

Budete potřebovat:

- na adrese <https://jupyter.korpus.cz/lemtag> nástroj LemTag
- na adrese <https://jupyter.korpus.cz/lemtag/md/help> je k němu návod včetně obecných pokynů k úkolu
- na [korpusovém cloudu](#) je ke stažení 22 sad po 100 nerozpoznaných tvarech, vybraných frekvenčně (ale i tak jsme se dostali už na úroveň hapaxů) a následně seřazených retrográdně
- soubor Excel (MS Office) nebo Calc (LiberOffice) – zapisování (kopírování) výsledků
- soubor Word (příp. jiný textový editor) – zapisování překlepů, chyb a nejasností
- v KonTextu korpus oral_v5 – vyhledání tvaru slova
- v KonTextu korpus syn_v4 – hledání spisovného lemmatu, hledání analogického tvaru
- v KonTextu korpus oral2013 – zvukové nahrávky, ověření výslovnosti slova
- Internetová jazyková příručka – ověření pravopisu, ověření slov ve slovnících SSČ, SSJČ (příp. DEBDict)
- Google – např. hledání názvů obcí (<http://mesta.obce.cz/>)
- tagset pozičního systému – ověření morfologické značky
https://wiki.korpus.cz/doku.php/seznamy:tagy#popis_jednotlivych_pozic_znacky

Co najdete na frekvenčním dně:

- 1) překlepy (tvar je obtížně vyslovitelný) – *věřejně, pravděpoboně, valsně, odečtě, budťě, skříňě, tedkoň*, příp. ověřit poslechem nahrávky, pokud je k dispozici
- 2) pravopisné a jiné chyby – *spontálně, zlíně*, chybné názvy obcí
- 3) nejasnosti – *angličinář, irinuješ, naexpuješ, nelábeš*
- 4) tvary bez spisovného lemmatu – vytvořit lemma nové – *kočena, dravkyně, čvachtaliště*
- 5) tvary s hláskovou změnou či koncovkou (hovorovou, nářeční) – spisovné lemma – *mléně* → *mlýn, dondeš* → *dojít, stróhneš* → *strouhnout*
- 6) složená slova – *spidervepř* → *vepř, celorodinně* → *rodinně*