

SEMINÁRNÍ ÚKOL – JARO 2017

Termín odevzdání: do **31. 5. 2017**, do Odevzdáárny v ISu u předmětu CJBB85

Materiál:

- korpus oral_v5 – spojení mluvených korpusů ORAL2006, ORAL2008 a ORAL2013
- spontánní dialogy, mluvčí z celého území ČR
- korpus zkušebně lemmatizován a morfologicky označován
- **slova s frekvencí – 1, u kterých se předpokládá, že lemmatizace a značkování nejsou možné**
- **přístup – <http://kontext.korpus.cz>, vlastní přihlášení**

Úkol:

Roztřídit slova do stanovených kategorií (s tímto číslováním), případně navrhnout kategorii novou.

1. překlipy (výskyty v ORAL2013 ověřit poslechem – jde o chybu v přepisu, nebo slovo bylo takto vysloveno?)
2. pravopisné a gramatické chyby
3. vlastní jména a názvy
4. zkratky a zkratková slova
5. cizojazyčné výrazy
6. citoslovce
7. záhady (určit slovní druh (příp. další morf. kategorie), odhadnout význam)
8. výslovnost
9. hovorové výrazy

Postup:

- ověřit tvar slova v korpusu, prohlédnout jeho kontext
- zařadit do kategorie, okomentovat
- pokud nelze slovo v korpusu nalézt, ověřte si jeho původní tvar v originálním souboru na <https://owncloud.korpus.cz/index.php/s/RwUHKZWkiMzcJfu>
- Internetová jazyková příručka (<http://prirucka.ujc.cas.cz>) – ověření pravopisu, ověření slov ve slovnících SSČ, SSJČ (příp. DEBDict)
- Google – např. hledání názvů obcí (<http://mesta.obce.cz/>)
- volitelně Český jazykový atlas (<http://cja.ujc.cas.cz/>)