

## Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: Doc. RNDr. Vladislav Kuboň, Ph.D.

Jméno a příjmení autora práce: Bc. Soňa Kyselová

Název práce: Srovnání (a historická podmíněnost) výstupů ze strojových překladačů

---

Předkládaná diplomová práce se zabývá problémem porovnání výstupů z nejoblíbenějších strojových překladačů Bing a Google Translate a jejich vývoje v průběhu zhruba jednoho roku. Jedná se o zajímavé a aktuální téma, jehož užitečnost ještě podtrhuje fakt, že minimálně u překladače Google Translate v průběhu sledovaného roku došlo pravděpodobně ke změně metody překladu ze statistické na neuronovou. Dalším pozitivním faktorem je volba jazykového páru, místo obligátního páru angličtina – čeština autorka zvolila méně frekventovaný pár němčina – čeština, což atraktivitu tématu ještě zvyšuje. Práce se skládá ze tří hlavních kapitol, úvodu, shrnutí výsledků a závěru. Zmíněné tři kapitoly jsou obsáhlé, první z nich se věnuje teoretickým východiskům, jmenovitě historii strojového překladu, představení překladačů Bing a Google Translate a rozboru metod hodnocení kvality překladu. Úroveň výkladu je zvolena dobře, autorka čtenáře seznamuje se základní problematikou přehlednou formou bez zbytečných technických detailů. Jediné, co by se této kapitole dalo vytknout, je citace nepůvodních zdrojů (např. diplomové práce Mirka Týnovského z roku 2007 místo původních článků) a občasně chybějící vysvětlení některých pojmů (např. transdukčního slovníku na str.16 či synchronní gramatiky na str. 17). Další kapitola se zabývá metodologickým základem experimentu. Autorka formuluje výchozí hypotézy, představuje dvoustupňový model hodnocení a seznamuje čtenáře s empirickým materiálem. Výběr testovacích textů je metodicky dobře zdůvodněn, co však poněkud zaráží, je jejich délka. Texty většinou obsahují pouze něco přes 300 slov, což nedává záruku, že budou zachyceny drobné změny kvality překladu, týkající se např. nějakých okrajových jazykových jevů nebo konstrukcí. Na druhou stranu je třeba přiznat, že ve světle detailního zpracování výsledků a klasifikace chyb je jen obtížné si představit, že by bylo možné v rámci diplomové práce podobně kvalitně zpracovat řádově rozsáhlejší texty. Kvalita je zde upřednostněna před kvantitou, což ale poněkud relativizuje dosažené výsledky.

Ve třetí hlavní kapitole autorka podrobně rozebírá získané výsledky pro jednotlivé druhy textu. Tato část je bohatě ilustrována grafy, znázorňujícími vývoj kvality překladu v čase. Některé výsledky jsou překvapivé, protože ukazují, že postupným vylepšováním systémů se překlad některých konkrétních vět může zhoršit natolik, že se to projeví i na celkovém skóre daného testovacího textu. Výsledky také potvrzují předpoklad, že v průběhu sledovaného období došlo ke změně kvality překladu u systému Google Translate, kde byly pravděpodobně nasazeny neuronové sítě.

V závěru práce autorka shrnuje dosažené výsledky a zasvěceně je komentuje. Její vývoje jsou dobře zdůvodněné, nezmiňují však jeden faktor, který má na kvalitu překladu také vliv. Jde o to, že velké firmy jako Microsoft a Google pro zkvalitnění překladačů průběžně doplňují korpusy paralelních textů, a tak se klidně mohlo stát, že testovací texty se v průběhu sledovaného období mohly dostat do trénovacích korpusů, což typicky má za následek zvýšení kvality překladu – zejména statistické systémy výrazně lépe překládají texty, které systém už jednou „viděl“. Domnívám se, že zejména řeč kancléřky Merkelové na důležitém summitu byla pravděpodobně přeložena do češtiny a publikována, díky čemuž se mohla dostat do paralelních trénovacích dat.

Po jazykové stránce je diplomová práce dostatečně kvalitně zpracována, text je plynulý a dobře čitelný. Autorka se však nevyvarovala některých překlepů (např. *Prag Dependency Treebak* místo *Prague Dependency Treebank* na str. 17, *PROMPT* místo *PROMT* na str. 27, *NMT* místo *HMT* na str. 25 apod.). Tyto drobné nedostatky však nemají vliv na celkově velmi dobrý dojem z předkládané práce.

### Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě a navrhuji známku výborně až velmi dobře.

V Praze dne: 22.1.2018

Podpis: Vladislav Kuboň, v.r.