

Neřízená závislostní analýza je alternativní způsob určování vztahů mezi slovy ve větě. Nepotřebuje žádný anotovaný závislostní korpus, je nezávislý na jazykové teorii a univerzální pro velké množství jazyků. Jeho nevýhodou je ale zatím relativně nízká úspěšnost.

V této práci diskutujeme některé předchozí práce a představujeme novou metodu neřízené analýzy. Náš závislostní model se skládá ze čtyř podmodelů: (i) hranový model, který řídí rozdělení dvojic řídicích a závislých členů, (ii) model plodnosti, který řídí počet členů závislých na uzlu, (iii) model vzdálenosti, který řídí délku závislostních hran a (iv) model vypustitelnosti. Tento model je založen na předpokladu, že slova která se mohou z věty vypustit, aniž by se porušila její gramatičnost jsou v závislostním slově listy.

Odvození závislostních struktur provádíme pomocí Gibbsova vzorkovače. Představujeme vzorkovací algoritmus, který zachovává projektivitu závislostních stromů, což je velmi užitečnou vlastností.

V našich experimentech na 30 jazycích srovnáváme výsledky pro různé parametry modelů. Naše metoda překonává dříve publikované výsledky pro většinu zkoumaných jazyků.