

**Mgr. Jiří Semecký:**  
**Verb Valency Frames Disambiguation**  
Disertační práce

(vyjádření školitele)

Předložená disertační práce se zabývá v zásadě rozlišováním významu sloves podle předem určeného repertoárů významů. Každý význam je přitom určen v korpusu jednoznačně pomocí odkazu do valenčního slovníku, a v tom to slovníku je i řada pomocných a do jisté míry využitelných údajů charakterizujících syntaktické a sémantické chování daného významu slovesa. Vzhledem k existenci korpusů s análně anotovanými významy uvedeného typu bylo možno postupovat pomocí statistických metod a výsledky řádně vyhodnotit pomocí standardních, světově uznávaných evaluačních kritérií.

Práce je rozdělena do osmi kapitol a dvou příloh (A a B). V úvodu autor charakterizuje cíl disertace a popisuje strukturu vlastní práce. Ve druhé kapitole popisuje metody strojového učení použité v práci; tato kapitola je důležitá pro případné následovníky z hlediska reimplementace systému. Třetí kapitola popisuje detailně cíl práce, a definuje použití termínů „word sense“ a čím je v práci aproximován (tj. právě valenčními rámci). Čtvrtá kapitola obsahuje popis použitých korpusů (tj. trénovacích a testovacích dat). V páté kapitole autor uvádí další přístupy k valenci i rozlišování významu u dalších autorů, jak pro češtinu, tak i pro jiné jazyky. Kapitoly 6 a 7 potom představují vlastní jádro práce, resp. vlastní přínos disertanta k danému cíli práce; v šesté kapitole jsou uvedeny všechny lingvistické vlastnosti daného slova a jeho kontextu, které jsou podle mínění autora relevantní pro řešení daného problému (ty pak tvoří tzv. „features“) a které také použil v programech pro strojové učení na základě daných korpusů. V kapitole sedmé jsou pak popsány relevantní experimenty (včetně přesného nastavení těchto experimentů a popisu použitých dat) a jejich evaluace pomocí standardních metrik. Všechny použité metody, ať už se liší principiálně nebo nastavením parametrů (features) jsou pak na základě této evaluace podrobně porovnány.

Práce je psána velmi dobrou angličtinou a dobře se čte, jednotlivé části textu na sebe logicky navazují, samotná struktura práce je dobře zvolena. Klíčová kapitola 6 velmi dobře popisuje zvolené jazykové i další „features“, i když z ní není zcela jasné, jakým postupem k nim autor došel (je ovšem zcela zřejmé, že právě tato část je v každém podobném systému právě oním „art“, neboli vědeckou intuicí, která pak dané výsledky má nebo nemá). Z kapitoly sedmé je ovšem víc než zřejmé (protože autor si dal ve svou práci s různými systémy kombinacemi oněch „features“), že výsledky se dostavily; zlepšení proti obvyklé baseline je nejen signifikantní, ale i velmi výrazné.

Práce je formálně zcela v pořádku, seznam citované literatury je dostatečný.

Předloženou disertaci Mgr. Jiřího Semeckého tedy považuji za práci, která splňuje nároky kladené na disertační práci a doporučuji ji k obhajobě.

Praha, 14.8.2007

