transformační pravidla. Pro jazyky,

⁵ HUTCHINS, W. John, The first decades of machine translation. *Early years in machine translation*, John Benjamins Publishing, 2000. ISBN 9027283710

⁶ FORMÁLNÍ GRAMATIKA. CzechEncy – Nový encyklopedický slovník češtiny [online]. Brno: Vladimír Petkevič, Karel Oliva, 2017 [cit. 2019-05-19]. Dostupné z: <https://www.czechency.org/slovník/FORMÁLNÍ%20GRAMATIKA>

jako je korejština nebo turečtina, bývá uplatňován přístup Ch. J. Fillmora, který v 60. letech představil svou pádovou gramatiku a později rámcovou sémantiku.⁷⁸

Dobře sestavený pravidlový překladač by si měl na rozdíl od přímého překladače poradit i s dvojicemi velmi odlišných jazyků, jako je korejština s angličtinou. Všechn vložený text musí nejprve projít procesem morfologické analýzy, během které se provádí tzv. tokenizace, při které se rozdělí jednotlivá slova, interpunkční znaménka atd. V korejštině, která je aglutinačním jazykem, je nutné v tomto procesu oddělit plnovýznamové morfémy od gramatické části, sufixů plnících nějakou funkci, jako je třeba nominativní partikule nŭn (는), honorifikační sufixy kkäsŏ (께서), nebo třeba koncovky interogativu. Tato část je díky omezenému a jasně oddělenému počtu morfémů o něco jednodušší než třeba u jazyků flektivních, přestože i v korejštině můžeme narazit na slova, u kterých dochází ke skloňování a změnám ve kmenu. Jako příklad může posloužit například sloveso čitta (짓다)⁹, které se při použití ve větě v minulém čase vyskytuje ve tvaru čičössŭmnida (지었습니다). Pokud by překladač oddělil sŭmnida (습니다) jako sufix značící úroveň zdvořilosti a minulého sufixu öss (었), tak by bez použití rozšířených pravidel v této části vytvořil slovo čida (지다), čímž by celá věta ztratila původní smysl.

Na souborech lingvistických pravidel je sestaven tzv. parser, který vložený text zdrojového jazyka zanalyzuje ze syntaktického hlediska a vytvoří z textu obvykle syntaktický strom, pomocí kterého abstraktně vyjádří vztahy mezi jednotlivými slovy. Takový strom může vypadat například jako na následujícím obrázku, kde je stromová reprezentace věty Kim Tā-džungŭn 1988 njŏn kukhŏŭwŏne tangsŏntöötgo, 1991 njŏn tchonggapjadangin mindžutangŭl čchangdanghajötta. (김대중은 1988 년 국회의원에 당선되었고, 1991 년 통합야당인 민주당을 창당하였다).¹⁰¹¹

⁷ CHÉRAGUI, Mohamed Amine. Theoretical overview of machine translation. Proceedings ICWIT, 2012, 160.

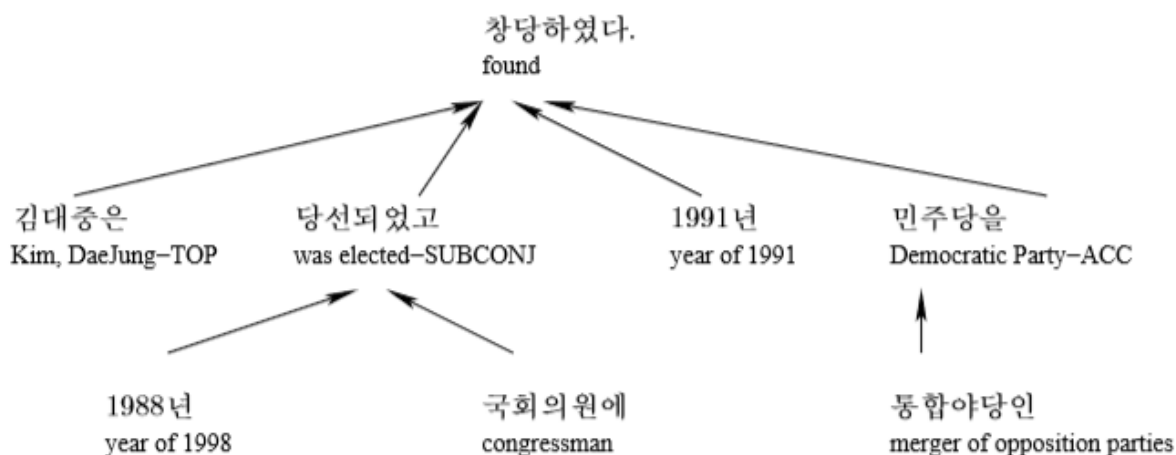
⁸ ČCHŎ Hočchŏl (et al.). Kigje pŏnjŏk_l ŭhan hangugŏ nonhang čhegje jŏngu. [Studie korejského argumentačního systému pro strojový překlad] Hangugŏ _imihak 3, 1998. Dostupné z: http://www.aistudy.co.kr/paper/pdf/korean_hong.pdf

⁹ stavět

¹⁰ Kim Te-džung byl v roce 1988 zvolen členem Národního shromáždění a v roce 1991 založil Demokratickou stranu spojením opozičních stran

¹¹ CHUNG Hoojung. Statistical Korean Dependency Parsing Model based on the Surface Contextual Information. Disertační práce. Korea University. 2004. Dostupné z:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.492.8537&rep=rep1&type=pdf>.



Pro vznik je nutné, aby slova byla nejprve „otaggovaná“, takže každé slovo by mělo mít značku, jestli jde o sloveso, jméno či jiný druh a následně se vyjádřily jejich vzájemné vztahy. Při překladu je přitom použit slovník, ve kterém jsou záznamy, které mají označení vyjadřující, jestli se jedná například o substantivum nebo sloveso a substantivum (či jiný slovní druh) pak dostane podle sémantického rámce přidělené ekvivalentní substantivum v druhém jazyce. Právě správné určení pádů hraje velký význam v analýze jazyka pro strojový překlad. V angličtině se dá určit významová role díky předložkám a pořádku slov ve větě, jelikož má poměrně pevný slovosled, kterým je subjekt – predikát – objekt. V korejské větě je však slovosled na první pohled značně volnější, což umožňují pádové partikule. Přesto se i zde se dají najít jistá pravidla, která se dají ve strojovém překladu využít. Prvním z nich je, že se predikát vždy vyskytuje až na konci věty. Dalším je, že rozvíjející větné vždy členy předchází slova, která rozvíjejí. Slova, na která je kladen důraz, bývají kladena blíže ke slovesu.¹²¹³Negace, která může být dalším problematickým jevem, se obvykle vyjadřuje na konci slovesa, zápornou *an* (안) nebo *mot* (못), případně může být vyjádřena

¹²Dennis Woojun Yang. Korean Language Generation in an Interlingua-based Speech Translation System. Massachusetts, 1996. Dostupné z

https://www.researchgate.net/publication/37996660_Korean_language_generation_in_an_interlingua-based_speech_translation_system. Diplomová práce. Massachusetts Institute of Technology.

¹³ Yang ve své práci zmiňuje příklad věty „Nanün onül 2 sie haksānghökwanesö čchinguwa čömsimül

mögöta.“ (나는 오늘 2 시에 학생회관에서 친구와 점심을 먹었다.) (Dnes ve dvě hodiny jsem ve studentské

budově s kamarádem jedl oběd.) V této větě je informací, na kterou klade mluvčí důraz, oběd. Přeskládáním slovosledu můžeme snadno zdůraznit, že mluvčí jedl oběd ve dvě hodiny, nebo že jedl oběd s kamarádem.

V angličtině, která nemá pádové partikule, slovosled tak snadno přeskládat nelze.

samotným slovesem *ǒbta* (없다). A nakonec jmenné fráze jsou následovány partikulemi *nŭn, ka* (는, 가).

Co se týče korejských postpozic, mohou stát za jménem, adverbium, a některá se mohou skládat za sebe. Můžeme je rozdělit do tří kategorií, a to spojovací koncovky, doplňující postpozice, a zejména postpozice určující roli slova, pádové partikule, což je část typická část modelu korejského jazyka, a který je pro jazyky, jako je angličtina, nepoužitelný. Vyskytují se za jmény, jmennými frázemi a gerundia vytvořeného ze slovesa. Jejich funkcí je označování subjektu pomocí *i/ga, ũn/nŭn, kkesǒ, esǒ, sǒ* (이/가, 은/는 께서, 에서, 서), přímého objektu *ŭl/rŭl* (을/를), nepřímého objektu *i/ga* (이/가), přivlastňovacího pádů *ŭi* (의), a hrají roli i v určování dalších rolí, jako je direktiv nebo instrumentál *e, ege, hantche, ŭro* (에, 에게, 한테, 으로), nebo také zvolání *a/ja* (아/야). K partikulím pak někteří řadí i sponu *ida* (이다).

Obtíže s pádovými partikulemi mohou vyplývat z toho, že korejské partikule mohou mít více významů (například partikule *e* (에), která označuje směr, místo, čas nebo příčinu, zatímco angličtina má pro každé z těchto označení jinou prepozici, pro směr *to*, místo *in*, čas *at* a příčinu *because of*). Krom toho se samozřejmě může překladáč setkat s případy, kdy je některá partikule vynechaná, nebo kde se naopak může vyskytovat více slov označených nominativní nebo akuzativní partikulí v jedné větě. Příkladem může být například věta *čchingu kchoga kchŭda* (친구가 코가 크다)¹⁴, kde se dvakrát po sobě vyskytuje partikule *가*, čímž překladáč ztrácí možnost řídit se čistě pravidlem partikule nominativu = nominativ a může docházet k nesmyslným překladům.¹⁵

Dalším problémem typickým pro korejštinu jsou kategoriální slovesa *hada* či *tōda* (하다 či 되다),¹⁶ kterým předcházejí nomina většinou s původem v čínštině, a která tvoří nezanedbatelnou část korejské slovní zásoby. Nezbytnou součástí parsování je totiž taggování, což je přiřazování

¹⁴ Kamarád má velký nos

¹⁵ PĀ Či-jŏn. Hanjŏng kigjepŏnjŏgŭi kjŏk silhjŏn jŏngu. [Studie o realizaci pádů v korejsko-anglickém strojovém překladu.] Sedžong Dáhakkjo. 2010. 12.

¹⁶ Dělat či stát se

jednotlivým slovům a částem věty označení, o jaké slovní druhy a kategorie se jedná, aby mohl překladatel vytvořit stromové grafy vyjadřující vztahy mezi slovy. Gramatika korejského jazyka značně zlehčuje rozeznání slovesa podle typických zdvořilostních a dalších koncovek. U sino-korejských slov ovšem nastává následující problém – jak překladatel rozezná, kdy jméno předcházející *hada* či *töda* (하다 nebo 되다) je součástí slovesa. Můžeme se setkat rovnou s několika příklady, jak může toto spojení vypadat:

- a) NP 하다
- b) NP + 하다
- c) NP 를 하다
- d) NP 를 NP 를 하다

Jména v tomto případě nejsou přímým objektem přestože mohou obsahovat pádovou partikuli 를. Zatímco v případě a) překladatel pravděpodobně přeloží sloveso správně, v případě b) hrozí, že překladatel vyhodnotí jméno a sloveso jako samostatná slova a vznikne díky tomu překlad, který může nepřirozeně. Podobný problém nastává v případě c) s tím rozdílem, že překladatel označí jméno jako akuzativ, a případ d) už se stává nepřeložitelným. Tyto situace si můžeme představit, pokud za NP¹⁷ doplníme například slovo 분석 (*analýza*)

Korejština	Překlad do angličtiny
Punsökhada (분석하다)	analyze
Punsök hada(분석 하다)	do analysis
Punsökül hada (분석을 하다)	do the analysis
Önörül punsökül hada (언어를 분석을 하다)	do the language the analysis

¹⁷ NP zde označuje jméno

Ve své práci *Light Verbs and Sino-Korean Verbal Nouns in Korean-English Machine Translation*¹⁸ jsou navrženy dva způsoby řešení: první je označit a pracovat s těmito konstrukcemi jako s idiomy, přičemž pádová partikule *by* byla označena jako proměnná, která tam může a nemusí být. Problémem by bylo ovšem vytvořit seznam osahující všechny možné případy a dalo by se předpokládat, že by se tak velké množství dat projevilo i na funkčnosti a rychlosti překladače. Dalším řešením by mohlo být specifikovat u jmen ve slovníku podkategorii, která by připouštěla možnost, že na sebe jméno může brát vlastnosti slovesa pod podmínkou výskytu jednoho s kategoriálních sloves.

Další typickou problematickou částí pro překladač je šest zdvořilostních stupňů, pro které nemá angličtina ekvivalentní vyjádření. Pokud jde tedy o překlad směrem z korejštiny do angličtiny, je možné do jisté míry vyjádření zdvořilosti ignorovat. O něco horší je pak ovšem překlad opačným směrem. Jednou z možností je držet se neutralizace zdvořilostního stupně typické pro psaný projev, ovšem i tak se mohou v překladu vyskytovat příklady, kdy to není možné obejít.

Neméně podstatným problémem pro strojový překlad může být fakt, že v korejštině na rozdíl od angličtiny existují gramaticky správné větné konstrukce bez vyjádřeného podmětu. Pokud takovou větu překládáme do angličtiny, musíme nutně podmět vyplývající z kontextu do věty dosadit, aby nebyla gramaticky chybná, a aby k takovým překladům nedocházelo, je nutné opět přidat další pravidla pro „obnovení argumentů“, kterými mohou být do anglické věty přidána např. zájmena. Přesto nejde o ideální řešení a je to jedna z částí překladu, která nabízí mnoho prostoru pro chyby.¹⁹

Vícevýznamová slova, která se vyskytují snad ve všech jazycích, jsou dalším problémem pro překlad počítačem. Příkladem může být v angličtině například slovo *book*, které může označovat podstatné jméno *kniha* a zároveň sloveso *rezervovat*, nebo třeba *bank* – *banka* nebo *břeh*. V korejštině se často vyskytují slova s více významy v jedné kategorii, například sloveso *ssŭda* (쓰다), které může znamenat *psát*, *používat*, *nosit*, *být hořký*... Bez toho, že by přístroj chápal kontext věty je opět velká pravděpodobnost, že překladač nevybere správný protějšek v cílovém jazyku.

¹⁸ RYU, B. R., & Kim, Y. K. *Light Verbs and Sino-Korean Verbal Nouns in Korean-English Machine Translation*. Taejon, 1999. Chungnam National University, Electronics and Telecommunications Research Institute (ETRI).

¹⁹ HAN, C. H., LAVOIE, B., PALMER, M., RAMBOW, O., KITTREDGE,... & KIM, M. Handling structural divergences and recovering dropped arguments in a Korean/English machine translation system. Conference of the Association for Machine Translation in the Americas. Springer, Berlin, Heidelberg. 2000, 10.

Z výše uvedených příkladů je zřejmé, že ani sadou složitých lingvistických pravidel a jazykových modelů není snadné podchytit všechna možná úskalí při převodu jednoho jazyka do jiného naprosto odlišného cílového jazyka. Při anglicko-korejském překladu se v této fázi musí brát v potaz například fakt, že korejština nepoužívá pomocná slovesa pro vyjádření budoucnosti nýbrž tvary slovesa, zatímco angličtina může vyjádřit budoucnost například slovem *will* a další odlišnosti. Po rozboru zdrojového jazyka je tedy nutné zvolit ještě správný přístup, jak informace převést do jazyka cílového. Quah²⁰ do tohoto typu strojového překladu řadí dva přístupy: transferový překlad (transfer-based) a interlingvu.

2.2.1 Transferový překlad

Jako u jiných pravidlových překladačů je zde prvním krokem je tokenizace, morfologická analýza a tagování. Pro tento typ systému je zásadní to, že probíhá na rozdíl od přímého překladu ve třech krocích, přičemž prvním z nich je převedení textu zdrojového jazyka do abstraktní reprezentace, která nese charakteristiky zdrojového jazyka a obvykle ji představují syntaktické stromové grafy, druhým krokem je převedení této reprezentace do ekvivalentní reprezentace cílového jazyka, a posledním krokem je vytvoření překladu do cílového jazyka na základě této reprezentace. Právě abstraktní reprezentace v tomto systému pomáhá překonat rozdíly mezi jazyky. Na rozdíl od interlingvy abstraktní reprezentace mezi jazyky není univerzální a musí se přizpůsobovat každému jazyku zvlášť.²¹

2.2.2 Interlingva

Tento přístup vychází z idey umělého jazyku interlingvy²², která by měla fungovat jako nezávislý univerzální jazyk, který tvoří prostředníka mezi ostatními jazyky. Překlad v tomto případě probíhá ve dvou fázích, přičemž první z nich je zpracování a převedení textu zdrojového jazyka do interlingvy, a druhá je generování textu z interlingvy do cílového jazyka. Na rozdíl od transferového překladu se tudíž reprezentace při překladu snaží pracovat s tím, co mají všechny jazyky společné. Výhodou tohoto přístupu je to, že je použitelný pro více jazyků, a v případě přidání dalšího jazyka do systému není nutné měnit jeho další části, ale stačí dodat funkci pro

²⁰ QUAH, C. K. Translation and technology. New York: Palgrave Macmillan, 2006. ISBN 14-039-1832-5.

