

## Posudek vedoucího diplomové práce

Miroslav Týnovský: Využití lingvistických informací při EBMT

Diplomová práce Miroslava Týnovského se zabývá problémem využití lingvistických informací při automatickém překladu založeném na příkladech (Example Based MT – EBMT). Metoda strojového překladu pomocí příkladů, původně popsána prof. Nagaem počátkem 80. let, byla využita v mnoha experimentech a její nejrůznější varianty byly popsány v celé řadě prací. Žádný z těchto experimentů a článků se však nezabýval češtinou, proto hlavním úkolem této práce bylo jednak prozkoumat možnost vytvoření jednoduchého experimentálního systému obsahujícího češtinu a obohatit tento systém o využití lingvistických informací.

Diplomová práce je přehledně uspořádaná, skládá ze šesti kapitol, z nichž první obsahuje úvod a poslední závěru. Ve druhé kapitole autor čtenáře stručně seznamuje s historií strojového překladu a snaží se na základě literatury vymezit strojový překlad na základě příkladů. Ve třetí popisuje vytváření dvou paralelních korpusů pro experimenty s překlady z češtiny do němčiny a angličtiny. Čtvrtá kapitola popisuje vytváření systému EBMT a uvádí čtenáře do problematiky hodnocení kvality automatického překladu. Pátá kapitola se zabývá experimenty s využitím lingvistických informací ve vytvořeném systému EBMT. Práce je doplněna třemi přílohami a datovým nosičem.

Po obsahové stránce práce jednoznačně splňuje zadání, v některých ohledech jej dokonce překračuje (pečlivé zpracování hned dvou jazykových párů a veškerá časově náročná a důkladná příprava dat pro tyto jazykové páry). Hlavním přínosem této diplomové práce je přehledný popis metody EBMT, a příprava a provedení experimentů, které dávají velmi dobrý základ pro další výzkum. Autor v práci také naznačuje i konkrétní směry, jakými by se další podrobnější zpracování tématu EBMT mezi typově odlišnými jazyky mohlo dále ubírat. I když vylepšení využívající lingvistické údaje v morfologii, která autor implementoval, nijak podstatně nezlepšují relativně nízkou kvalitu překladu metodou EBMT, je z práce celkem jasné, že hlavní problém leží jinde – v kvalitnějším alignmentu a rekombinaci, která je velmi podstatným, přitom v literatuře hojně opomíjeným prvkem, jenž má zřejmě největší vliv na kvalitu překladu. Velmi cenným vedlejším výsledkem práce jsou paralelní data pro pár němčina-čeština. Autor dokázal pro svou práci využít parseru skupiny NaTS z Univerzity v Hamburku, což bylo první praktické použití tohoto spíše experimentálního parseru někým mimo okruh pracoviště, na kterém parser vznikl. Je třeba také ocenit velmi dobrou práci s literaturou, kdy autor dokázal využít i velmi čerstvých a aktuálních výsledků (ACL 07). Dokázal také samostatně vyhledat a přiblížit čtenáři méně známou metodu vyhodnocování kvality překladu METEOR.

Po formální a typografické stránce je práce také na velmi dobré úrovni, neobsahuje gramatické ani stylistické chyby. Literatura je použita v hojné míře a je správně citována. Jedinou výjimkou, kde by bylo vhodné přesněji citovat, je str. 26, kde chybí citace McDonaldova parseru a systému TectoMT. Příložené CD obsahuje veškerý materiál použitý v diplomové práci s výjimkou autorsky chráněného korpusu PCEDT.

Vzhledem k výše uvedeným skutečnostem mohu konstatovat, že práce splňuje po formální i obsahové stránce veškeré nároky kladené na diplomovou práci z oboru lingvistiky a strojového překladu.

V Praze dne 3.9.2007

Vladislav Kuboň  
ÚFAL MFF UK