

## Základy korpusové lingvistiky

### Úkoly na 3. dubna 2017

1. Zjistěte v korpusu SYN2015, co je nejčastěji považováno za „těžké“ a „lehké“, tj. která substantiva se pojí s adjektivy *těžký* a *lehký*. Do frekvenčního seznamu zahrňte dvojice s frekvencí 20 a vyšší.

2. Vyhledejte v korpusu SYN2015 přechodníky přítomné. Využijte k tomu hledání pomocí morfologické charakteristika „tag“ – slovní druh + detailní určení slovního druhu.

Poté:

- Vytvořte frekvenční seznam lemmat, tj. zjistěte, od kterých sloves se využívají přechodníky nejčastěji – které mají vyšší frekvenci než 50.
- Vytvořte frekvenční seznam slovních tvarů, tj. zjistěte, které přechodníky a v kterých tvarech jsou v češtině nejvíce frekventované (mají vyšší frekvenci než 50).
- Projděte si prvních 100 slovních tvarů podle frekvence a pokuste se vyvodit, který tvar přechodníku podle rodu a čísla se užívá nejčastěji.