²¹ NOONE, Grace. Machine Translation A Transfer Approach. Computer Science, Linguistics and a Language (CSLL) Department, University of Dublin, Trinity College, 2003.

²² Mezi příklady pokusů o interlinguu můžeme řadit například esperanto nebo ido, což jsou umělé jazyky vytvořené převážně na základě západoevropských jazyků

analýzu a generování nového jazyka, což šetří čas i peníze. Problémem je však vytvoření samotné univerzální interlingvy, což mnozí považují za nemožné.²³

2.2.3 Výhody a nevýhody překladač založeného na pravidlech

Lingvistika a hloubková analýza vět zdrojového jazyka je hlavní rozdíl od přímého překladač. Je zřejmé, že takový překladač dokáže mnohem lépe podchytit rozdíly odlišných jazyků, stále a je zde vyšší pravděpodobnost, že pokud dostane gramaticky správnou větu ve zdrojovém jazyce, dokáže vytvořit gramaticky správnou větu i v jazyce cílovém. Pokud se v překladač vyskytne chyba, dá se snadno zjistit, proč a jakým způsobem vznikla a následně odstranit. Mezi nevýhody tohoto překladače však patří fakt, že vytvoření takového překladače vyžaduje odbornou lidskou sílu, finance a čas a je náročné je taky jeho další rozšiřování a údržba. Navíc i velmi kvalitní překladač bude mít problém vytvořit kvalitní překlad, pokud je zadaný text zdrojového jazyka negramatický. Ačkoliv je možné zařadit další části programu, jako je například kontrola pravopisu, je to další zátěž pro počítač, který má data zpracovat, krom toho, že poskytuje další prostor pro chyby, které by mohly zapříčinit opravu správného textu na nesprávný, případně nechat špatný text neopravený. Další problematickou částí je tvoření delšího plynulého textu, jelikož řeší konstrukce zvlášť a nebere v potaz kontext celého textu. Jak jde vidět v předchozích příkladech, množství problémů a odlišností, se kterými se program musí vypořádat, je obrovský a jen stěží se dá podchytit každá možná varianta.. Navíc i takový překladač by se musel vypořádat například se slovy, které mají více významů, případně s odlišnostmi jako je vypuštěný podmět. Proto se dá předpokládat, že pravidlově založený překladač by mohl mít spíše úspěch v konkrétní oblasti, kde je malá pravděpodobnost, že budou výrazy použity v jiném kontextu než s jakým se překladač setká.

Příkladem využití úspěšného pravidlového překladače může být Météo System, kanadský překladač pro překlad meteorologických zpráv vytvořený v roce 1977, který překládá dvojici francouština-angličtina. Výhodou je, že termíny, se kterými se setkává, patří k jednomu kontextu – počasí. Pokud dostane nejednoznačnou větu na zpracování, systém tento text odmítne a pošle konkrétní část textu lidskému překladateli. Díky tomu byl překladač poměrně spolehlivý a ušetřil tak velké množství času a práce.²⁴ Nicméně ani takový překladač není pro běžného uživatele

²³ DORR, Bonnie Jean. Machine translation: a view from the lexicon. Cambridge, Mass.: MIT Press, c1993. Artificial intelligence (Cambridge, Mass.). ISBN 0262041383. Dostupné z: <https://www.aclweb.org/anthology/J94-4009>

²⁴ Viz 21.

dostatečný, a proto se výzkum vydal v 80. letech novým směrem, který do překladu zapojuje algoritmy a statistiku.

2.3 Statistický překlad

80. léta jsou dobou, kdy se v jižní Koreji začaly řešit technické problémy počítačového zpracování jazyka, jako je písmo, jeho fonty atd.²⁵ Výzkumu strojového překladu se ujalo Japonsko, a do konce 70. let představovat hlavní proud transferový překlad a interlingva. S příchodem 80. let se však výzkum, který tou dobou vedla především firma IBM a Japonsko, začal stáčet směrem k metodou překladu založeném na statistice pomocí strojového učení. Tato změna souvisí mimo jiné se samotným vývojem počítačů – výkonnější počítač dokázal rychleji zpracovat složitější algoritmy, které jsou potřeba ve statistických metodách. Pravidlový překlad potřebuje pro správnou funkčnost slovníky, ve kterých jsou dodány informace o typech jednotlivých slov, o jejich funkci ve větách a o jejich možném chování. Oproti tomu statistický překladač potřebuje především bilinguální i vícejazykové korpusy, ze kterých sám dostane informace nutné k překladu. Zjednodušeně řečeno když dostane takový překladač text, podle dat, které měl k dispozici rozhodne, jaký text cílového jazyka se rovná textu jazyka zdrojového, ať už jde o jednotlivá slova nebo delší segmenty. V podstatě tímto způsobem získává z korpusu vlastní slovník tím, že zjišťuje, jaký překlad je ze všech možných překladů daného segmentu nejpravděpodobnější, čímž řeší nejen problém synonym, ale i homonym. Aby toho byl ale schopen, je zřejmé, že zde záleží především na kvalitě dat, ze kterých informace pro překlad využívá a na kterých se takový překladač trénuje.

Ideální korpus pro překladač by měl splňovat jistá kritéria, mezi které patří například žánrová vyváženost a dostatečné množství dat. První kritérium je nutné především v případě slovníků, které by měly sloužit široké veřejnosti. Pokud by byl překladač trénován například jen na textech týkajících se vojenství, pravděpodobně by se často setkával s výrazy, které nikdy nebyly součástí korpusu, ze kterého data získává, a nebyl by schopen je přeložit, nebo by je, jako slova, která mohou mít ve vojenství jiný význam než v běžní řeči, přeložil špatně. Velkost navíc kromě problému s neznámými výrazy hraje pro statistický překladač roli, protože stroj pravděpodobnost vypočítává právě z počtu výskytů, kdy byl výraz přeložen konkrétním způsobem a počtu výskytů daného výrazu v celém korpusu. Vytvořit kvalitní korpus přitom není snadná záležitost. Hlavním

²⁵ PARK, Se-Young; OH, Gil-Rok. Machine translation in Korea.7. MT Summit, 1999, 100-104.

zdrojem vícejazykových textů, ze kterých může být korpus vytvořen, je internet, ze kterého lze snadno získat velké množství dat v elektronické podobě.

Nevýhodou ovšem je, nejednotnost formátu a další technické problémy, ale především lze mezi takovými texty snadno nalézt volnější překlady, případně i překlady naprosto chybné, kterých by měl překladač získat pokud možno co nejméně. I když je možné na vytřídění takových textů využít opět algoritmy, které zmenší riziko výskytu špatných dat, přesto neexistuje způsob, jakým by bylo možné je zcela eliminovat, aniž by zároveň došlo ke ztrátě užitečných dat, a není v lidských silách veškeré dostupné texty zkontrolovat a opravit, nehledě na fakt, že se texty na internetu často aktualizují, což může být ovšem zároveň výhodou kvůli aktuálnímu slovník.²⁶

Neméně obtížným úkolem je navíc dodávání informací o gramatice, sémantice apod. Co se týče korejského jazyka, tvorbou korejských korpusů, které by byly vhodné pro výzkum a účely zpracování přirozeného jazyka (NLP), se začaly zabývat univerzity a společnosti věnující se NLP už v 80. letech. V roce 1998 pak vláda spustila tzv. 21st Century Sejong Project, což byl projekt, v jehož rámci byl vytvořen Korean National Corpus (KCN, Sejong Corpus)²⁷. Korpus obsahuje jak mluvenou tak psanou řeč, paralelní korpus skládající se z korejštiny a dalších jazyků, jako je angličtina a japonština, a korpus obsahující morfologicky označená slova. Krom toho obsahuje i informace o homonymech a věty, které prošly procesem parsování, o velikosti 150 tisíc slov,²⁸ což jsou všechno data, na kterých se překladač může učit překládat. Mimo tuto hlavní část pak korpusu pak KCN obsahuje i historické korpusy nebo také korpusy určitých regionů (Severní Korea, zahraničí).

Překladač jako takový můžeme pak rozložit na více komponentů, mezi které patří jazykový model cílového jazyka, který udává pravděpodobnost pro každou možnou sekvenci slov, komponent pro doplňování slov např. v případech, kdy je ve zdrojové větě nevyjádřený podmět, který musí být v cílovém jazyce vyjádřený což je časté právě v případě korejštiny – angličtiny, lexikální model, který řeší uspořádání slov a další. V každém komponentu hledá překladač segment s největší pravděpodobností a dohromady pak tyto komponenty vygenerují cílový překlad.

²⁶ BOJAR, Ondřej. Čeština a strojový překlad: strojový překlad našincům, našinci strojovému překladu. Praha: Ústav formální a aplikované lingvistiky, c2012. Studies in computational and theoretical linguistics. ISBN 978-80-904571-4-0.

²⁷ Korpus je dostupný z <http://www.sejong.or.kr/user/main.do>

²⁸ Kang, B. and Kim H. (2004). Sejong Korean Corpora in the Making. Paper presented at the International Conference on Language Resources and Evaluation. Retrieved February 2, 2015, from

Mezi významné modely statistického překladače patří zejména frázový překlad (phrase-based MT), překlad založený na slovech (word-based MT) a na syntaxi (syntax-based MT), kde je rozdílem především nejmenší překladová jednotka. Každý z nich má své vlastní nedostatky a výhody. Překladač, který bere jako základní překladovou jednotku slovo, byl z těchto modelů vyzkoušen jako první, nicméně nebyl příliš úspěšný, na rozdíl od frázového překladače. Ten už nepracuje se samostatnými slovy, ale s delšími úseky ve větě, tedy frázemi. Díky tomu je větší pravděpodobnost, že vezme v potaz správný kontext a nepřeloží homonymní slovo špatně. Zdrojový text je nejprve rozdělen na menší části, fráze, což může být jakákoliv sekvence slov bez ohledu na fráze v lingvistickém kontextu, a tyto jednotky jsou nezávisle přeloženy. Z každé fráze je potom odvozen soubor lexikálních jednotek a následně dojde ke jejich přeskládání do větného pořádku cílového jazyka, podle přednastavených jazykových modelů pro větný pořádek.

Třetí ze zmíněných modelů, překladač založený na syntaxi, využívá více odlišný přístup.²⁹ Jedním z hlavních nedostatků čistého statistického překladače je, že sice využívá tzv. jazykové n-gram modely³⁰, díky kterým je kontext omezen na n sousedících slov, což je žádané proto, že se překladač může pochopitelně setkat i se správnými větami, které se v korpusu nevyskytují, ale zároveň však může být pro takový model problematické, pokud se sématicky či syntakticky související slova nachází dále od sebe. Tyto problémy částečně řeší syntaktický model, který využívá větný rozbor překládaného textu, který místo toho, aby předpovídal následující slovo podle slov předešlých předpovídá možné slovo podle slova řídícího.

Je nutné také zmínit, že lingvistiku přitom statistický překladač jako takový úplně nezavrhuje, jak by se mohlo na první pohled stát. Podobně jako pravidlový překladač text parsuje a využívá syntaktické stromy, díky kterým zabraňuje tomu, aby v překladu vznikaly nesmyslné a negramatické věty skládáním za sebe slova, která spolu mohou často souviset. Zároveň může mít dodané další části, jako je spellcheck, který omezí nesmyslná slova, a další komponenty.³¹

²⁹ Ahmed, Amr, and Greg Hanneman. "Syntax-based statistical machine translation: A review." *Computational Linguistics* (2005).

³⁰ Sekvence po sobě jdoucích slov, případně značek

³¹ STYMME Sara, AHRENBERG Lars, *Using a Grammar Checker for Evaluation and Postprocessing of Statistical Machine Translation*. Linköping University, Švédsko, 2010. Dostupné z: https://www.researchgate.net/publication/220746465_Using_a_Grammar_Checker_for_Evaluation_and_Postprocessing_of_Statistical_Machine_Translation

2.3.1 Překlad založený na příkladech (Example-Based MT)

Překlad založený na příkladech je často zařazován do skupiny překladačů založených na datech, protože podobně jako statistický, potřebuje bilinguální korpus a slovník. Na první pohled je takový překladač mnohem jednodušší, než je překladač statistický, přesto ve skutečnosti může využívat jak metody ze statistiky, tak modely a pravidla z pravidlových překladačů. Bývá také označován jako překladač, pro který je základní ideou využití analogie. Zatímco statistický překladač se na bilinguálním textu učí a využívá i data jako jsou parametry a pravděpodobnosti, příkladový překladač využívá zejména samotný text. Tohle se může ve výsledku projevit tak, že zatímco statistický překladač může vytvořit jiný překlad textu než je ten který má k danému zdrojovému textu k dispozici, příkladový využije právě tohoto existujícího překladu a přeloží větu identicky. Jak již název napovídá, funguje na jednoduchém principu příkladových vět z bilinguálního korpusu, podle kterých tvoří výsledné překlady. Názornou ukázkou může být věta *Čönün sakwarül möggo siptcha*. (저는 사과를 먹고 싶다.)³² Překladač může najít v korpusu buď přesný příklad takovéto věty, kdy stačí použít tento existující překlad, nebo větu podobnou – například *Čönün päřül möggo siptcha* (저는 배를 먹고 싶다.)³³ V takovém případě si překladač může vzít jen známou část věty z příkladu, tedy „*Čönün möggo siptcha*“ (저는 먹고 싶다). a překlad pro jablko najít ve slovníku a doplnit jej do věty na místo, které obsadilo slovo hruška.

Stejně jako u statistických překladačů je hlavním problémem vytvoření dostatečně kvalitního korpusu, kterým se může řídit, a pokud se setká s neznámými typy souvětí, bude mít problém si s nimi poradit. Tento typ překladače byl vyvíjen podobně jako statistický také zejména v 80. letech 20. století. Na rozdíl od statistického se ovšem neujal se v takové míře, jako je statistický frázový překladač.³⁴

³² Chei sníst jablko

³³ Chei sníst hrušku

³⁴ Zatímco se tento typ překladače příliš nerozšířil, překladatelé v technických odvětvích často využívají pomocný strojový překlad, translation memory (TM), který funguje na velmi podobném principu. Výhodou může být například přidělování procent, na kolik je pravděpodobně string správně přeložen. 100% můžou být označené stringy, které se ve stejné podobě nachází mezi dříve přeloženými stringy a mohou být bez úprav přeloženy totožně. 101% pak označuje stringy, které byly nalezeny ve stejné podobě v paměti TM a zároveň je stejný i string předchozí a následující, což zvyšuje pravděpodobnost správného kontextu. Krom toho může být TM využita i jako součást strojového překladače.

2.3.2 Výhody a nevýhody překladače statistického překladače

Podobně jako pravidlový překladač i překladače využívající data z korpusu, jako je statistický nebo příkladová mají své nedokonalosti. Zřejmou výhodou oproti pravidlovému překladu může být větší využití kontextu, větší pravděpodobnost, že dokáže zpracovat i ne zcela správné věty, a že si může v korpusu dohledat i data, pomocí kterých dokáže přeložit i idiomy. Problematická část pro tento překlad však nastává ve chvíli, kdy se překladač učí překládat na nekvalitních datech. I do statistických překladačů bývá zakomponována lingvistika, zejména funkční generativní popis pro vytvoření paralelního treebanku, který obsahuje větné rozборы, na kterých se může překladač naučit parsovat – podobně jako u pravidlových překladačů. U frázových překladačů může nastat chyba, pokud je mezi výrazy, které jsou na sebe vázané, existuje velká vzdálenost, což je, jak ve své knize uvádí Bojar, také jeden ze způsobů, jak poznat kvalitní překladač. V korejštině se přitom takové věty vyskytují poměrně často, což je způsobeno odsouváním řídicího slovesa až na konec věty, přičemž se mezi řídicím jménem a slovesem můžou vyskytovat celá souvětí.

2.4 Neurální strojový překlad

Nejnovejší technologií, která ovládla začátek 20. století, jsou tzv. umělé neurální sítě. Přestože můžeme počátky vývoje neurálních sítí najít už v minulém století, kvůli technologické limitaci se začaly ve větším měřítku využívat až po roce 2010. Tato technologie má využití nejen ve strojovém překladu, ale také v mnoha dalších oblastech umělé inteligence, jako je vizuální rozpoznávání, rozpoznání řeči, ve hrách a uplatnění najde třeba i na sociálních sítích. Inspirace pro tento systém je živý mozek. Podobně jako v mozku si můžeme představit neuronové sítě jako systém složený ze zjednodušených umělých neuronů spojených „synapsemi“, které přenáší signál mezi neurony. Tyto umělé neurony, jejichž úkolem je „vytvářet výstupní signál na základě vnitřního potencionálu, který je určen sumou ohodnocených příchozích signálů a práhem. Výstupní signál je předán dalším neuronům, které jsou mezi sebou propojeny ohodnocenými synapsemi.“³⁵ Podobně jako se učí lidský nebo zvířecí mozek, se tato velmi zjednodušená napodobenina postupně sama učí a její schopnosti se tak se zpracovanými daty zlepšují. Využívá k tomu rekurentní neuronové sítě, díky kterým se učí i ze svých vlastních chyb. Oproti hojně užívanému statistickému strojovému překladu by se tak měl v ideálním případě, podobně jako člověk, brát v potaz kontext

³⁵ MYŠKA V.: REKURENTNÍ NEURONOVÉ SÍTĚ PRO KLASIFIKACI TEXTŮ, Brno 2018, dostupné z https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=171408

celé věty, zvládat i náročnější větné konstrukce a je teoreticky možné, aby se do určité míry naučil rozeznávat například i stupeň zdvořilosti, nebo alespoň udržoval konzistentní úroveň zdvořilosti v daném textu, která hrají v korejštině významnou roli.

