Klíč k rychlému přizpůsobení jazykových technologií pro libovolný jazyk závisí na dostupnosti základních nástrojů a datových zdrojů, jako jsou
jednojazyčné nebo paralelní korpusy, anotované korpusy, značkovače slovních druhů, syntaktické analyzátory, a podobně. Jazyky, pro něž tyto základní zdroje neexistují, označujeme jako zdrojově chudé jazyky.
V této práci se zabýváme otázkou závislostního syntaktického rozboru zdrojově chudých jazyků za pomoci zdrojů pro jiné jazyky. Pro nalezení závislostní struktury používáme tři postupy: (i) promítnutí závislostí ze zdrojově bohatého jazyka do zdrojově chudého jazyka za pomoci slovního zarovnání v paralelním korpusu (ii) analýze pod-zdroji jazyků pomocí parserů, jejichž modely jsou vyškoleni na stromových korpusů z jiných jazyků, a nedívejte se na skutečných slovních forem, ale pouze na POS kategorie. Zde se zabýváme problémem neslučitelnosti různých anotačních stylů používaných zdrojovými analyzátory a cílovými závislostně anotovanými korpusy používanými pro evaluaci, který řešíme pomocí harmonizace anotací do jednotného standardu; a konečně (iii) zavádíme nový postup, ve kterém pro promítnutí závislostí do zdrojově chudého jazyka používáme paralelní korpusy vytvořené pomocí strojového překladu namísto lidského překladu.
Výše uvedené postupy jsme použili na pět indických jazyků: hindštinu, urdštinu, telugštinu, bengálštinu a tamilštinu (seřazeno sestupně podle dostupnosti závislostně anotovaných dat). Abychom prokázali použitelnost uvedených postupů v praxi, vyvinuli jsme závislostně anotovaný korpus pro tamilštinu, pro niž dosud žádný takový zdroj neexistoval, a takto získaná data využíváme pro evaluaci a také jako zdroj pro závislostní rozbor jiných indických jazyků. Nakonec jsme seznam se strategie, které může být použit k získání závislost struktury pro cílových jazyků pod jiný scénáře s omezenými zdroji.