

Cieľom tejto diplomovej práce je navrhnúť, implementovať a empiricky evaluovať klasifikátory pre rozpoznávanie sémantických patternov anglických slovies. Ako trénovacie a testovacie údaje používame konkordancie z pilotnej kolekcie 30 anglických slovies, ktorá bola spracovaná metódou Corpus Pattern Analysis. Modely klasifikátorov tvoríme pomocou algoritmov strojového učenia s učiteľom. Experimentujeme s rozhodovacími stromami, algoritmom k najbližších susedov (kNN), podpornými vektormi (SVM) a Adaboostom. V práci sa, okrem iného, zameriavame na návrh vhodnej množiny rysov pre strojové učenie (feature selection). Experimentujeme s množinami morfo-syntaktických i sémantických rysov. Naše výsledky ukazujú, že morfo-syntaktické rysy sú najdôležitejšie pre sémantickú desambiguáciu, hoci pre niektoré slovesá hrajú sémantické rysy dôležitú úlohu.