Na rozdíl od statistického překladače, který využívá prediktivní algoritmy a potřebuje mít k dispozici bilinguální korpus, na základě kterého tyto algoritmy předvídají správný překlad, neurální sítě využívají hloubkové učení, může využívat jako učební data statistické modely, na kterých se sám naučí lingvistická pravidla, a dokáže se učit také ze svých vlastních chyb. Tréninková data by přitom měl opět ideálně tvořit co nejkvalitnější bilinguální korpus, ale je možné mu dodat i další data, jako jsou glosáře, lingvistická pravidla pro strukturu věty a podobně. Zásadní také je, že bere jako základní jednotku celou větu, což by mělo napomáhat dosažení lingvistické kvality i lepší práce s kontextem.

Překladač má na první pohled také jednodušší strukturu, než překladač statistický, který se skládá z mnoha samostatných částí. Charakteristické jsou pro něj tzv. vektorové reprezentace. Neobsahuje ovšem samostatné jazykové nebo lexikální modely. Patří k tzv. enkodér - dekodér modelům, kdy enkodér fáze zahrnuje analýzu, výsledkem které je vytvoření vektorů, které mají stanovenou určitou hodnotu určenou podle „embedding“ tabulky, která by měla zahrnovat informace, jako pozice ve slovníku, ale i sémantické vlastnosti a hodnotu kontextu předchozího a následujícího segmentu. U velmi podobných vektorů je pak pravděpodobné, že jsou to synonyma.

Dekodér fáze pak zahrnuje vygenerování textu v cílovém jazyce dekodováním dříve vytvořených vektorů, přičemž bere v potaz opět kontext předchozích a následujících segmentů již přeloženého textu. Kontext přitom zahrnuje díky již dříve zmíněným rekurentním neurálním sítím, což je specialita tohoto překladače. Dalo by se říct, že překlad slova probíhá v opakujících se cyklech, kdy jsou při každém cyklu brány v potaz předchozí sekvence a je tak upravována hodnota vektoru. Pro překlad je pak zejména hojně využívaný typ sítě LSTM (long short term memory). Tenhle druh rekurentních neurálních sítí využívá „brány“ vstupní, výstupní a zapomínací. Tyto brány umožňují zapomínání sekvencí, které mají opakovaně příliš malou pravděpodobnostní hodnotu, a zvyšování váhy u správných sekvencí. Dochází zde také k rozhodnutí, které termíny uložit do dlouhodobé paměti, a které do krátkodobé. Ve vstupní bráně je pak rozhodnuto, které nové informace budou využity v následujících krocích, a které budou vyřazeny. Výchozí brána pak rozhoduje, které hodnoty mohou být výstupem buňky. Výhodou tohoto typu sítě je, že zvládá brát v potaz i kontext souvisejících, ale vzdálených sekvencí.

Dalším využívaným typem je GRU, gated recurrent unit, která má na rozdíl od LSTM jen dvě brány, update a reset gate, přičemž první ze zmiňovaných funguje podobně jako zapomínací a vstupní brána a reset gate má funkci výstupní brány. Zjednodušením LSTM na GRU je pak dosaženo urychlení tréninku a je méně náročný na výkon počítače.

2.4.1 Výhody a nevýhody neurálního překladače

Jak statistický překladač, tak i neurální přispěli podle většiny zdrojů zlepšení překladů zvláště co se týče kontextu, plynulosti a přirozenosti překladu, ačkoliv mají stále daleko do překladů produkovaných člověkem.³⁶ Přestože je neurální strojový překladač nepochybně krokem vpřed, má také své nevýhody. Jednou z nich může být delší čas nutný k tomu, aby se překladač „vytrénoval“ a kvůli své složitosti i nárok na výkon stroje. Další nezpochybnitelnou překážkou může být nepředvídatelnost. Jelikož se překladač učí „sám“, pokud se v datech, ze kterých se učí, vyskytne chyba, je schopný nejen tuto chybu reprodukovat, ale také může ovlivnit i jiné překlady, které by byly správné, pokud by se překladač nenaučil pracovat právě s tou chybou. Když je taková chyba nalezena, těžko se pak dohledává, odkud vznikla, a ještě hůř se taková chyba opravuje. Příkladem jeho samostatnosti může být třeba Google překladač, který si poté, co Google přešel ze statistiky na neurální síť, vytvořil svou vlastní řeč pro překlad jazyků, pro který nebyl dostatečně vytrénován. V podstatě by se dalo říct, že si vytvořil svou vlastní interlingvu, aniž by se o to cíleně pokoušeli vývojáři, kterou začal používat například pro překlady mezi japonštinou a korejštinou, přestože by vyučen pro to překládat oba jazyky jen v páru s angličtinou.³⁷ Na druhou stranu je tento příklad dobrý i pro demonstraci výhod oproti statistickým modelům, které se bez správných vstupních dat nemohou mezi neznámými páry naučit tyto páry přímo překládat.

Jelikož se neurální překladač dokáže na příkladech naučit i další pravidla kromě syntaxe, jako je morfologie, dokáže si mnohem lépe poradit s různými tvary slov, i když v korpusu všechny možnosti, jak může dané slovo vypadat, nemusí vůbec mít. I díky tomu považuje většina zdrojů neurální strojový překlad za lepší z hlediska plynulosti, správného kontextu a dokáže si lépe poradit

³⁶ V roce 2017 byla uspořádána soutěž v překladech lidí proti umělé inteligenci. Přestože dostaly překladače Google, Naver a Systran mezi 10-15 body z 30 možných, lidské překlady dostali v průměru 25 bodů. Problémem se zdála být jak gramatická struktura, tak i kontext. Očekává se však, že vzhledem k povaze těchto překladačů, které využívají hloubkové učení, se časem tyto problémy vytráť.

Humans Defeat Machine Translation In Korean Competition. Professional Translation & Interpretation Services | ULG [online]. 2019 United Language Group. All Rights Reserved. [cit. 13.06.2019]. Dostupné z: <https://unitedlanguagegroup.com/blog/humans-defeat-machine-translation-korea-competition/>

³⁷ Zero-Shot Translation with Google's Multilingual Neural Machine Translation System [online]. Google, 2016. Dostupné z: <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>

s lexikální diversitou. Studie, jako je Neural versus Phrase-Based Machine Translation Quality: a Case Study³⁸ však ukazují, že i neurální překladač má stále prostor pro zlepšování například co se týče delších vět a uspořádání slov ve větě, což by mohlo být opět významným problémem právě v překladu korejštiny. To ukazuje ve své práci i Pak Ok-su³⁹, který porovnává chyby v překladech v závislosti na zdrojovém textu. Pak v práci uvádí například to, že pouhé rozdělení souvětí na dvě věty může změnit kvalitu překladu. Věnuje se i doplňování nevyjádřeného podmětu, což, jak z práce vyplývá, může také znamenat pro překladač podstatný problém. Další zařazené problémy v překladu z korejštiny do angličtiny jsou zde opět slova s více významy případně špatná interpretace. Krom toho i další výzkumy ukazují, že neurální síť může i přes všechny tréninky mít problém například s morfologií v určitých jazykových párech, jako je pár angličtina – lotyština.⁴⁰

Nejmodernější překladače dnes využívají především neurální sítě. Přestože jde o jeden typ překladače, který využívá podobné metody, záleží zde i na drobných rozdílech v algoritmech a zvláště na datech, na kterých je takový překladač trénován. V následující kapitole bych proto ráda porovнала moderní dostupné překladače Google Translate a Naver Papago v překladu z korejštiny do angličtiny na různých typech textu..

³⁸ BENTIVOGLI, Luisa, et al. Neural versus phrase-based machine translation quality: případová studie, 2016. [online] Dostupné z: <https://arxiv.org/pdf/1608.04631.pdf>

³⁹ PAK Ok-su. Hanjǒng kigjǽpǽnjǽgesǽ Stui jugjǽng tchǔkǽinge ttarǔn pǽnjǽk orju punsǽk. [Analýza chyb podle typologické charakteristiky zdrojového textu v korejsko-anglickém strojovém překladu.] The Society For Humanities Studies In East Asia, 2017. 12 [online] Dostupné z: http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07289490&language=ko_KR

⁴⁰ Humans Defeat Machine Translation In Korean Competition. Minneapolis [online]. 2019 United Language Group. All Rights Reserved. [cit. 13.06.2019]. Dostupné z: <https://unitedlanguagegroup.com/blog/humans-defeat-machine-translation-korea-competition/>

3. Srovnání překladačů Naver Papago a Google Translate

Angličtina dnes zastává pozici světového jazyka, a tudíž jsou překlady právě do a z angličtiny velmi žádané a právě automatizace překladu je dnes kvůli velkému množství překládaných textů téměř nutností. Strojový překlad ovšem využijí nejenom komerční subjekty, ale může být pomocníkem i pro běžného uživatele, který korejštinu neovládá, nebo využije překladač jako pomocníka pro urychlení práce. V předchozí kapitole byly představeny základní přístupy strojových překladačů v chronologickém pořadí, a také problémy, se kterými se tyto překladače potýkají při překladu korejštiny a angličtiny. Nejnovější technologie neurálních sítí spojuje principy předchozích modelů; za prvé se sama učí lingvistická pravidla, na kterých je založen pravidlový překladač, aniž by byla nutná větší interference člověka. Zároveň by měla umět překonat rozdíly mezi jazyky, jako jsou zmíněná sino-korejská slova, a poradit si s homonymy a mnohoznačnými výrazy díky využití okolního kontextu. Přitom využívá podobně jako statistický překlad již existující překlady jako tréninková data, a může tak využít i samotný existující statistický model, který by měl předčít schopností učit se a zapomínat. Je tedy zřejmé, že neurální překladače představují přelom ve schopnostech automatického překladu.

Jestli si ale skutečně tyto překladače dokáží poradit s výše zmiňovanými problémy a s jakými nedostatky se stále při překladu korejštiny do angličtiny umělá inteligence musí vypořádat bych chtěla zjistit v následující kapitole s využitím moderních veřejně dostupných překladačů. Mezi takové překladače, které jsou pro běžného zdarma dostupné, patří korejský Naver Papago a americký Google Translate. Přestože jsou oba překladače dnes založené na neurálních sítích, každý byl vytrénovaný jiným způsobem a na jiných datech, a proto jsou i vyprodukované překlady často odlišné.

V následující kapitole krátce představím oba zmíněné překladače. Poté nechám oba překladače přeložit vybrané druhy textu a rozeberu konkrétní chyby, na které je možné v těchto překladech narazit. Zaměřím se zejména na jejich schopnosti v rámci výběru správného překladu u vícevýznamových slov, určení rolí jednotlivých členů ve větě a překlad odborných a méně častých termínů.

3.1 Naver Papago

Naver Papago je veřejně dostupný překladač korejské společnosti Naver Corporation, který byl poprvé jako testovací verze zveřejněn v roce 2016 ve formě mobilní aplikace.⁴¹ Dnes již existuje i veřejně dostupná webová verze překladače dostupná na adrese <https://papago.naver.com/>. Překladač momentálně umí překládat mezi 15 jazyky zahrnujícími korejštinu, japonštinu, tradiční a zjednodušenou čínštinu a další asijské jazyky a pak také angličtinu, francouzštinu, španělštinu, ruštinu, němčinu, italštinu a portugalštinu. Krom toho obsahuje další užitečné funkce, jako je překlad celé webové stránky, která je momentálně dostupná pro korejštinu, angličtinu, japonštinu a čínštinu, a nebo také funkci *Honorific*, která při překladu z angličtiny do korejštiny umožňuje uživateli zvolit zdvořilostní úroveň v přeložené korejštině. Bez této funkce užívá Naver při překladech obvykle neutrální stupeň zdvořilosti, přičemž pokud funkci zapneme, začne překladač používat u sloves koncovku neformální zdvořilosti *jo* (요) a přizpůsobovat i další větné členy, jako např. zájmena, kdy výraz pro *ty* změní z *nŏ* (너) na *tangsin* 당신.⁴²

Anglická věta	I'm really happy to meet you.
Překlad bez funkce <i>honorific</i>	Nŏrŭl mannasŏ čŏngmal kippŭda.(너를 만나서 정말 기쁘다.)
Překlad s funkcí <i>honorific</i>	Tangsinŭl mannasŏ čŏngmal kippŏjo. (당신을 만나서 정말 기뻐요.)

Jak již bylo výše zmíněno výše, Naver Papago je založený na neurálních sítích, a je proto pro tento překladač zásadní mít kvalitní tréninková data, ale ideálně i zpětnou vazbu, díky které se může nadále zlepšovat. Naver využívá pomoci svých uživatelů aplikací GYM. V této aplikaci jsou registrovaným uživatelům nabízeny samostatné věty v korejštině a několik variant překladu těchto vět. Uživatel může pak sám dodat vlastní návrh, jak by věta měla být přeložena. Aplikace pak každý den vyhodnotí, který z překladů dostal od uživatelů nejvíce bodů.

Překladač korejské společnosti Naver Papago sice nenabízí překlad do tolika jazyků jako Google Translate, přesto se podle Naveru stala jejich mobilní verze oblíbenější mezi korejskými uživateli. Počet aktivních uživatelů za měsíc podle průzkumu App Annie vystoupal v březnu 2019

⁴¹ South Korea's Naver launches official version of AI translator 'Papago'. THE NATION, 2017 [online]. Dostupné z: http://www.nationmultimedia.com/detail/Startup_and_IT/30321483

⁴² Naver updates translation service for Korean honorifics. The Korea Herald, 2019 [online]. Dostupné z: <http://www.koreaherald.com/view.php?ud=20190117000606>

až k 5,65 milionům uživatelů, což je skoro o milion více než počet uživatelů Google Translate v Koreji. Důvodem je podle Naveru to, že dosahuje lepší lingvistické kvality překladů mezi korejštinou, japonštinou a čínštinou. V počtu aktivních uživatelů webové verze ovšem stále vede Google Translate.⁴³

3.2 Google Translate

Počátky dnes celosvětově nejpoužívanějšího překladače Google Translate dostupného na <https://translate.google.com/> sahají mnohem dál než historie korejského překladače, a to až do roku 2006. Překladač byl založený na statistickém systému společnosti SYSTRAN, a to až do roku 2007, kdy si Google vytvořil vlastní statistický model. Tento model nedokázal překládat přímo mezi jinými páry jazyků kromě páru angličtina – jiný jazyk, takže při překladu např. z japonštiny do korejštiny tento překladač text v systému přeložil nejprve do angličtiny a poté z angličtiny do korejštiny, což je sice jednodušší a finančně méně náročné, ovšem dává do další prostor pro to, aby se v procesu překladu ztrácely informace.

V roce 2016 pak Google oznámil, že začne postupně překladač převádět rovněž na neurální systém, a korejština byla mezi prvními 8 jazyky, kterých se tato změna týkala. Výhodou je, že neurální sítě zde můžou využít existujícího statistického modelu. Výhodou se stal tzv. Zero-Shot Translation, tedy schopnost překládat přímo mezi různými jazykovými dvojicemi bez nutnosti převádění textu do angličtiny.⁴⁴ Nespornou výhodou oproti překladači od Naveru je schopnost překládat do více než 100 jazyků. Funkce pro překládání webu není součástí překladače, ale byla integrována do prohlížeče Google Chrome. Navíc nabízí možnost přeložit celé dokumenty po nahrání souboru do překladače. Z pohledu korejštiny je naopak nevýhodou Googlu fakt, že neumožňuje uživateli ani základní volbu zdvořilosti, jako funkce *Honorific* od Naveru, a z pohledu zdvořilosti tak text z angličtiny Google překládá značně nekonzistentně.

Podobně jako Naver má Google zájem na tom získávat zpětnou vazbu pro trénink překladače a to uživatelům umožňuje pomocí aplikace *Komunita*, kde mohou uživatelé buď hodnotit správnost překladů nebo mohou sami překládat slova a fráze či celé věty.

⁴³ Naver's Papago more popular than Google Translate among Koreans. The Korea Herald, 2019 [online]. Dostupné z: <http://www.koreaherald.com/view.php?ud=20190423000792>

⁴⁴ Google Translate: How does the search giant's multilingual interpreter actually work?. Independent [online]. Londýn: Independent, 2018 [cit. 2019-07-01]. Dostupné z: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-translate-how-work-foreign-languages-interpreter-app-search-engine-a8406131.html>

3.3 Porovnání schopností překladačů

Ačkoliv oba zmíněné překladače využívají neurálních sítí, musí mezi nimi nutně existovat rozdíly, což je dané nejen vlastním nastavením překladače, ale také daty, na kterých byly překladače trénovány, zpětnou vazbou uživatelů a všechny další informace, které byly překladačům předány. Jako důkaz těchto rozdílů nám postačí porovnání překladu výrazů s více významy, tedy homonymy, která najdeme ve všech jazycích včetně korejštiny. Zatímco člověk obvykle správný význam souzvučného výrazu pochopí díky kontextu, počítač sám o sobě významům rozumět nemůže a proto má se správným výběrem významu problém. Výjimkou přitom stále ještě nejsou ani neurální strojové překladače. Jak tyto chyby mohou vypadat si můžeme ukázat na následujícím příkladu s korejským slovem *pä* (배), které může znamenat hruška, žaludek, břicho, loď a další významy.

