

POSUDEK BAKALÁŘSKÉ PRÁCE

Název práce: Hodnocení automatického překladu SMT systémů (Google translate, Bing) z francouzštiny do češtiny: dezambiguace vybraných polysémních výrazů

Diplomantka: Kateřina Hodková

Pracoviště: Ústav romanistiky FF JU v Českých Budějovicích

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

Rozsah: 90 stran vč. bibliografie a vlastních příloh

Cílem předkládané bakalářské práce je vyhodnotit, s jakou úspěšností se podaří automatickým SMT překladačům provést dezambiguaci vybraných polysémních výrazů v kontextu. Úspěšnou dezambiguací se zde rozumí výběr adekvátního ekvivalentu daného polysémního výrazu v kontextu bez ohledu na ostatní faktory (morfologická podoba výrazu, syntaktické zapojení do kontextu, atp.).

V úvodní části práce najdeme stručné fundované shrnutí teorie strojového překladu (s. 2-6), principů dezambiguace v počítačové lingvistice (s. 8-16) a teorie týkající se vyhodnocování kvality strojového překladu (s. 17-21). Autorka zde čerpala z cizojazyčné literatury a prokázala, že se v dané problematice orientuje. Drobnou kritickou poznámku bych měl k pasáži o dezambiguaci prováděné člověkem (s. 11), kde autorka čerpá aktuální informace z pramenu starého téměř třicet let (Hirst, 1987). Za velmi přínosné naopak považuji to, že teorii vhodně propojuje s vlastním výzkumem, promýšlí dopad jednotlivých hypotéz a experimentálně je ověřuje. Zjistila například, že použité překladače podle všeho nepracují s principem „one sense per discourse“ (s. 24); na konkrétním příkladu ověřila, že používání angličtiny jako interlingvy by mohlo vysvětlit zdánlivě nelogický překlad výrazu *règlement* (s. 25-26); konstatovala také, že slova v bezprostředním kontextu nepomohou dezambiguovat sloveso *acquitter* (s. 23). V posledně jmenovaném příkladu by bylo zajímavé možná jít ještě dále a klást si otázku, jak velká vzdálenost dezambiguujícího výrazu by se na dezambiguaci mohla projevit. V pramenech o korpusové statistice se totiž uvádí, že pravděpodobnost souvýskytu slov je extrémně vysoká (a tedy prediktabilní) do vzdálenosti 3-4 grafických slov, ale s větší vzdáleností strmě klesá.

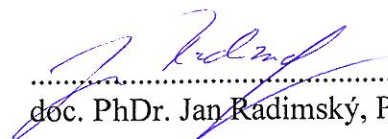
Pro vlastní výzkum si vybrala deset francouzských výrazů, které se v oblasti práva a obchodu používají alespoň ve dvou významech. Na příkladech kontextualizovaných vět pak ověřovala, zda SMT systém přeložil (dezambiguoval) testovaný výraz korektně. Celkové výsledky jsou přehledně shrnuty v tabulce na s. 30, následují komentáře k překladu jednotlivých slov. Z metodologického hlediska mám k empirické části jednu kritickou připomínku. Z komentářů např. na s. 34 je patrné, že si diplomantka určila sadu „přijatelných“ a „nepřijatelných“ překladových ekvivalentů předem podle ekvivalentů slovníkových, aniž by znala kontext. Např. pro *article* chápe jako přijatelné ekvivalenty slova *zboží* a *položka*, nikoli už *předmět* (v rámci stejné množiny příbuzných významů). Takto mimo kontext ale ekvivalenty určovat nelze. Ve vzorku na s. 57 pak totiž čteme, že v příkladu 12 hodnotí negativně překlad *nádobí* pro *articles culinaires* (ačkoli z kontextu je zřejmé, že právě o nádobí jde); podobně na s. 58 v příkladu 17 hodnotí jako přijatelný překlad *položky* pro *articles de mode*, ačkoli v daném kontextu by šel překlad *položky* použít jen velmi obtížně. Drobnou poznámku mám ještě k výrazu *kolokace*, jak jej autorka užívá na s. 29 a 33 – podle kontextu soudím, že jde o *fráze* jakožto jednotky SMT překladu založeného na frázích (tzv. *phrase-based SMT model*), jak o něm sama autorka píše na s. 4.

Uvedené připomínky nic nemění na tom, že autorka předložila kvalitní a rozsáhlou práci, která tvůrčím způsobem zpracovává náročné téma.

Po formální stránce práce odpovídá požadavkům.

Závěr:

Předkládanou bakalářskou práci doporučuji k obhajobě a hodnotím jako **výbornou**.


.....
doc. PhDr. Jan Radimský, Ph.D.

V Českých Budějovicích, dne 20. května 2016