

## Abstrakt

Syntaktická analýza je jedním z nejdůležitějších kroků počítačového zpracování přirozených jazyků. V této práci se zaměřujeme na formalismus závislostní gramatiky, protože jeho hlavní principy, zejména vztah řídicího a závislého uzlu, se ukázaly užitečné pro řadu rozdílných jazyků, se zvláštním zřetelem na vysvětlení slovosledu a vztahu mezi povrchovou strukturou a významem. Většina moderních efektivních algoritmů závislostní syntaktické analýzy je založená na faktorizaci závislostních stromů. Ve většině těchto přístupů analyzátor (parser) ztrácí značnou část kontextové informace během procesu faktorizace. V této práci zkoumáme, jak syntakticko-sémantické rysy ovlivňují metody diskriminativního strojového učení vyššího řádu pro závislostní syntaktickou analýzu. Ukazujeme, že lingvistické rysy v mnoha případech přinášejí významné zlepšení úspěšnosti.

Nejdříve přinášíme přehled několika diskriminativních metod učení pro grafové statistické závislostní parsery a vysvětlujeme koncept vyššího řádu, což je zobecnění práce (Koo a Collins 2010) a (McDonald et al. 2006). To nás dovede k jádru práce – rysovému inženýrství pro závislostní parsery vyššího řádu. Experimentujeme s několika syntakticko-sémantickými rysy a snažíme se vysvětlit jejich teoretické základy. Pokusy provádíme na dvou odlišných jazycích – angličtině a češtině – a srovnáváme několik výsledků získaných různými algoritmy.

**Klíčová slova:** závislostní syntaktická analýza, parsing, diskriminativní učení, vyšší řád, sémantické rysy.