Pokud tento výraz zadáme samostatně do překladačů, které v tomto případě nemají žádný kontext, oba překladače ho přeloží jako *ship* (loď). Google Translate v tomto případě pak nabídne další možné varianty překladu slova *pe* seřazené podle četnosti, tedy jak často se tento překlad vyskytuje ve dostupných dokumentech (viz obr. 2).

Chybu v překladu můžeme najít již v jednoduché větě “Marie snědla hrušku”.

Korejská věta	Marianün pärül mögötta, 마리아는 배를 먹었다.
Anglický překlad	Maria ate a pear.
Naver Papago	Maria ate a pear.
Google Translate	Mary ate the ship.

Rekurentní neuronové sítě by měly být schopné využít okolních slov k určení správného překladu i v případě, že se ve větě vyskytnou homonyma. Jak však můžeme vidět, zatímco Naver bez problémů změnil překlad slova na hrušku, Google překvapivě nedokázal rozpoznat, že slovo jíst nedává ve spojení s argumentem loď smysl, a slovo hruška přeložil jako loď.

Ani Naver se ovšem těmto chybám nevyhne, jak lze vidět na příkladu věty „Marie si sedla a snědla hrušku.“.

Korejská věta	Marianŭn andžasŏ pārŭl mögŏtta,
	마리아는 앉아서 배를 먹었다.
Anglický překlad	Maria sat down and ate a pear.
Naver Papago	Maria sat down and ate her stomach.
Google Translate	Mary sat down and ate the ship.

Zatímco Google se zde stále drží výrazu loď, překladač Naveru zde přidal navíc přivlastňovací zájmeno její v kombinaci s argumentem žaludek. Naopak správné výsledky od obou překladačů získáme, pokud do věty vložíme navíc slovo *chutný*.

Korejská věta	Marianŭn andžasŏ masinŭn pārŭl mögŏtta,
	마리아는 앉아서 맛있는 배를 먹었다.
Anglický překlad	Maria sat down and ate a delicious pear.
Naver Papago	Maria sat down and ate a delicious pear.
Google Translate	Mary sat down and ate a delicious pear.

Na těchto příkladech lze vidět, že překladače skutečně pracují při překladu se zbytkem věty, ačkoliv v těchto příkladech vykazují značné nedostatky, kdy nedokáží s jistotou určit správný význam homonyma podle slovesa a vyžaduje více kontextu. V této kapitole proto nechám představené překladače přeložit několik druhů textů, z korejštiny do angličtiny, které již existují v obou jazycích nebo jsou doplněné vlastním překladem. Mezi tyto texty zařadím publicistiku, komerční stránky, korejskou ústavu, běžnou řeč, korejskou ústavu. Vycházím z předpokladu, že díky různorodosti textů se překladače setkají s více variantami větné struktury a s rozmanitější slovní zásobou. Zaměřím se na chyby, které v podaných textech překladače produkují a porovnáám v tomhle ohledu překladač Naver Papago a Google Translate.

3.3.1 Publicistický text

První z typu textů, který nechám překladače zpracovat, je publicistický styl. Vybraný článek *Strengthen the alliance* pochází z webu Korea Joongang Daily⁴⁵, který obsahuje bilinguální sloupek. Bohužel překlad článků nemůžeme považovat za bilinguální zarovnaný text,

⁴⁵ <http://koreajoongangdaily.joins.com/news/article/Article.aspx?aid=3064889>

jelikož se některé informace v korejském textu v anglické verzi nevyskytují. Nabízený existující překlad na stránce proto nutné považovat jen za referenční a bylo nutné pro potřeby této práce vytvořit překlad vlastní.⁴⁶ Z článku o velikosti zhuba 500 slov jsem pak do práce z důvodů dodržení rozsahu práce zařadila první odstavec korejského textu, ve kterém jsou zřetelně vidět některé nedostatky obou překladáčů.

- Korejský text:

Mun Dže-in tätongrjõngi ilbon osakkaesõ jõllin 20 kæguk(G20) cõngsanghõüi ilcõngül mamurlhagu 29 il kügukhätta. Ibõn hõüinün oldürõ pihäkhwa hjõpsangesõ milljõna ittõn tähanmingugi jõlgangdülkwaüi cõkdžõp tähwarül tchonghä õggorjõgül hõbokhal kihõro jõgjõdzjõtta. Kürõna kjõlkwarül pomjõn sõngkwanün mimigä poinda. Usõn kãchõgugin ilbonkwaüi cõsanghõdambutchõ jõllidži mothätta. Tadžahõüi cõhamsõkõcha ilbon ttangül palkün uri täтчongrjõngi ilbon cõchongriwa jangdža hõdamül katdži mothan kõn jurjega obtta. Han · il kwangjega õlmana simgakhan sudžunüro cõchurakhänündži tandžõküro pojõcunda. Panmjõn Abe Shinzõ ilbon cõchongrinün Donald Trump miguk täтчongrjõngül pirothä jak 20 kã nara cõngsangina kukdžekigu sudžangkwa jangdža hõdamül kačjõtta. Mun täтчongrjõngi 7 cõcharje jangdža hõdamül kadžin kõtkwa täbitõnda. Kedaga G20 hõüimada jõlligon hättõn han · mi · il cõngtanghõüido musantwättko mi · il · indo 3ça cõngsanghõüga kü çarirül däsinhätta. Mi · il cõngsanghõdam tü kongdongsõngmjõngesõdo `han · mi · il kongdžo `ran pchjohjõni ppadzjõtta. Tongasia cõngcõhiesõ pjõnpangüro milljõnan hanguküi hjõndžusoga cõnnarahage türõnan kõsida.⁴⁷

⁴⁶ Všechny vlastní překlady byly přeloženy samostatně a následně opraveny korejskou rodilou mluvčí Da-ä Ho

⁴⁷ Prezident Mun Če-in se 29. vrátil domů po ukončení summitu Skupiny 20 (G20) v Osace, Japonsku. Toto setkání bylo považováno za příležitost díky přímému dialogu se světovými mocnostmi obnovit diplomatické schonosti Korejské republiky, která byla vyřazena z letošních vyjednávání o denuklearizaci. Avšak když se podíváme na výsledky, úspěchy se zdají být zanedbatelné. Zaprvé prezident neměl příležitost vést dvoustranný dialog s hostujícím Japonskem. Je nebývalé, aby jihokorejský prezident, který cestoval do Japonska, aby navštívil vícestranná setkání, nevěděl dvoustranný dialog s japonským premiérem. Ukazuje to, v jak vážném stavu jsou korejsko-japonské vztahy. Premiér Abe Šinzó se zúčastnil 20 summitů s lidry cizích zemí a představiteli mezinárodních organizací, včetně prezidenta USA Donalda Trumpa, oproti sedmi dvoustranným setkáním prezidenta Mun Če-ina. K tomu bylo navíc zrušeno trojstranné setkání mezi jižní Koreou, Spojenými státy a Japonskem – které se konalo na všech předchozích summitech G20, a bylo nahrazeno setkáním USA, Japonska a Indie. Společné prohlášení Japonska a USA navíc vypustilo část „spolupráce jižní Koreje, USA a Japonska“. Tím byl ukázán současný stav Koreje, která byla odsunuta na samý okraj východoasijské politiky.

문재인 대통령이 일본 오사카에서 열린 20 개국(G20) 정상회의 일정을 마무리하고 29 일 귀국했다. 이번 회의는 올들어 비핵화 협상에서 밀려나 있던 대한민국이 열강들과의 직접 대화를 통해 외교력을 회복할 기회로 여겨졌다. 그러나 결과를 보면 성과는 미미해 보인다. 우선 개최국인 일본과의 정상회담부터 열리지 못했다. 다자회의 참석차 일본 땅을 밟은 우리 대통령이 일본 총리와 양자 회담을 갖지 못한 건 유례가 없다. 한·일 관계가 얼마나 심각한 수준으로 추락했는지 단적으로 보여준다. 반면 아베 신조 일본 총리는 도널드 트럼프 미국 대통령을 비롯해 약 20 개 나라 정상이나 국제기구 수장과 양자 회담을 가졌다. 문 대통령이 7 차례 양자 회담을 가진 것과 대비된다. 게다가 G20 회의마다 열리곤 했던 한·미·일 정상회의도 무산됐고 미·일·인도 3 자 정상회의가 그 자리를 대신했다. 미·일 정상회담 뒤 공동성명에서도 '한·미·일 공조'란 표현이 빠졌다. 동아시아 정치에서 변방으로 밀려난 한국의 현주소가 적나라하게 드러난 것이다.

- Překlad vytvořený člověkem

President Moon Jae-in has returned home on 29th after wrapping up his attendance of the Group of 20 (G-20) summit in Osaka, Japan. This meeting was seen an opportunity for Republic of Korea, which has been excluded from this year denuclearization negotiations, to restore its diplomatic capability through direct dialog with world powers. However, looking at the outcome, results seem to be insignificant. Firstly, he could not have bilateral talks with the host country Japan. It is unprecedented that the South Korean president, who traveled to Japan to attend multilateral meetings, failed to hold bilateral talks with the Japanese prime minister. This shows how serious is the state of Korea-Japan relationship. On the other hand, Japanese Prime Minister Abe Shinzo held 20 summits with foreign leaders or heads of international organizations including US. President Donald Trump. It contrasts with President Moon's sever rounds of bilateral talks. Moreover, a tripartite summit among South Korea, the United States and Japan — which was held at every G-20 Summit was cancelled and replaced by a summit among the United States, Japan and India instead. The joint statement after the U.S.-Japan summit also

omitted the expression "cooperation between South Korea, the U.S. and Japan." The current state of Korea, which has been relegated to the periphery of East Asian politics, has been clearly revealed.

- Naver Papago

Moon Jae-in president to finish the conference (g20) 20 countries summit held in Osaka, Japan, 29, returned from abroad. The meeting was seen as an opportunity for the Republic of Korea, which had been pushed out of the denuclearization negotiations this year, to restore its diplomatic prowess through direct talks with the powers. Looking at the results, however, the results appear to be insignificant. First of all, it failed to hold a summit with host Japan. It is unprecedented that the South Korean president, who traveled to Japan to attend multilateral meetings, failed to hold bilateral talks with the Japanese prime minister. It shows how serious the Korea-Japan relationship has been. Japanese Prime Minister Shinzo Abe, on the other hand, held bilateral talks with U.S. President Donald Trump and heads of about 20 countries or international organizations. It contrasts with President Moon's seven rounds of bilateral talks. Moreover, the Korea-U.S., Japan-Japan summit, which used to be held at every G20 meeting, was cancelled, and the U.S., Japan, and India replaced the meeting. The joint statement after the U.S.-Japan summit also omitted the expression "cooperation between South Korea, the U.S. and Japan." The current state of Korea, which has been relegated to the periphery of East Asian politics, has been clearly revealed.

- Google Translate

President Moon Jae-in returned to Japan on Friday after finishing the 20th G20 summit in Osaka, Japan. The meeting was considered an opportunity for the Republic of Korea, which had been left out of negotiations for denuclearization, to recover diplomatic power through direct talks with the powers. However, the results seem to be insignificant. First, it did not open since the summit with Japan, the host country. It is unusual for our president, who has stepped up to Japan to attend a multilateral meeting, to have no bilateral talks with Japanese prime minister. It shows clearly how serious the relationship between Korea and Japan has fallen. Prime Minister Shinzo Abe, meanwhile, held bilateral talks with US President George W. Bush and top leaders of international organizations, including some 20 nations. Compared to Mr. Moon 's seven bilateral talks. In addition, the summit meeting between the two Koreas, the US, Japan, and

Japan, which had been held every G20 meeting, was abolished, and the three-way summit meeting of the US, Japan, and India took its place. In the joint statement after the US-Japan summit, the expression "Korea-US-Japan cooperation" is missing. The present state of Korea, which has been pushed to the edge of East Asian politics, has been exposed nakedly.

Jak můžeme jasně vidět, v obou překladech jsou velké rozdíly. Naver Papago vytváří hned v první větě negramatickou část *president to finish the conference*. Problémem se zdá být sloveso *mamurihago* (마무리하고), jehož součástí je pomocná partikule *hago* (하고), kterou zde překladač nerozpoznal a nevyužil. Jelikož je tato věta v korejštině gramaticky správná a nemá ambiguidní význam, dalo by se předpokládat, že pravidlový překladač by si s touto větou co se týče gramatiky poradil lépe. Otázkou ovšem zůstává, co by takový překladač udělal se zkratkou G20 – jak můžeme vidět, i neurální překladač si tuto zkratku nesprávně upravil na g20. Následujícímu větě můžeme vytknout snad nezvyklý výběr slov, ale zastavit se můžeme u věty třetí, kde překladač dvakrát v krátkém rozmezí použije stejný překlad pro slovo *kjŏlkwa* (결과) a *sŏngkwa* (성과). Přestože se oba termíny dají přeložit stejně, použití synonyma by zvyšovalo přirozenost výsledného překladu. Hned v další větě si pak můžeme všimnout věty s nevyjádřeným podmětem v korejštině. Není moc překvapivé, že si strojový překladač nedokázal z kontextu vyvodit, že by tam mohl být např. prezident, a pro vytvoření správné věty doplnit zájmeno středního rodu, *it*.

Typický problém prozrazující, že je překladač založený ne na pravidlech, nýbrž závisí na korpusu a dalších příkladových datech, je třetí věta od konce, *Kedaga G20 höüimada jölligon hättön han · mi · il čöngtanghöüido musantwättko mi · il · indo 3ča čöngsanghöüga kü čarirül däsinhätta*, kde si překladač špatně do věty doplnil dvakrát Japonsko. Nakonec nás zde může zarazit naopak absence slova setkání či meeting v téže větě, kdy meeting USA – Koreje – Japonska nahradily USA – Japonsko – Indie, nikoliv jejich meeting. Nakonec můžeme tento překlad shrnout tak, že Naver dokázal z většiny zachovat původní smysl textu, ačkoliv výběrem terminologie může překlad působit značně nepřirozeně. Nejvýraznějšími chybami jsou v jednom případě gramatické, kdy je nezvládnutý rozbor věty, nedostatečné využití kontextu pro doplnění nevyjádřeného podmětu, vložení slova navíc a naopak eliminace slova.

Překlad Google Translate si narozdíl od Naveru dokázal poradit s pomocnou partikulí *hago* (하고), nicméně dopustil se zde podobně závažných chyb, ne-li závažnějších. Místo toho, aby se zde prezident vracel z Japonska, se vrací do Japonska a překladač navíc rozhodl, že 29. přeloží jako konkrétní den, a to pátek, což může být další typickou chybou překladače závislého na datech. Nakonec ještě v téže větě vypustí slovo *skupina* a číslo 20 vyhodnotí jako řadové číslo, čímž se opět dopouští zkreslování informací. Podobně jako u Naveru i zde má překladač problém s třetí větou s nevyjádřeným podmětem. Google zde ovšem ještě k doplnění středního zájména překládá větu doslovně, což je chyba, která by se dala čekat právě spíše od pravidlového překladače, a která opět mění význam celé věty.

Závažnou chybou opět typickou pro překladače založené na datech je ve větě o setkání Abe Šinzóa s Donaldem Trumpem. Překladač zde naprosto chybně přeložil jméno amerického prezidenta jako George W. Bush. Hned v další větě pak můžeme vidět příklad nedodání nevyjádřeného podmětu, což zase dělá větu negramatickou. Chybu s dodáním slova Japonska v následující větě jsme se již setkali u Naveru. Nakonec nás zde může zarazit poslední poněkud nadbytečné slovo *nakedly*.

Překlad Googlu se od Naveru v mnohém liší. Všeobecně zde Google používá přirozenější slovní zásobu a některé úpravy, jako jedna eliminace slova *results* ve třetí větě pomáhají plynulosti textu.

Je však zřejmé, že zde má Google problémy nejen s využitím kontextu pro doplnění nevyjádřeného podmětu, ale i se samotnou stavbou věty s nevyjádřeným podmětem, a taktéž se zde vyskytuje stejný případ vložení slova navíc. Na rozdíl od Naveru se zde vyskytuje více vyloženě nesprávných překladů, jako 29. – pátek, Skupina 20 – dvacátý nebo Donald Trump – George W. Bush.

Celkově zde tedy můžeme říct, že Google dosahuje lepších výsledků z hlediska využití slovní zásoby a přirozenosti textu, objevují se zde však závažnější chyby z hlediska přesnosti překladu a gramatické struktury. Naver, ač méně přirozeným způsobem, zde naopak dokázal lépe převést stěžejní informace článku.

