

Spolu se vznikem stále dalších jazykových zdrojů – slovníků, lexikálních databází, korpusů, treebanků – roste i potřeba jejich účinného propojování, které by umožnilo snadné využití veškerých shromážděných vlastností a informací. V tomto ohledu je také aktuální téma univerzálních lexikografických formátů.

Tato práce zkoumá metody automatického propojování jazykových dat. Představíme zde systém na propojování slovníků, jakými jsou například VALLEX, PDT-Vallex, FrameNet, nebo SemLex, které poskytují syntaktickou informaci o svých heslech. Systém je automatický, umožňuje tudíž opakovanou aplikaci na novější verze vyvíjejících se jazykových zdrojů. Na základě syntaktické informace obsažené ve slovníku víceslovných výrazů SemLex navrhujeme metodu vyhledávající tyto výrazy v automaticky anotovaném textu.

Praktickým výstupem potvrzujícím úspěšnost použitých metod je mj. propojení slovníků VALLEX a PDT-Vallex vedoucí k doplnění desítek tisíc anotovaných vět z treebanků PDT a PCEDT do VALLEXu.