

Petr Němec: Automatic analysis of temporal relations within a discourse and its application within a machine translation framework.

Posudek doktorské disertační práce (Petr Sgall)

Předložená práce předkládá autorův originální systém automatické analýzy časových vztahů v české promluvě a charakterizuje možné uplatnění tohoto systému při syntéze anglických vět v rámci strojového překladu z češtiny.

Práce, psaná anglicky, je rozdělena do sedmi kapitol:

Kapitola 1: Introduction (s. 7-13) stručně popisuje projekt Prague Dependency Treebank (PDT) s jeho různými rovinami (tektogramatickou, morfematickou a pomocnou analytickou) i pražský česko-anglický závislostní korpus.

Kapitola 2: Annotations of temporal relations (s. 14-35) je věnována rozboru časových vztahů v promluvě a charakteristice autora anotovaného korpusu časových vztahů ve vztahu k některým dosavadním koncepcím.

Kapitola 3: Automatic analysis of temporal relations (s. 36-59) obsahuje z lingvistického hlediska jádro práce; rozbírá především přístup J. Panevové k rekurzivní povaze časových vztahů a k opozici základního a relativního času. Probírají se tu možnosti upřesnění a rozšíření tohoto přístupu, s tím výsledkem, že jde o přístup velmi spolehlivý.

Kapitola 4: Text generation within a machine translation framework (s. 60-70) uvádí model syntézy anglických časů na tektogramatické, tzv. povrchové i morfematické úrovni i jeho empirické zhodnocení.

Kapitola 5: Tree searching/rewriting formalism (s. 71-82) se zabývá zejména popisem formalismu pro vyhledávání komplexních jazykových jevů ve stromech a přeměnou tektogramatických stromů na jiné roviny, a také implementací tohoto formalismu.

Kapitola 6: Using temporal information for text generation (s. 83-90) stručně charakterizuje lingvistické problémy dosud otevřené (rozlišení jednoduchých a perfektních i průběhových časů) i další (plusquamperfektum, nepřímá řeč) a jejich zpracování.

Kapitola 7: Conclusion (s. 91-92) podává stručné závěry celé práce.

V přílohách (s. 93-117) jsou obsaženy gramatika analýzy časových vztahů, odhad rozsahu časových funkcí, dvojjazyčný seznam subfunktorů, přehled implementovaného softwaru a rozsáhlá bibliografie.

Práce je velmi zdařilým pojednáním o daných problémech a v mezích možností je velmi soustavně zpracovává jak lingvisticky, tak algoritmicky, i v podobě automatického systému. Moje připomínky se netýkají podstaty těchto problémů, které jsou tu zpracovány velmi dobře, ale spíš záležitostí okrajových a jednotlivých míst.

Obecnější připomínka je jen jedna, snad relevantní hlavně z hlediska eventuální další publikace textu; jde o určitou neúplnost obsahové koherence disertace: její text věnuje

z lingvistického hlediska pozornost především časovému systému češtiny, zatímco angličtina je tu pojednána jen na základě porovnání s češtinou. Má to ovšem své dobré důvody, dané zaměřením práce na strojový překlad z češtiny do angličtiny. Je tu obsaženo i rozlišení problémů, které jsou při tomto postupu řešitelné a které nikoli (v kap. 6).

Poznámky k jednotlivým místům:


K odd. 2.1.1 – úvaha o vidu by mohla být prohloubena jednak bližším porovnáním klasifikace dějů u Vendlera ad. s opozicí dokonavosti v češtině (např. stav se snad vyjadřuje jen slovesnými tvary dokonavými, činnost je primárně nedokonavá) a jednak přezkoumáním hypotézy, že se tato opozice v češtině vyjadřuje prostředky obvykle sloužícími lexikálnímu odvozování, ale sama o sobě (na tektogramatické rovině) má povahu gramatickou;

k odd. 4.3 (u pozn. 1) – jestli dobře rozumím, užívají se i hodnoty gramatémů (např. času).

Celkově je práce dobrým dokladem autorovy schopnosti samostatné vědecké činnosti a velmi významným přínosem k lingvistickému i infromatickému zpracování otázek slovesného času. Důležité je kritické ověření předcházejících výsledků (zejména J. Panevové) i autorův vlastní velmi závažný přínos ke zpracování vztahů mezi časovými významy v češtině a v angličtině jak po stránce lingvistické, tak i algoritmicky a pro účely strojového překladu.

Jsem proto jednoznačně přesvědčen, že práce po všech stránkách splňuje požadavky na doktorské disertace kladené a navrhuji, aby jejímu autorovi, RNDr. Petru Němcovi, byl po úspěšné obhajobě udělen titul Ph.D. (doktor) v oboru i-3 „Matematická lingvistika“.

V Praze 24. 10. 2007



sgall@ufal.mff.cuni.cz