3.3.2 Komerční text

Dalším z textů, které nechám překladače zpracovat, je text týkající se služeb cloudových služeb Azure od Microsoftu, který je dostupný na <https://azure.microsoft.com/ko-kr/overview/what-is-cloud-computing/>. Vybraná pasáž textu má jen cca 60 slov a je lokalizovanou verzí původně anglického textu, který je možné najít na téže stránce.⁴⁸

- Korejský text:

Kchlaud kchömpjutchingiran?

Süpgə marhamjön kchlaud kchömpjutchingün intchönet („kchlaud“) ül tchonghşa söbi, stchoridži, deitchöbeis, netchwökching, sopchütchüweö, punsök, intchellidžöns tüngüi kchömpchjutching söbisürül čekonghanün kösimnida. Kchlaud kchömpjutchingül tchonghä tö pparün, hjöksikkwa jujönhan risosürül čekonghago täkjumo kjöngjögüi iik hjokwarül nuril su ittke tümnida. Ilbandžogüro sajonghan kchlaud söbisüe tähäšöman jogümül čibulhamüro, unjöng pijongül natčchugo inpchrarül poda hjojulčogüro unjönghal su issümnida. Kűppunman anira, pidžnisü jogu sahangü pjönhwae ttafa kjumorül čočönghal su issümnida.⁴⁹

클라우드 컴퓨팅이란?

쉽게 말하면 클라우드 컴퓨팅은 인터넷(“클라우드”)을 통해 서버, 스토리지, 데이터베이스, 네트워킹, 소프트웨어, 분석, 인텔리전스 등의 컴퓨팅 서비스를 제공하는 것입니다. 클라우딩 컴퓨팅을 통해 더 빠른 혁신과 유연한 리소스를 제공하고 대규모 경영의 이익 효과를 누릴 수 있게 됩니다. 일반적으로 사용한 클라우드 서비스에 대해서만 요금을 지불하므로, 운영 비용을 낮추고 인프라를 보다 효율적으로 운영할 수 있습니다. 그뿐만 아니라, 비즈니스 요구 사항의 변화에 따라 규모를 조정할 수 있습니다.

⁴⁸ Stránku lze přepnout do angličtiny přepsáním jazykové zkratky ko-kr v adrese webu na en-us.

⁴⁹ Co je cloud computing? Jednoduše řečeno, cloud computing je doručování výpočetních služeb, včetně serverů, úložišť, databází, sítí, softwaru, analytických nástrojů a inteligentních funkcí, přes internet („cloud“) a nabízí rychlejší inovace, flexibilitu prostředků a cenové výhody. Obvykle platíte jenom za cloudové služby, které skutečně využijete, což pomáhá snižovat provozní náklady, efektivněji provozovat infrastrukturu a škálovat s ohledem na měnící se obchodní potřeby.

- Překlad vytvořený člověkem

What is cloud computing?

Simply put, cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping you lower your operating costs, run your infrastructure more efficiently, and scale as your business needs change.

- Naver Papago:

What is cloud computing?

Simply put, cloud computing is about delivering computing services such as servers, storage, databases, networking, software, analytics, and intelligence over the Internet ("cloud"). Cloud computing enables you to deliver faster innovation, more flexible resources, and the benefits of large scale operations. Typically, you pay only for the cloud services you use, so you can lower your operating costs and run your infrastructure more efficiently. Plus, you can scale as your business needs change.

- Google Translate:

What is Cloud Computing?

In a nutshell, cloud computing is to provide computing services such as servers, storage, databases, networking, software, analytics, and intelligence through the Internet ("cloud"). Cloud computing can provide faster innovation, more flexible resources, and the benefits of large-scale operations. You pay only for commonly used cloud services, so you can lower your operating costs and run your infrastructure more efficiently. Not only that, you can scale as your business requirements change.

Na první pohled můžeme u obou překladačů vidět relativně dobrý překlad, který je plynulý a smysuplný. Oba překladače produkují velmi podobný překlad až do třetí věty, která může být interpretována více způsoby, a proto je vzhledem k její ambiguitě až překvapivé, že zde Naver správně vyhodnotil, že výraz *ilbančögüro* (일반적으로) zde znamená „typicky“. Fakt, že překladač vyhodnotil výraz správně, zde může být zásluhou právě neurálních sítí a statistiky,

která se již s tímto případem setkala a překladač tudíž nepřekládal slova samostatně a doslovně. Google Translate přeložil tento výraz jako „běžně“ a spojil ho s výrazem „používané“, čímž význam věty, že „uživatel platí jen za služby, které využívá“, na „uživatel platí jen za běžně užívané služby“, což může čtenáře bez další specifikace, že jde o služby využívané právě uživatelem, v tomto případě zmást. Naver tedy v této situaci zvládl vyprodukovat přesnější překlad.

Důvodem, proč si s tímto textem poradily oba překladače relativně dobře mohou být krátké jednoduché věty a slovní zásoba přejatá z angličtiny. Následující text ovšem obsahuje už složitější větné konstrukce a technickou terminologii. Vybraná pasáž o přibližně 160 slovech pochází z webové stránky společnosti Samsung, který je dostupný na:

<https://www.samsung.com/sec/aboutsamsung/business-area/device-solution/>. Jelikož stránka neobsahuje shodující se anglickou verzi tohoto textu, byl překlad vytvořen samostatně.

- Korejský text

Samsöngčöndžanün 1993 njön irä segje memori pandočche sidžangesö söndurül judžihamjö tokbočögin kjöngdžängrjöküro küllöböl IT sidžange kijöhago issümnida. 2016 njönenün pandočche kisulü hangjerül tolpcahan 10 nanogüp 16Gb mobail D rämül jangsanhattgo, 3sedä (48tan) 3D V nändürül kibanüro süthoridži punja jökä čhögo jongrjangin 15.36TB SAS SSD rül jangsanhanün tüng čhasedä pchürimiöm čepchumül čhulsihässümnida. Ttohan čhogosok, čchosohjöng, čchočölcön čhasedä sistchäm sölgjerül ühan 8GB HBM2 D räm, 8GB LPDDR4/4X mobail D räm, M.2 NVMe SSD 960 PRO 2TB, muge Igü pchäkchidzie modün pandočcherül čipčökhan 512GB BGA NVMe SSD tüng ipkjesö juilhage konggüphamjö pchürimiöm memori sidžangül čchangčchulhago issümnida. 2017 njönenün mobail, PC, entchöpchraidžü, kchlaud söbisü, kasanghjönsil (VR), čüngganghjönsil (AR), Automotive tüng töuk tajanghan punjaesöü kogäk jogue täünghagi ühä 4sedä(64 tan) 3D V nändü kiban SSD jangsanül sidžaküro, 512GB 3D V nändü kiban čhogojongrjang söbi SSD, čchosohjöng eUFS, čhogosok Z-SSD tüng kidžon täbi sokdowa jongrjangül töuk nopchin čhasedä rainöbül sönböil jedžöngimnida. Tchükhi čhasedä 10 nanogüp D rämüwa 5 sedä 3D V nändü tüng čhömtan

*kongdžöngül čokkie kabalhe pchürimiöm memori sidžang söngdžangül čisokčögüro čundohal kösimnida.*⁵⁰

삼성전자는 1993 년 이래 세계 메모리 반도체 시장에서 선두를 유지하며 독보적인 경쟁력으로 글로벌 IT 시장의 성장에 기여하고 있습니다. 2016 년에는 반도체 기술의 한계를 돌파한 10 나노급 16Gb 모바일 D 램을 양산했고, 3 세대(48 단) 3D V 낸드를 기반으로 스토리지 분야 역대 최고 용량인 15.36TB SAS SSD 를 양산하는 등 차세대 프리미엄 제품을 출시했습니다.

또한 초고속, 초소형, 초절전 차세대 시스템 설계를 위한 8GB HBM2 D 램, 8GB LPDDR4/4X 모바일 D 램, M.2 NVMe SSD 960 PRO 2TB, 무게 1g 의 패키지에 모든 반도체를 집적한 512GB BGA NVMe SSD 등을 업계에서 유일하게 공급하며 프리미엄 메모리 시장을 창출하고 있습니다. 2017 년에는 모바일, PC, 엔터프라이즈, 클라우드 서비스, 가상현실(VR), 증강현실(AR), Automotive 등 더욱 다양한 분야에서의 고객 요구에 대응하기 위해 4 세대(64 단) 3D V 낸드 기반 SSD 양산을 시작으로, 512GB 3D V 낸드 기반 초고용량 서버 SSD, 초소형 eUFS, 초고속 Z-SSD 등 기존 대비 속도와 용량을 더욱 높인

⁵⁰ Samsung Electronics si udržuje vedení na celosvětovém trhu čipů již od roku 1993 a s bezkonkurenční soutěživostí přispívá růstu globálního trhu IT. V roce 2016 společnost začala ve velkém vyrábět 10nm typ 16GB mobilních DRAM pamětí, které prolomily limity technologie čipů a zveřejnila novou generaci prémiových produktů, jako jsou 15,36TB SAS SSD s nejvyšší úložnou kapacitou založené na třetí generaci 48 vrstevné 3D V-NAND. Firma je také jediná ve svém odvětví, která také nabízí 512GB BGA NVMe SSD a další, který spojuje všechny 8GB HBM2 DRAM, 8GB LPDDR4/4X mobilní DRAM, M.2 NVMe SSD 960 PRO 2 TB pro další generace ultra-rychlého, ultra-malého a energii šetřícího systémového designu čipů v 1 gramovém balení. Abychom odpověděli na požadavky zákazníků z počítačového a firemního odvětví, cloudových služeb, virtuální reality (VR) augmentované reality (AR) automobilové aplikace a mnohých další odvětví, počínaje masovou produkcí čtvrté generace 64-vrstevných SSD založených na 3D V-NAND a společnost v roce 2017 představí sestavu dalších generací s vyšší rychlostí a kapacitou ve srovnání s předchozími generacemi, jako vysokokapacitný SSD servery založené na 512GB 3D V-NAND, ultrakompaktní eUFS, vysokorychlostní Z-SSD a další. Nadále budeme také vylepšovat naši 10nm technologii pro řešení DRAM a technologie V-NAND příští generace, abychom si udrželi naše postavení a přispívali k růstu světového trhu paměťových řešení.

차세대 라인업을 선보일 예정입니다. 특히 차세대 10 나노급 D램과 5세대 3D V낸드 등 첨단 공정을 적기에 개발해 프리미엄 메모리 시장 성장을 지속적으로 주도할 것입니다.

- Překlad vytvořený člověkem

Samsung Electronics has been maintaining a leadership in the global memory chip market since 1993 and contributes to global IT market growth with unrivaled competitiveness. In 2016 the company mass-produced 10-nanometer class 16GB mobile DRAM which breaks through the limits of chip technology and released next generation of premium products such as 15.36TB SAS SSD with the highest storage capacity based on third generation (48-layer) 3D V-NAND. The company is also the only one in industry that also supplies 512GB BGA NVMe SSD etc. which aggregates all 8GB HBM2 DRAM, 8GB LPDDR4/4X mobile DRAM, M.2 NVMe SSD 960 PRO 2 TB for next generation of ultra-fast, ultra-small and power saving system design chips in 1g weight package. To respond to customer needs in various fields as , PC, enterprise, cloud servis, virtual reality (VR), augement reality (AR), Automotive and others, the company will in 2017 start with mass production of 4th generation (64-layer) 3D V-NAND-based SSDs and introduce next-generation lineup with higher speed and capacity in comparison to previous generations, as ultra-high capacity server SSD based on 512GB 3D V-NAND, ultra-compact eUFS, high-speed Z-SSD and more. We will especially continue to lead the premium memory market growth by developing advanced processes at the right time as new generation of 10-nanometer DRAM and fifth-generation 3D V-NAND etc.

- Naver Papago

Samsung Electronics has been leading the global memory chip market since 1993, contributing to the growth of the global IT market with its unrivaled competitiveness. In 2016, the company mass-produced 10-nano 16Gb mobile DRAM, which surpassed the limits of semiconductor technology, and released its next-generation premium products such as 15.36TB SAS SSD, the highest capacity ever in storage, based on the third-generation (48-layer) 3D V-NAND. It also provides 8GB HBM2 DRAM, 8GB LPDDR4/4X mobile DRAM, M.2 NVMe SSD 960 PRO 2TB, and 512GB BGA NVMe SSD with all semiconductors in a 1g package to create a premium memory market. In 2017, the company will introduce next-generation lineups with higher speed

and capacity compared to the existing 512GB 3D V-NAND-based ultra-high capacity server SSD, ultra-small eUFS, and ultra-fast Z-SSD, starting with the mass production of 4G (64-layer) 3D V-NAND-based SSD to respond to customer needs in more diverse areas such as mobile, PC, enterprise, virtual reality (VR), AR, AR, AR, AR, AR, AR, and automotive. In particular, we will continue to lead growth of premium memory markets by developing high-tech processes such as next-generation 10-nano DRAM and fifth-generation 3D V-NAND at the right time.

- Google Translate

Samsung Electronics has maintained its leading position in the global memory semiconductor market since 1993 and has contributed to the growth of the global IT market with unparalleled competitiveness. By 2016, Samsung Electronics has mass-produced 10Gb mobile 16Gb DRAM, We launched next-generation premium products, such as mass production of 15.36TB SAS SSD, the highest capacity storage area ever, based on 3G (48-stage) 3D V NAND.

In addition, 8GB HBM2 DRAM, 8GB LPDDR4 / 4X mobile DRAM, M.2 NVMe SSD 960 PRO 2TB, and 512GB BGA NVMe SSD, all in one package weighing 1g, are designed for next-generation ultra-fast, And is creating a premium memory market.

In order to respond to customer needs in various fields such as mobile, PC, enterprise, cloud service, virtual reality (VR), augmented reality (AR) and automotive in 2017, mass production of 4th generation (64 stages) 3D V NAND- Starting with a 512GB 3D V NAND-based ultra high capacity server SSD, ultra-small eUFS and ultra-fast Z-SSD,

In particular, we will continue to lead the growth of the premium memory market by developing cutting-edge processes such as next-generation 10-nm DRAM and fifth-generation 3D V-NAND in a timely manner.

Použitý text je o poznání složitější než předchozí jak kvůli své struktuře, tak slovní zásobě. Naver Papago si v první části vedl až překvapivě dobře. Od třetí věty dále si však můžeme všimnout vynechávání informací, že společnost je ve svém odvětví jedinou, která dané produkty nabízí, nebo výčet pozitivních vlastností představených paměťových čipů.

V komerčním prostředí jsou ovšem podobné informace pro zákazníka zásadní. Toto vynechávání částí vět se objevuje ve strukturálně náročnějších pasážích, které neurální překladač nedokáže zpracovat. Ukazuje to na nedostatečně vycvičený neurální systém v lingvistickém zpracování

struktury. Další velmi výrazný problém vidíme v předposlední větě, kde překladač místo augmentované reality nesmyslně opakuje zkratku AR. Původ této chyby není bohužel dohledatelný. Naopak zkratky odkazující k technickým názvům nepředstavují pro tento překladač žádnou překážku.

Google Translate má k technickým výrazům a složité struktuře naprosto odlišný přístup. Přestože zkratky zde existují ve stejné podobě i v anglickém jazyce, kvůli svým algoritmům se snaží tyto zkratky překládat a dopuští se tak chyb, kdy přepisuje jednotky a z technického hlediska pak vytváří nesmysly. Ještě větší problémy pak ovšem nastávají od třetí věty, kdy se objevuje složitější struktura korejské věty. Podobně jako Naver si s ní Google nedokáže moc dobře poradit, avšak na rozdíl od Naveru se své nesmyslné části překladu snaží spojit s částmi textu, se kterými by si teoreticky poradit mohl, a ve výsledku zde vnikají mnohem větší nesmyslné chyby. Věty na sebe nenavazují, v některých chybí řídicí sloveso a nejsou oddělené tečkami, nýbrž čárkami. Krom toho má zřejmě Google větší problém i s technickými výrazy, jako 48-vrstevný apod. Můžeme tak říct, že především prostřední část textu není dostatečně srozumitelná pro potenciálního zákazníka a překladač tak v tomto překladu nesplnil svůj účel.

Oba překladače se nezvládly vypořádat s technickým textem pro komerční účely na sto procent a u obou je nutná další úprava člověkem. Přesto se dá z počtu a povahy chyb říct, že Naver dokázal text zpracovat přeci jen o něco lépe než konkurenční Google.

3.3.3 Mluvená řeč

Od předchozích textů se mluvená řeč liší jak strukturálně, tak tématicky. Takový text proto může poskytnout další pohled na práci strojových překladačů. Následující vybraný text byl převzat z webu projektu Humans of Seoul dostupného na <https://humansofseoul.com/>. Lidé v rozhovorech jsou náhodně vybraní a přepsaný korejský text je pak zveřejněn v korejštině i angličtině. Většina rozhovorů je zveřejněna na facebookové stránce <https://www.facebook.com/seoulhumans>. První zvolený text má v korejštině rozsah 167 slov.

- Korejský text

„Näga sŭmulhan sare hangŭrŭl čchöŭm päwössöjo. Kŭrŭl nŭtke päwötdžiman näga mwöl čal ssönjahamjŏn, jŏnäpčhŏndžirŭl kiga makhage čal ssössöjo. Mullon čigŭm jetnal sikŭro

jönäpchjödži ssümjön ,aigo nömu koritchapungada‘ kūröl köjejo. Jetnarenün jusikhäpoige sadžasöngörül manhi ssötkötünjo. Jerül tūlmjön härodonghjöl. Päkñjönül katči saldaga han mudöme türökandanün ttütiejo. Jönäpchödzie irön köl jangnjömsam ssümjön kkömpäck čugöjo. Näka sūmul sönösalcčümüöna. Kačchi pūročchi čangsa hättön ihongsöbiranün äga issönünde tangsi Hong-söbiga čosönpangčikiranün kchün hösa taninün agassirül kkosirjögo hanün köja. Kjörhonhal ttäčkači näga jönäpchjödžirül täsin ssöčwössöjo. Kü ihuro sewori manhi činačjo. Öñunal pumonim musjöhöhün Hapčchön sansorül tanjöhönünde čchinguhantche čonhwaga on köjejo. Hong-söpiga amūro čugöttago. Nanün Pusane inün čchinguga čugötta hamjön čigūmirado nārjögänün saramiejo. Kūräsö paro čangñjesike kaččjo. Jöngčöng apcheso ,ja čchinguja, öttökhe hömuhake kanja‘ hago nökdurirül hadaga, param ssörjögo pakke nagassöjo. Kūnde Hong-söpü manuraka ttara nawa irökhe marhadäjo. ,Adžossi čham komapsūmnida, Söuleso jögikkadži wadžusigo. Kūnde näga adžossihantche ttak hankadži muröpol ke issöjo. Jennare uri ägi abodžihago jönähal ttä kü pchjödži adžossiga ssössöčjo?‘ Öju... Nanün nollačjo. Morül čul aranūnde. Čagin ta algo issöttäjo, čchöūbutchö. Kibuni čom isanghadä. “⁵¹

“내가 스물한 살에 한글을 처음 배웠어요. 글을 늦게 배웠지만 내가 뭘 잘 썼냐하면, 연애편지를 기가 막히게 잘 썼어요. 물론 지금 옛날 식으로 연애편지 쓰면 ‘아이고 너무 고리타분하다’ 그럴 거예요. 옛날에는 유식해보이게 사자성어를 많이 썼거든요. 예를 들면 해로동혈. 백년을 같이 살다가 한 무덤에 들어간다는 뜻이에요. 연애편지에 이런 걸 양념삼아 쓰면 껌뻑 죽어요. 내가 스물 서너살쯤이었나. 같이 브로치 장사 했던

⁵¹ Naučil hangül když mi bylo 21 let. Přestože jsem se to naučil pozdě, uměl jsem docela dobře psát, obzvlášť milostné dopisy. Samozřejmě kdybych se pokusil takový dopis napsat dnes, lidé by říkali, „Páni, jak staromódní.“ Tehdy jsme psali čtyřslabičné idiomy ve snaze vypadat vzdělanými. Například „hä-ro-dong-hjöl“. Znamená to být pohřbeni společně v jednom hrobě poté, co spolu strávili 100 let. Okořenit dopisy něčím takovým dělalo z dopisu pecku. Když mi bylo 23 nebo 24, pracoval jsem s jakýmsi I Hong-söbem v podniku s brožemi. Hong-söb se tehdy snažil randit s dívkou, která pracovala ve velké firmě Čosönská textilní manufaktura. Psal jsem pro něj milostné dopisy dokud se nevzali. Od té doby uběhlo hodně času. Jednoho dne, když jsem se vrátil z návštěvy hřbitova Hapčchön, kde jsou pohřbeni mí rodiče, mi zavolal kamarád. Řekl, že Hong-söb zemřel na rakovinu. Jsem člověk, který když některý z mých přátel z Pusanu zemřel, tam jel celou tu cestu, aby je vyprovodil. Takže jsem rovnou vyrazil na pohřeb. Trochu jsem mluvil před pamětní fotkou a říkal jsem něco jako „Ach, příteli, jak jsi mohl takhle odejít?“ a podobně. Poté jsem šel ven nadýchat se čerstvého vzduchu. Hong-söbova žena mě následovala a řekla „Děkuji, že jsi sem jel celou tu cestu až ze Söulu. Mám na tebe jednu otázku. Když jsme tehdy já a Hong-söb randili, byl jsi to ty, kdo psal ty dopisy, že?“ To mě tedy překvapilo. Myslel jsem, že se to nikdy nedozvěděla. Řekla, že to věděla celou dobu už od začátku. Bylo to trochu zvláštní.

이홍섭이라는 애가 있었는데 당시 홍섭이가 조선방직이라는 큰 회사 다니는 아가씨를 꼬시려고 하는 거야. 결혼할 때까지 내가 연애편지를 대신 써줬어요. 그 이후로 세월이 많이 지났죠. 어느날 부모님 모셔놓은 합천 산소를 다녀오는데 친구한테 전화가 온 거예요. 홍섭이가 암으로 죽었다고. 나는 부산에 있는 친구가 죽었다 하면 지금이라도 내려가는 사람이에요. 그래서 바로 장례식에 갔죠. 영정 앞에서 ‘야 친구야, 어떻게 이렇게 허무하게 가냐’ 하고 녀두리를 하다가, 바람 쐬려고 밖에 나갔어요. 근데 홍섭이 마누라가 따라 나와 이렇게 말하대요. ‘아저씨 참 고맙습니다, 서울에서 여기까지 와주시고. 근데 내가 아저씨한테 딱 한가지 물어볼 게 있어요. 옛날에 우리 애기 아버지하고 연애할 때 그 편지 아저씨가 썼었죠?’ 어유...나는 놀랐죠. 모를 줄 알았는데. 자긴 다 알고 있었대요, 처음부터. 기분이 좀 이상하대.”

- Překlad vytvořený člověkem

"I learned Hangul for the first time when I was 21 years old. Though I learned late, I wrote quite well, especially love letters. Of course if I try to write a love letter like that today people would say, "Wow, how old fashioned." Back then, we'd write with four-character idioms to try and seem educated. For example, 'Hae Ro Dong Hyul.' It means being buried together in one grave after being together for your whole lives. Spicing up a love letter with something like this was sure to be a killer. When I was 23 or 24, I worked with this kid Hong-seob Lee in the brooch business. Back then, Hong-seob was trying to date this girl who worked at a big company called Joseon Textile Manufacturing. I wrote love letters for him until they married. It's been a long time since then. One day, after coming back from visiting the Hapcheon graveyard, where my parents are buried, a friend called me. Said Hong-seob had passed away from cancer. If any of my Busan friends pass away, I'm the kind of person who still goes all the way to see them off. So I went straight to the funeral. I talked a bit in front of his memorial picture, saying, 'Ah, friend, how could you go like this?' and such. Then I went outside to get some air. Hong-seob's wife followed

me out, and said, "Thank you for coming all the way from Seoul. I just have one question for you. Back then when Hong-seob and I were dating, you wrote those letters, right?" Whew, was I surprised. I thought she'd never know. She said she had known all along, from the very beginning. That felt a little strange."

- Naver Papago

"When I was 21, I learned Hangul for the first time. I learned how to write late, but I wrote a great love letter. Of course, if you write a love letter in the old fashioned way, you'll be like, "Oh, it's too much of a fuss." In the old days, we used a lot of idioms to make them look good. Harrocyclic blood, for example. It means that they will live together for 100 years and then go into a grave. If you use this as a spice in a love letter, you'll die. I was about twenty-four. There was a girl named Lee Hong-seop who sold brooch with us. At that time, Hong-seop was trying to lure a girl from a big company called Joseon Bangjik. I wrote a love letter for her until she got married. Many years have passed since then. One day, I was visiting Hapcheon Oxygen where my parents were staying, and my friend called me. Hongsub died of cancer. If my friend in Busan is dead, I'm going down now. So I went to the funeral right away. I was wondering, "Hey, buddy, how can I go like this?" And I went outside to get some air. But Hongsub said his wife followed him out and said, Thank you so much. Please come all the way from Seoul. But I have only one question for you. You wrote that letter when you were dating my baby's father, right? I was surprised. I thought you wouldn't know. He knew everything from the beginning. They're feeling a little weird."

- Google Translate

"I first learned Hangul at the age of twenty - one. I learned to write late, but if I wrote well, I wrote my love letter wonderfully. Of course, if you write a love letter in an old fashioned way, it will be "I am so annoying". In the old days, I used a lot of lion sex to look familiar. For example, the blood of the sea. It means living in a grave for a hundred years. If you write this in your love letter, you will die. I was about twenty or three. There was a boy called Lee Hongseop who broke a brooch, but he was trying to cheat on a lady who was a big company called Chosun weaving at that time. Until I got married, I wrote my love letter instead. It has been a long time since then. One day, my friend is taking oxygen from my parents. He died of cancer. I am a person who goes down now if a friend in Busan is dead. So I went to the funeral. In front of Yongjeon, he said,

'Hey, man, how are you going to be like this?' I complained and went outside to air it. But Hong-seop follows his wife and says, 'Thank you very much, Mr. But I just have one thing to ask you. I used to write that letter when I was in love with my father 's father in the old days. I thought you did not know. You knew everything, from the beginning. I feel a little strange. "

Původní korejský text nepatří k nejsnažším pro překlad. Obsahuje idiom, vlastní jména a střídá se v něm přímá řeč s nepřímou. Hned v druhé větě si můžeme všimnout, že měl překladáč Naver s tímto textem problémy. Místo původního významu, že autor „psal docela dobře a obzvláště milostné dopisy“ zkrátil informaci na to, že autor napsal jeden dobrý dopis. Korejský text nespecifikuje, jestli byl dopis jeden či sto, ovšem informace, že autor psal dobře, je zde úplně vynechaná, tudíž překlad není možné považovat za správný. Hned v další větě pak můžeme vidět chybný překlad slova *koritchabunhada* (고리타분하다)⁵², které překladáč přeložil chybně jako frázi *be too much of a fuss*⁵³. Zde se můžeme jen domnívat, zdali bylo faktorem, který překladáč zmatl, zvolání *aigo* (아이고)⁵⁴. Ačkoliv to byla očekávaná chyba, překladáč si neporadil ani s čtyřslabičným idiomem, který přeložil jako *Harrocytic blood*, což je spojení, které ani neexistuje. V následující větě *Päknjönül katči saldaga han mudöme türökandanün ttütiejo*. (백년을 같이 살다가 한 무덤에 들어간다는 뜻이에요) vynechává i informaci ve vysvětlení idiomu, že „dotyční jsou pohřbení do stejného hrobu“ Následné varování, že pokud tímto okořeníte dopis, zemřete, je zase ukázkou toho, že si překladáč nedokázal vzít v potaz jiný než doslovný význam, ovšem opět jde o chybu, která je typická spíše pro pravidlový překladáč než pro statistický nebo neurální. Případů, kdy překladáč vynechá podstatnou informaci pak můžeme vidět více ve větách jako je *Näka sümul sönösalččümiöna*. (내가 스물 서너살쯤이었나.)⁵⁵, *Jöngčöng apcheso ja čchinguja, öttökhe hömuhake kanja hago* (영정 앞에서 ‘야 친구야, 어떻게 이렇게 허무하게 가냐’ 하고)⁵⁶ nebo *Nanün Pusane inün čchinguga čugötta hamjön*

⁵² Být staromódní

⁵³ Je kolem toho moc povyku

⁵⁴ Zvolání

⁵⁵ Bylo mi 23 nebo 24 let.

⁵⁶ Před vzpomínkovou fotkou jsem říkal „Ach, příteli, jak jsi mohl takhle odejít?“

čigūmirado nārjōganūn saramiejo (나는 부산에 있는 친구가 죽었다 하면 지금이라도 내려가는 사람이에요)⁵⁷. Přesně tenhle druh chyby je typický pro překladače, které jsou založené na datech. Oproti tomu se chyby v určení pohlaví autorova přítele, nepřesné vyjádření o podniku, ve kterém autor pracoval, nebo nepřírozený výběr slova *lure*⁵⁸ pro vyjádření, že se muž dvořil jisté dívce, se nezdají tak závažné. Zvláštní je, že si překladač nedokázal poradit se slovy jako je hřbitov nebo pohřbení, místo kterých do překladu vložil slova *oxygen*⁵⁹ a *were staying*⁶⁰, čímž opět vznikl nesmyslný překlad původního textu. Poslední tři věty textu ukazují chyby jak ve špatném rozboru korejské věty a určení závislostí jednotlivých větných členů, jako například ve spojení *Hong-sōpi manuraka* (홍섭이 마누라가)⁶¹, nerozpoznání přímé řeči a špatném doplňování nevyjádřeného podmětu.

Je zřejmé, že Naver Papago v překladu tohoto typu textu selhal. Překlad od Google Translate však dosahuje ještě horších výsledků. Přestože si na rozdíl od Naveru poradil alespoň přibližně s prvními dvěma větami, z celkového počtu 19 vět můžeme 14 označit jako chybných. Většina těchto vět neodpovídá gramatické struktuře zdrojového textu, v případě idiomu doplňuje víceméně náhodná slova jako *lion sex* pro překlad samotného slova idiom a *blood of the sea* pro zmíněný idiom. U Google překladače se dá předpokládat, že se se slovem idiom už v tréninkových datech setkal, nicméně se zde projevují vlastnosti statistiky a neurálních sítí, kde algoritmy nepracují se samostatnými slovy, nýbrž s pravděpodobnostmi, že se věta či část věty objeví v téhle konkrétní podobě. Nemá proto smysl zabývat se každou špatně přeloženou větou v tomto překladu. Zajímavostí ovšem je, že na rozdíl od Naveru můžeme v textu najít pokus o překlad slova *adžossi* (아저씨)⁶² pomocí zkratky Mr.

Ani Google Translate ani Naver Papago nedokázali z většiny zvládnout přenést smysl zdrojového textu do cílového jazyka. Překlady se ovšem liší prvotně v chybách, kterých se dopustily – zatímco Naver informace, se kterými si nedokáže poradit, spíše vynechává, Google je většinou nevynechává, ale výsledné překlady nenesou původní smysl. Chyb se navíc dopuští na

⁵⁷ Jsem člověk, který když některý z mých přátel z Pusanu zemřel, tam jel celou tu cestu, aby je vyprovodil.

⁵⁸ Vábit, lákat

⁵⁹ Kyslík

⁶⁰ Pobývali

⁶¹ Hong-sōpova žena

⁶² Pán

různých místech, což opět potvrzuje domněnku, že měly překladače při učení k dispozici odlišná data. Většina zmíněných chyb je dobrou reprezentací slabin překladačů založených na statistice. U neurálního překladače, kterými oba testované překladače jsou, by se ovšem měly díky rekurentním sítím naučit tyto chyby eliminovat. Oba překladače zde ovšem dokazují, že ani tato nejnovější technologie není schopná si s mnohými překážkami stále poradit.

3.4 Shrnutí výsledků

V přechozích podkapitolách jsme si na několika částech textu ukázali konkrétní vyprodukované překlady překladačů Naver Papago a Google Translate. Je zřejmé, že si přes aplikaci neurálních sítí nejsou překladače schopné vytvořit překlady na takové úrovni, aby byl zaručen přenos všech informací z korejštiny do angličtiny. Překladačům byly předloženy tři druhy textů, a to publicistický, komerční se zaměřením na techniku a přiřozená řeč. Beletrie ani poezie mezi příkladové texty nebyly zařazeny, protože zde nemůžeme počítat s pararélním překladem a zvlášť poezie není typ textu, na který by bylo při stavbě strojového překladače brán větší zřetel.

Mezi nejčastější chyby, se kterými jsme se v práci setkali, můžeme zařadit vynechávání částí textů, kdy překladač neznámou část jednoduše v cílovém jazyce vůbec nezachytí, což je chyba, která se vyskytovala častěji u překladače Naver Papago. Oba překladače pak vytvářely věty, kde kvůli neúplné informaci nebo ignoraci gramatické funkce vznikaly části, které původní význam věty buď nevystihly, nebo význam otočily. Nakonec se zde také často objevovaly chyby terminologické, kdy překladač nezvolil správný význam u vícevýznamového slova, což mohlo opět vést k tvorbě vět s odlišným významem než je ve zdrojovém textu. Na komerčním překladu technického rázu je také krásně ilustrováno, jaký problém můžou pro překladače znamenat delší souvětí, zvlášť, pokud se dají vyložit více způsoby. K chybám u takových struktur se zdá být náchylnější Google Translate, který při nejasném významu začal tvořit nesmyslné a negramatické věty, které na sebe významově ani strukturou nenavazovaly.

Po tomto textu se zdál být nejproblematičtější text poslední, který obsahuje přiřozenou řeč. Oba překladače selhaly v přesnosti, přičemž Google Translate přeložil chybně většinu textu. Ani jeden z překladačů nedokázal správně zpracovat přímou řeč.

Většina chyb, se kterými jsme se zde setkali, má zřejmou souvislost s nedostatky u statistiky a neurálních sítí jak bylo prezentováno v teoretické části, přesto jsme se setkávali i s chybami, které by podle teorie měly být díky algoritmizaci eliminovány.

Podle článku na webu Hankyoreh z roku 2017 byly oba překladače otestovány a srovnány z hlediska kvality překladu. Ve výsledku měl údajně Naver zvládat lépe slangové výrazy a zkratky, zatímco Google Translate si vedl lépe u dlouhých vět. Oba překladače pak vykazovaly dobré výsledky v případě vět krátkých.⁶³ Pokud tyto výsledky porovnáme s výsledky naší práce, můžeme s některými body souhlasit, a to především lepší schopnosti Naveru u zkratk. Naopak Google Translate na těchto ukázkách nevykazoval lepší schopnosti u dlouhých vět, ale spíše naopak. Musíme brát ovšem v potaz, že článek vyšel v době, kdy oba překladače teprve začínaly s novou technologií neurálních sítí, tudíž se dá předpokládat, že se jejich schopnosti překladu z větší části změnily.

Na závěr můžeme podotknout, že je zde zřejmý prostor pro další a podrobnější výzkumy, které by mohly srovnat aktuální stav těchto překladačů a zjistit původce opakujících se chyb,

⁶³ Google Translate fares better on long sentences, Naver Papago better on slang. The Hankyoreh [online]. Sŏul: The Hankyoreh, 2017 [cit. 2019-07-01]. Dostupné z: http://english.hani.co.kr/arti/english_edition/e_business/782430.html

4. Buocnost strojového překladač

Jak lze z předchozích kapitol vidět, technologie strojového překladač se za přibližně 70 let své existence výrazně posunula kupředu. Obrovský skok v přesnosti překladač zaznamenaly především neurální strojové překladače, což je případ porovnávaných překladačů Naver Papago a Google Translate. Od výše zmíněné soutěž v překladech mezi lidmi a počítači,⁶⁴ kde byly neurální překladače lidmi překonány, uplynuly již dva roky, což pro strojové překladače znamená 2 roky času na další učení a vylepšování schopností v překládání. I na malém vzorku příkladů uvedených v druhé části této bakalářské práce však můžeme vidět, že se překladače stále dopouští mnoha chyb, které by se ve smysluplném překladač neměly vyskytovat, a v mnoha případech obrací nebo zcela znemožňují uživateli pochopit původní smysl věty.

Strojové překladače jsou však dnes již nutnou součástí překladačů profesionálních lokalizačních firem. Nesmíme ovšem zapomenout na fakt, že pokud je cílem správný a kvalitní překladač, musí být strojem vyprodukované překladač následně zkontrolovány a zeditovány lidskými překladačeli. Vzhledem k současné kvalitě strojového překladač se zdá být nepravděpodobné, že by měly překladače v nejbližší době být schopné nahradit lidskou sílu. Musíme ovšem brát v potaz, že překladače mohou mít přístup k novým textům vloženým na internet, na kterých své schopnosti mohou dále vylepšovat, každá zpětná vazba uživatelů může přispět ke zkvalitnění překladačů a samotný výzkum umělé inteligence neboli AI se také stále posouvá směrem dopředu a může překladačům poskytnout dostatečné zázemí pro to, aby se dále vyvíjely samotné překladače. Že se technologie umělé inteligence bude stále vyvíjet dokazuje i fakt, že mnozí přední vědci, mezi které patří například Steven Hawking, vyjádřili z její nadějně budoucnosti obavy, že se stane natolik samostatnou, že by mohla vědomě ohrozit lidstvo.⁶⁵

Pokud se umělá inteligence dokáže jen přiblížit této úrovni, dá se předpokládat, že by ve strojovém překladač neměla mít problémy ani při jazycích natolik rozdílných, jako je korejština a angličtina. Přesto je tohle zatím hudba budoucnosti. Již teď je ovšem jasné, že si i přes své nedokonalosti strojový překladač najde uplatnění v mnoha oborech vzhledem k poptávce o překladač a lokalizaci v co nejkratším čase za co nejmenší cenu. Výdaje, které jsou potřebné na následné opravy a kontroly pak mohou být jedním z faktorů, které povedou ke zlepšení

⁶⁴ Humans beat AI in language translation [online]. Söul: Yonhap News Agency, 2017 [cit. 2019-07-01]. Dostupné z: <https://en.yna.co.kr/view/AEN20170221012500320>

⁶⁵ Stephen Hawking warns artificial intelligence 'may replace humans altogether' [online]. Londýn: Independent, 2017 [cit. 2019-07-01]. Dostupné z: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html>

překladače samotného. Nadějnou cestou také můžou být hybridní přístupy, tedy kombinace předchozích modelů strojového překladu s nejnovějšími technologiemi, a využít tak silné stránky každého z nich.⁶⁶ Musíme však brát na vědomí, že momentálně není cílem strojového překladu kompletně nahradit lidskou sílu, nýbrž poskytovat asistenci, a to zejména z důvodu stále většího objemu překladů, což se dnes týká právě i Koreje, která je představitelem technologicky vyspělé civilizace, což vede opět k další nutnosti produkovat více textů a dokázat je představit světu.

⁶⁶ Future of machine translation [online]. Connecticut: Technology Marketing Corporation, 2018 [cit. 2019-07-01]. Dostupné z: <https://www.tmcnet.com/topics/articles/2018/12/26/440734-future-machine-translation.htm>

5. Závěr

Práce na téma Strojový překlad a korejština byla nečekaně náročná vzhledem k prolínání více oborů, a to oboru jazykového i oboru informatiky. Strojové překladače využívají formální gramatiky pro zpracování přirozeného jazyka, zejména novější modely strojového překladu využívají více matematického pohledu a statistiky a neurální sítě pak využívají nejnovější technologii umělé inteligence. Hlavní výzvu zde proto představovalo hlubší zkoumání terminologie a teorie, které jsou pro pochopení celé problematiky nezbytnou nutností. Všechny tyto technologie jsem se v práci pokusila popsat zejména z pohledu překladu korejského jazyka, který i dnes představuje pro překladače výrazné problémy. K pochopení práce bylo nutné projít velké množství odborné literatury, která byla dostupná především v anglickém jazyce a korejském jazyce. Na korejštinu konkrétně se většinou zaměřovaly zejména disertační práce v korejštině. K problematice jsem nebyla schopná najít žádnou knižní korejskou publikaci, která by téma zpracovávala, a proto předpokládám, že taková práce zatím nebyla v korejštině vytvořena, nebo není dostupná širší veřejnosti. Dostupné korejské zdroje se obvykle zaměřují na konkrétní specifikum, jako jsou korejské pádové partikule ve strojovém překladu nebo práce představující některé z chyb, které ve strojovém překladu vznikají. Velká část se pak na problematiku zaměřuje spíše z informatického hlediska. V češtině byla jediným významným zdrojem kniha Ondřeje Bojara Strojový překlad našincům, našinci strojovému překladu, která představuje zejména problematiku statistických překladačů.

V první teoretické části jsem se pokusila využít znalosti nasbíraných ze zmíněných zdrojů a chronologicky představit jednotlivé významné typy překladačů, jejichž vliv je vidět i v nejnovějších dostupných modelech s využitím umělé inteligence. Pravidlové překladače byly založeny na sadě lingvistických pravidel, které musely být manuálně vytvořeny ve spolupráci s lingvisty, což je komplikovaný a drahý proces. Obtížné je také najít metodu, která by zabránila překladači vybírat nevhodná slova a fráze pro určitý typ kontextu, případně určování rolí ve větě u jmen, která v kombinaci se slovesem tvoří predikát. Statistické překlady by měly být řešením pro tento problém, avšak jak se ukázalo, problémem mohou být jak nedostatečně kvalitní bilinguální data, tak neschopnost poradit si s neznámými větnými strukturami. Neurální překladač by v teorii měl řešit jak problémy pravidlového překladače, tak statistického, jelikož podobně jako pravidlový překladač využívá lingvistická pravidla s tím rozdílem, že se je učí sám,

a na základě již naučených vzorců by měl být schopen tato pravidla implementovat i v neznámých podmínkách.

Na tuto část navazuje část praktická, ve které jsou stručně představeny veřejně dostupné překladače Naver Papago a Google Translate, které využívají nejnovější technologii neurálních sítí. Oba překladače byly uvedeny zhruba v podobnou dobu a využívají pomoci uživatelů pro další zlepšování překladu. Na jednoduchém testování s homonymním výrazem však bylo předvedeno, že se tyto překladače mohou značně lišit. V následujících kapitolách jsem se proto na vybraných korejských textech i na rozdíly v přístupu překladačů. Mezi úryvky textů jsem zařadila publicistický styl, komerční texty s technickým zaměřením, na kterých bylo možné zjistit, jak se překladače vypořádají s technickými zkratkami a složitější větnou strukturou, a nakonec také přepsané vyprávění reálného člověka, na kterém je možné uplatnit schopnosti překladačů u běžného přirozeného jazyka. Pro rozsah bakalářské práce jsem vybrala jen menší úseky textů, které umožňují demonstraci problémů, se kterými se strojové překladače potýkají. Nutno podotknout, že obtížné bylo samotné vyhledání textů, které by byly v korejštině i v angličtině a překlady se v některých částech nelišili, tudíž bylo v některých případech nutné vytvořit si vlastní referenční překlad.

Z těchto příkladů můžeme vyvodit, že si korejský překladač od společnosti Naver všeobecně vedl lépe. V bakalářské práci jsem se ovšem snažila o představení problémů, které se v textech vyskytují, nikoliv o vytvoření statistiky, k čemuž by bylo nutné využít mnohem většího množství bilnguálních textů.

V poslední části jsem uvedla krátkou úvahu na téma budoucnosti strojového překladu. Předchozí kapitola ilustruje, že strojový překladač má stále prostor pro zlepšování, a směr, kterým se ubývá nejen toto odvětví, ale i spousta technologií, je umělá inteligence.

Téma bakalářské práce jsem si zvolila, protože otázka strojového překladu je aktuální a dostatečně neprobádané téma. Zatímco tato bakalářská práce představuje jen povrchově, co si pod strojovým překladem představit a jaké má využití. Proto doufám, že tato práce poskytne mimo jiné i inspiraci pro ostatní, kteří se budou zajímat i o technický aspekt korejského jazyka.

Seznam použité literatury

- **Knižní zdroje:**

BOJAR, Ondřej. Čeština a strojový překlad: strojový překlad našincům, našiinci strojovému překladu. Praha: Ústav formální a aplikované lingvistiky, c2012. Studies in computational and theoretical linguistics. ISBN 978-80-904571-4-0.

DORR, Bonnie Jean. Machine translation: a view from the lexicon. Cambridge, Mass.: MIT Press, c1993. Artificial intelligence (Cambridge, Mass.). ISBN 0262041383. Dostupné z: <https://www.aclweb.org/anthology/J94-4009>

FELICE, Mariano. Linguistic indicators for quality estimation of machine translations. Universitat Autònoma de Barcelona & University of Wolverhampton, 2012.

HUTCHINS, W. John, HAROLD L. S. An introduction to machine translation. 2. vydání. Londýn: Acad. Press, 1992. ISBN 012362830X. S. 2.

HUTCHINS, W. John, The first decades of machine translation. Early years in machine translation, John Benjamins Publishing, 2000. ISBN 9027283710

LÜDELING, Anke; KYTÖ, Merja (ed.). Corpus linguistics. Walter de Gruyter, 2008.

Dostupné z:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.9749&rep=rep1&type=pdf>

QUAH, C. K. Translation and technology. New York: Palgrave Macmillan, 2006. ISBN 14-039-1832-5.

SAINT-DIZIER, P. (edit.). Predicative Forms in Natural Language and in Lexical Knowledge Bases. Dordrecht: Springer Netherlands, 1999. ISBN 9401727465

STREITER, Oliver. Linguistic modeling for multilingual machine translation. Als Ms. gedr. Aachen: Shaker, 1996. ISBN 3826519280. Dostupné z:

[https://s3.amazonaws.com/academia.edu.documents/1700111/57ptsf5r401m7ut9.pdf?response-](https://s3.amazonaws.com/academia.edu.documents/1700111/57ptsf5r401m7ut9.pdf?response-content-)

[disposition=inline%3B%20filename%3DLinguistic_Modeling_for_Multilingual_Mac.pdf&X-Amz-](https://s3.amazonaws.com/academia.edu.documents/1700111/57ptsf5r401m7ut9.pdf?response-content-disposition=inline%3B%20filename%3DLinguistic_Modeling_for_Multilingual_Mac.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-)

Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190701%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190701T001936Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=6204689077d8f4dcd22ef622b69cac217701a409526455371c623f0fccf75d16

WEISSER, Martin. Practical corpus linguistics: An introduction to corpus-based language analysis. John Wiley & Sons, 2016.

- **Zdroje v korejštině:**

BONG Ŏ-hü, et al. Čungganönö kigjepönjökpangsigül ühan öhudžisik pchjočhegje kwanhan jöngu. [Studie lexikálně znalostní reprezentace pro interlinguální strojový překlad]. Korea Information Science: Language and Technology. 1995. [online] Dostupné z: <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01064737&language=koŭKR>

ČCHÖ Dong-ik. An Analysis of Errors in Translating English Sentences with Inanimate Subjects into Korean by Machine Translation, The Joongwon Linguistic Society of Korea, 2013.10; ISSN 1975-8251

ČCHÖ Ho-čchöl (et al.). Kigje pönjökül ühan hangugö nonhang čhegje jöngu. [Studie korejského argumentačního systému pro strojový překlad] Hangugö üimihak 3, 1998. Dostupné z: <http://www.aistudy.co.kr/paper/pdf/koreanŭhong.pdf>

ČCHÖ Ho-čchöl (et al.). Takugö kigje pönjökül ühan čungganönö mohjönggwa pangpömnnon jöngu (Výzkum modelů a metod mezijazyka pro vícejazyčný strojový překlad). Kchömpchjutchöwa immunhak čchongsö 8, Korjö tähakkjo mindžokmunhwa jönguwön, 2003

ČCHWÖ Süng-kwon; Hong Mun-pchjo. Linguistic Modeling for Target Word Selection of Korean Adverbial Postpositions in a Multilingual MT-System. Hanguk čöngbokwa hakhö ono konghak jönguhö, 2001.10

HŬNG Mun-pchjo: Han-dok kigjepönjöküi täjögö sontchäk mundžee tähajö. [Problematika výběru strojového překladu pro jazyky korejštinu-němčinu]. Němčina a německá literatura, 2004 [online] dostupné z <http://210.101.116.28/Wŭfiles/kiss5/22101804ŭpv.pdf>

- I Min-häng, Či Kwang-san, Čöng So-u. Kigje Pönjök sistäm čchükdžöng čangčchi jöngu. [Výzkum testovacích případů pro strojový překlad]. Korean Society for Language and information. 1998, ISSN: 1226-7430
- KIM Čöng-rjöl: öhüginüngmunbop (Lexical-Functional Grammar)e küngöhan han-jöng jangqanggjang kigje pönjögüi önhakdžök kusöng. [Lingvistický design obousměrného korejsko-anglického systému strojového překladu založeného na lexikálně-funkční gramatice]. Korea National University of Education. 1999.
- KIM Deok-bok, et al. Jöngghan pönjöküi önhakdžök pchjöngga model jöngu. A Linguistic Evaluation of English-to-Korean Translation - Centered on Machine Translation. The Korean Society for Cognitive Science. 2001. 12
- KIM Jun-džöng, OK Čchöl-jong: Hangugö sösolöwa nonhjangdül saiüi üimijök. [korejské sémantické role subkategorizace]. Korean Language Processing Lab. University of Ulsan, Korean. 2014
- KIM Süng-üi, et al. Tagugo čadongtchongjök kisuldonghjang mit üngjong. [Trendy a aplikace v multilinguální automatické technologii pro překlad řeči]. 2011
- O Jöng-hun. Linguistische Probleme in der maschinellen Übersetzung -Topik und Fokus. Hanguk Dogohakhö, 2003, ISSN 1229-6333
- PÄ Či-jön. Hanjöng kigjepönjögüi kjök silhjön jöngu. [Studie o realizaci pádů v korejsko-anglickém strojovém překladu.] Sedžong Dähakkjo. 2010. 12.
- PAK Ok-su. Hanjöng kigjepönjögösö Stüi jugjöng tchükčinge ttarün pönjök orju punsök. [Analýza chyb podle typologické charakteristiky zdrojového textu v korejsko-anglickém strojovém překladu.] The Society For Humanities Studies In East Asia, 2017. 12
- PAK Söng, et al. Hanjöng kigjepönjogeso kjolčong tchüri haksübe ülhanhangugö pusakjök čosaüi üimi čungüisöng häso. Disambiguace sémantických rolí pro korejské adverbiální postpozice v korejsko-anglickém strojovém překladu, založená na rozhodovacím stromě. 2000. 6. Dostupné z: <https://bi.snu.ac.kr/Publications/Journals/Domestic/KISS27-6.pdf>

SI Čong-kon. Hangugo uimijök kjedžung kudžowa päjöl sunsö [Hierarchie a uspořádání sémantické role v korejštině]. Hanguk Indžikwahakhö, 1997 pp.239 – 248, HSS-Conference Papers

SI Čong-kon; Kim Won-kjöng; Ko Čchang-su. Jöng – Hanjöng kigje pönjök söngnüngpchjöngga pangan jöngu. [Studie metod hodnocení korejsko-anglického strojového překladu], Onowa čongbu, 2000-12, v.4, no.2, pp.1 – 26, HSS-Journal Papers; ISSN 1226-7430

- **Disertační práce**

CHUNG Hoojung. Statistical Korean Dependency Parsing Model based on the Surface Contextual Information. Disertační práce. Korea University. 2004 [obline]. Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.492.8537&rep=rep1&type=pdf>.

MYŠKA V.: REKURENTNÍ NEURONOVÉ SÍTĚ PRO KLASIFIKACI TEXTŮ, Brno 2018 [online], dostupné z https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=171408

PARK, Hyun Seok. Korean grammar using tags. Disertační práce. Philadelphia. 1994. [online] University of Pennsylvania Institute for Research in Cognitive Science Technical Report. Dostupné z: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1173&context=ircs_reports

VARIŠ D.: Japonsko-český strojový překlad. Praha, 2014. [online] Dostupné z: <http://www1.cuni.cz/~obo/vyuka/projekty/varis-japonsko-cesky-preklad.pdf>. Bakalářská práce. Ústav formální a aplikované lingvistiky. Vedoucí práce RNDr. Ondřej Bojar, Ph.D.

- **Ostatní články**

BENTIVOGLI, Luisa, et al. Neural versus phrase-based machine translation quality: případová studie, 2016.[online] Dostupné z: <https://arxiv.org/pdf/1608.04631.pdf>

CHÉRAGUI, Mohamed Amine. Theoretical overview of machine translation. Proceedings ICWIT, 2012, 160.

CHOI, Key-Sun, et al. An English-to-Korean machine translator: MATES/EK. In: Proceedings of the 15th conference on Computational linguistics-Volume 1. Association for

Computational Linguistics, 1994. [online] Dostupné z:
http://delivery.acm.org/10.1145/1000000/991907/p129-choi.pdf?ip=46.135.82.149&id=991907&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1561978379_62362d180e86dd7d7b936e998a66335d

CHOI, Sung-Kwon, et al. English-to-Korean Web Translator:“FromTo/Web-EK”. In: Machine Translation Summit VII. 1999. [online]. Dostupné z:
<https://pdfs.semanticscholar.org/471c/bc9d82458c06b92bad1a3e9bbe6d59324732.pdf>

CHOI, Sung-Kwon, et al. Hybrid approaches to improvement of translation quality in Web-based English-Korean machine translation. In: Proceedings of the 17th international conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 1998. [online] Dostupné z: http://delivery.acm.org/10.1145/990000/980886/p251-choi.pdf?ip=46.135.82.149&id=980886&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1561977787_5d65def0516b8d512489f34238a8500a

CHOI, Sung-Kwon; KIM, Young-Gil. Customizing an English-Korean Machine Translation System for Patent Translation. *Záznamy z 21. Pacific Asia Conference on Language, Information and Computation*. 2007. p. 105-114. [online] Dostupné z:
<https://www.aclweb.org/anthology/Y07-1010>

CLIFFORD, Ray, et al. The Effect of Text Difficulty on Machine Translation Performance--A Pilot Study with ILR-Rated texts in Spanish, Farsi, Arabic, Russian and Korean. DEFENSE LANGUAGE INST MONTEREY CA, 2004. [online]. Dostupné z:
<https://apps.dtic.mil/dtic/tr/fulltext/u2/a511696.pdf>

EGEDI, Dania, et al. Korean to english translation using synchronous tags. University of Pennsylvania, 1994. [online]. Dostupné z: <https://arxiv.org/pdf/cmp-1g/9410023.pdf>

ERIGUCHI A. Learning to Parse and Translate Improves Neura Machine Translation. , TSURUOKA Y. CHO Kyunghyun. Tokyo: The University of Tokyo. 2017. [online] Dostupné z <https://arxiv.org/pdf/1702.03525.pdf>

HAN, C. H., Lavoie, B., Palmer, M., Rambow, O., Kittredge,... & Kim, M. Handling structural divergences and recovering dropped arguments in a Korean/English machine translation system. Conference of the Association for Machine Translation in the Americas. Springer, Berlin, Heidelberg. 2000, 10.

HAN, Chung-hye, et al. Development and evaluation of a Korean treebank and its application to NLP. SIMON FRASER UNIV BURNABY (BRITISH COLUMBIA) DEPT OF LINGUISTICS, 2002. [online] Dostupné z <https://apps.dtic.mil/dtic/tr/fulltext/u2/a457100.pdf>

HONG, Munpyo; KIM, Chang-Hyun. Controlled Korean for Korean-English MT. on Language, Information and Computation. 2008. [online] Dostupné z: <https://www.aclweb.org/anthology/Y08-1040>

HUTCHINS, John. The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954. [online] Dostupné z:: [http://sts.bdtf.hu/btk/flli/romanisztika/OKTATSARS%20DOCENDI/TANANYAGOK%20\(O KTATÓ%20SZERINT\)/ANTONIO%20SCIACOVELLI/TRADUTTORI%20INTERPRETI/c orrispondenza%20commerciale/traduzione%20elettronica_articolo_english.pdf](http://sts.bdtf.hu/btk/flli/romanisztika/OKTATSARS%20DOCENDI/TANANYAGOK%20(O KTATÓ%20SZERINT)/ANTONIO%20SCIACOVELLI/TRADUTTORI%20INTERPRETI/c orrispondenza%20commerciale/traduzione%20elettronica_articolo_english.pdf)

HUTCHINS, W. John. "Example-Based Machine Translation: A Review and Commentary." Machine Translation 19, no. 3/4 2005. [online] Dostupné z:: <http://www.jstor.org/stable/20060481>.

KIM, Sung-Dong. Syntactic Category Prediction for Improving Translation Quality in English-Korean Machine Translation. In: Záznamy 23rd Pacific Asia Conference on Language, Information and Computation, Volume 2. 2009. [online] Dostupné z: <https://www.aclweb.org/anthology/Y09-2032>

LEE, Ki-Young; KIM, Young-Gil. Applying Statistical Post-Editing to English-to-Korean Rule-based Machine Translation System. In: Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation. 2012. [online]. Dostupné z: <https://www.aclweb.org/anthology/Y12-1034>

LIU, Ding; GILDEA, Daniel. Semantic role features for machine translation. In: Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010. [online] Dostupné z:

http://delivery.acm.org/10.1145/1880000/1873862/p716-liu.pdf?ip=46.135.82.149&id=1873862&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1561978025_8652e8abfd668631439c0e3b91674c21

MIKOLOV, T.; Yih, W.-t.; Zweig, G.: Linguistic Regularities in Continuous Space Word Representations. Zázpis z konference 2013 North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2013. [online] Dostupné z: <http://www.aclweb.org/anthology/N13-1090>

NOONE, Grace. Machine Translation A Transfer Approach. Computer Science, Linguistics and a Language (CSLL) Department, University of Dublin, Trinity College, 2003.

PARK Se-Young; OH, Gil-Rok. Machine translation in Korea.7. MT Summit, 1999, 100-104.

RYU, B. R., & Kim, Y. K. Light Verbs and Sino-Korean Verbal Nouns in Korean-English Machine Translation. Taejon, 1999. Chungnam National University, Electronics and Telecommunications Research Institute (ETRI).

RYU, Byong-Rae, et al. FromTo K/E: A Korean-English machine translation system based on idiom recognition and fail softening. In: MT Summit. 1999. [online] Dostupné z: https://www.researchgate.net/profile/Byong-Rae_Ryu/publication/2335450_KE_A_Korean-English_Machine_Translation_System_based_on_Idiom_Recognition_and_Fail_Softening/links/004635214d90b7ec4a000000.pdf

SENNRICH, Rico; HADDOW, Barry. Linguistic input features improve neural machine translation. 2016. [online] Dostupné z: <https://arxiv.org/pdf/1606.02892.pdf>

SEO, Young, et al. Korean-to-English Machine Translation System based on Verb-Phrase:'CaptionEye/KE'. In: Proceedings of the Korea Information Processing Society Conference. Korea Information Processing Society, 2000.

STYMME Sara, AHRENBERG Lars, Using a Grammar Checker for Evaluation and Postprocessing of Statistical Machine Translation. Linköping University, Švédsko, 2010. [online] Dostupné z:

https://www.researchgate.net/publication/220746465_Using_a_Grammar_Checker_for_Evaluation_and_Postprocessing_of_Statistical_Machine_Translation

WEINSTEIN, Clifford J., et al. Automated English-Korean Translation for Enhanced Coalition Communications. Lincoln Laboratory Journal, 1997. [online] Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.62.6997&rep=rep1&type=pdf>

- **Internetové stránky**

Humans Defeat Machine Translation In Korean Competition. Minneapolis [online]. 2019 United Language Group. All Rights Reserved. [cit. 13.06.2019]. Dostupné z: <https://unitedlanguagegroup.com/blog/humans-defeat-machine-translation-korea-competition/>

Zero-Shot Translation with Google's Multilingual Neural Machine Translation System [online]. Google, 2016. Dostupné z: <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>

Comparative Evaluation: Statistical vs. Neural Machine Translation | tilde.com. Tilde | tilde.com [online]. Dostupné z: <https://www.tilde.com/about/news/316>

South Korea's Naver launches official version of AI translator 'Papago'. THE NATION, 2017 [online]. Dostupné z: http://www.nationmultimedia.com/detail/Startup_and_IT/30321483

Naver updates translation service for Korean honorifics. The Korea Herald, 2019 [online]. Dostupné z: <http://www.koreaherald.com/view.php?ud=20190117000606>

Naver's Papago more popular than Google Translate among Koreans. The Korea Herald, 2019 [online]. Dostupné z: <http://www.koreaherald.com/view.php?ud=20190423000792>

Google Translate: How does the search giant's multilingual interpreter actually work?. Independend [online]. Londýn: Independend, 2018 [cit. 2019-07-01]. Dostupné z: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-translate-how-work-foreign-languages-interpreter-app-search-engine-a8406131.html>

Google Translate fares better on long sentences, Naver Papago better on slang. The Hankyoreh [online]. Söul: The Hankyoreh, 2017 [cit. 2019-07-01]. Dostupné z: http://english.hani.co.kr/arti/english_edition/e_business/782430.html

Humans beat AI in language translation [online]. Söul: Yonhap News Agency, 2017 [cit. 2019-07-01]. Dostupné z: <https://en.yna.co.kr/view/AEN20170221012500320>

Stephen Hawking warns artificial intelligence 'may replace humans altogether' [online]. Londýn: Independend, 2017 [cit. 2019-07-01]. Dostupné z: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html>

Future of machine translation [online]. Connecticut: Technology Marketing Corporation, 2018 [cit. 2019-07-01]. Dostupné z: <https://www.tmcnet.com/topics/articles/2018/12/26/440734-future-machine-translation.htm>

FORMÁLNÍ GRAMATIKA. CzechEncy – Nový encyklopedický slovník češtiny [online]. Brno: Vladimír Petkevič, Karel Oliva, 2017 [cit. 2019-05-19]. Dostupné z: <https://www.czechency.org/slovník/FORMÁLNÍ%20GRAMATIKA>

What is Machine Translation? Rule Based Machine Translation vs. Statistical Machine Translation [online]. Paris: SYSTRAN, 2019 [cit. 2019-07-01]. Dostupné z: <http://www.systransoft.com/systran/translation-technology/what-is-machine-translation/>

A Gentle Introduction to Neural Machine Translation [online]. Vermont Victoria: © 2019 Machine Learning Mastery Pty. Ltd. All Rights Reserved., 2017 [cit. 2019-07-01]. Dostupné z: <https://machinelearningmastery.com/introduction-neural-machine-translation/>

Is neural machine translation always the best option? [online]. Ghent: YAMAGATA EUROPE, 2018 [cit. 2019-07-01]. Dostupné z: <https://www.yamagata-europe.com/en-gb/blog/is-neural-machine-translation-always-the-best-option>

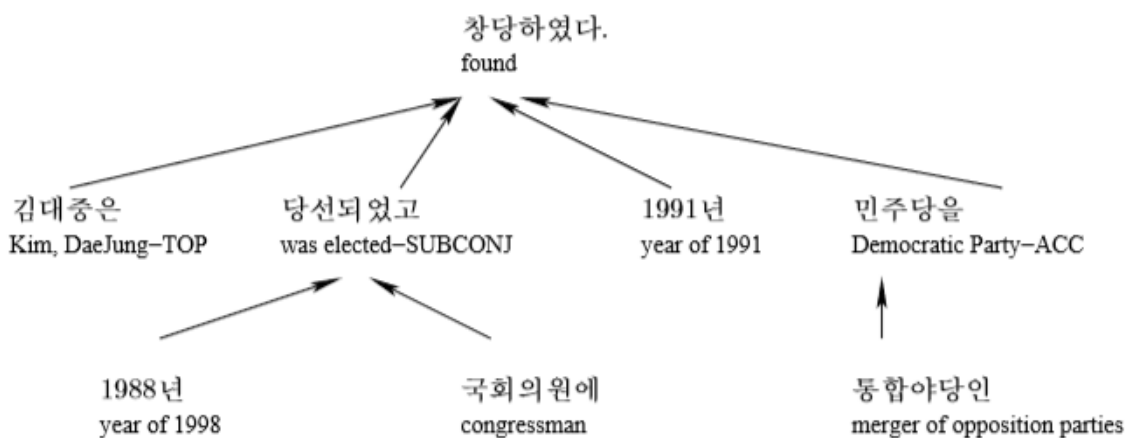
How does Neural Machine Translation work?: The representation of meaning in Neural, Rule-Based and Phrase-Based Machine Translation [online]. Paris: SYSTRAN, 2016 [cit. 2019-07-01]. Dostupné z: <http://blog.systransoft.com/how-does-neural-machine-translation-work/>

Keras LSTM tutorial – How to easily build a powerful deep learning language model [online]. Adventures in Machine Learning, 2018 [cit. 2019-07-01]. Dostupné z: <https://adventuresinmachinelearning.com/keras-lstm-tutorial/>

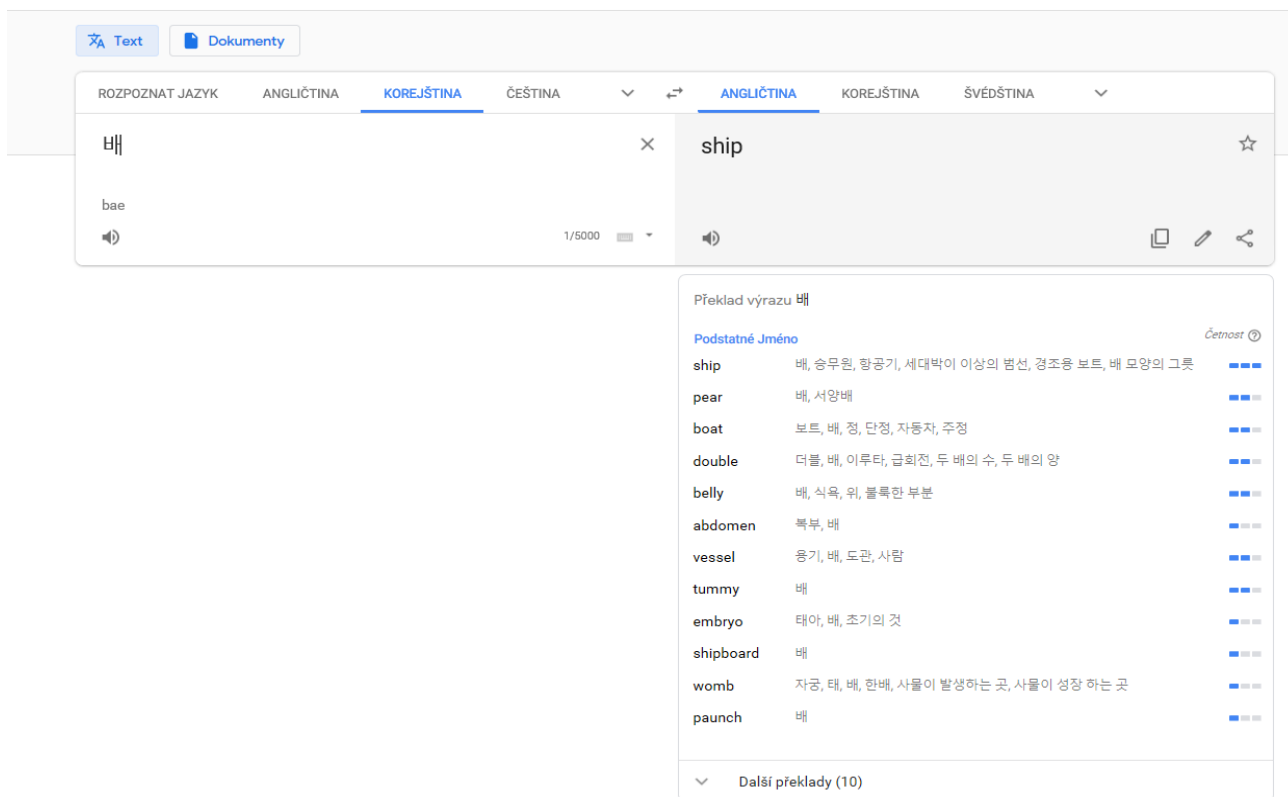
Illustrated Guide to LSTM's and GRU's: A step by step explanation [online]. Michael Nguyen, 2018 [cit. 2019-07-01]. Dostupné z: <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>

Пřílohy

Obr. 1



Obr. 2



Odeslat zpětnou vazbu