

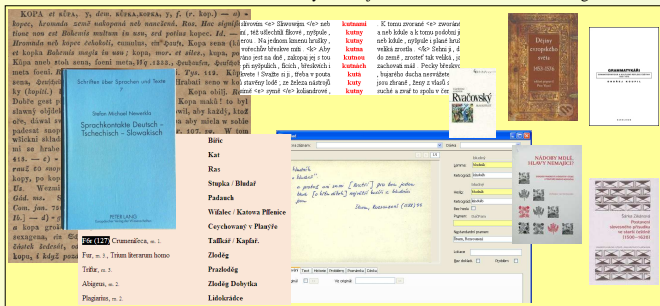
# Co říkají slovníkové a korpusové zdroje o humanistické češtině

Mgr. František Martínek, Filozofická fakulta Univerzity Karlovy, Ústav českého jazyka a teorie komunikace; Ústav pro jazyk český Akademie věd České republiky, v. v. i., oddělení vývoje jazyka, Praha; [fmart@seznam.cz](mailto:fmart@seznam.cz)

Tento poster je jedním z výsledků doktorandského grantového projektu č. 16 809 *Lingvistická analýza českých humanistických textů* uděleného Grantovou agenturou Univerzity Karlovy.

## Co se dozvídáme o humanistické češtině...

- ze slovníkových zdrojů?
- z elektronických zdrojů?
- z edic?
- z monografií?



## ... a co bychom se ještě potřebovali dozvědět?

Většina poznatků o humanistické češtině byla vyvozena z nespolehlivých zdrojů (čest výjimkám – Porádkovi, Zikánově, Michálkovi, Šlosarovi aj.). Mezi nejdůležitější limity poznatků, příčiny a důsledky těchto limitů patří:

- zaměření diachronní jazykovědy na zkoumání staré češtiny
- neexistence slovníku střední češtiny, celkového gramatického popisu atd.
- velmi špatná integrace zdrojů, popř. jejich faktická nepřítupnost
- malý rozsah a nereprezentativnost elektronických korpusů historické češtiny
- nejednotný ediční přístup k jazykovým památkám

## Potřebujeme popis všech plánů jazyka! Ale kde začít?

Chybí spolehlivý, reprezentativní a dostatečně rozsáhlý zdroj jazykových dat. To nutí každého badatele investovat čas do vytváření vlastního korpusu. Pokusili jsme se aspoň částečně přispět k řešení tohoto problému.

## Tvorba textového korpusu humanistické češtiny



Tisky z humanistického období nebo jejich části se přepisují podle konvencí diachronní složky Českého národního korpusu. V textu se užívají značky označující mj. tituly a nadpisy, foliáci či paginací tisku, cizojazyčná slova užitá citátově a transliteraci sporných míst.

**Příklad zeditovaného textu:**  
<<1a</s> <k> <o>Cyri paedia</o>, hadno-věrná starožitná historia o chvalitebném ve velikých <e>wesselských</e> knížecích <e>Knižecích</e> ctnostech vycvičení a zvedení, o slavných skutcích, vítězých válekách a právě heroickém <l>heroický</l> šlechtěném životu Cyra Staršího, prvního monarchy perského. </k>

- Prameny zahrnuté do korpusu (výběr):**
- anonym, Smlouva o ručnicích, 1524
  - Bartoš pisař, Kronika pražská od r. 1524 do r. 1530
  - Václav Miřinský, Písne gruntovní a velmi utěšené, 1531
  - Brikcí z Licska, Práva městská, 1536
  - Václav Hájek z Libočan, Kronika česká, 1541
  - Isidor Sevilský, Knižka velmi utěšená, 1551
  - Mikuláš Šúd ze Semanina, Novina a přihoda ..., 1551
  - Sebastian Münster – Zikmund z Püchova, Kozmografie česká, 1554
  - Tomáš Bavorovský, Zrcadlo onoho věčného a blahoslaveného života, 1561
  - Václav Zelotýn, Kalendář hvězdářský na r. 1575, 1574
  - Goerl z Goerlštejna, Aritmetika ..., 1577
  - Daniel Adam z Veleslavína, Kalendář historický, 1578
  - anonym, Agenda česká, 1581
  - Jan z Bakova, Dialog sedláka s doktorem o moru, 1582
  - anonym, O zachování dobrého zdraví knižka školy salernitánské, 1588
  - Matouš Hosius z Vysokého Mýta, Kronica moskevská, 1590
  - Heinrich Buenting, Itinerarium sacrae scripturae, 1592
  - Václav Březan, Život Viléma z Rožmberka, por. 1592
  - Bartoloměj Paprocký, Kvalt na pohany, 1595
  - Jan Kobiš z Bitýšky, O měřích vinných sudův, 1596
  - Giovanni Boccaccio, Velmi pěkná nová kronika, 1600
  - Arnošt Ferdinand Hess, Bič Židovský, 1603
  - Xenofon, Cyrípaidea, hodnověrná starožitná historie ..., 1605
  - Theodor Maius, Krátká zpráva o kometě, 1607
  - Kryštof Harant, Putování aneb cesta z království českého ..., 1608
  - Jiří Tesák Mošovský, Collis Vinearius, 1611
  - anonym, Div veliký a pravdivý při jedné sprostné děvečce sedlské, 1612
  - Viktorín A. Mantinský, Monumentum statečného a urozeného rytíře, 1612
  - Havel Žalanský (Phaeton), O ctých manželkách těhotných ..., 1615
  - Augustin Mitis, Kázání o povětří, 1619
  - Fridrich Falcký, Otevřené rozepsání, 1619
  - Jan Jesenius, K jich milostem ..., 1619
  - Kašpar Kargesius, Summa hry o svatém Eliášovi proroku, 1620

## K čemu byl, je – a bude? – korpus dobrý?

Byl vytvořen vyvážený korpus českých humanistických textů (vč. úryvků rozsáhlejších děl) o rozsahu 600 000 slovních tvarů, který bude díky ochotě příslušných lidí moci být zveřejněn v diachronní složce Českého národního korpusu. Texty jsou transkribovány. Jakékoliv nejasnosti či potenciální zdroje budoucích pochyb jsou zároveň zaznamenány v transliteraci. Pravopisné jevy je tedy na textech možné zkoumat jen velmi omezeně, zato je korpus užitečný zejména pro analýzu jevů hláskoslovných, syntaktických, výběrově pak morfologických a lexikálních.

Při tvorbě korpusu byla zpsněna transkripční pravidla. Postup transkripce byl předložen k diskusi na brněnské konferenci *Dějiny českého pravopisu (do r. 1902)*.

Řešitel grantu vede na FF UK volitelný seminář *Lingvistická analýza českých humanistických textů*. Výsledky analýz dílčích jazykových jevů – např. deklinace zájmen *on*; lexikální význam polysémních sloves; stavba, rozšíření a stylovorní role analytických verbonominálních spojení, slovosled valenčních doplnění; odkazování v textu – jsou průběžně přednášeny na studentských i vědeckých konferencích a publikovány v příslušných sbornících.

## Co můžeme vykázat?

- MARTÍNEK, František: Poznámky k jazyku a stylu Hájkovy *Kroniky české*. In: *Pokoušení Jaroslava Kolára. Sborník k osmdesátinám*. Praha: ÚČL AV ČR – FF UK, 2009, s. 115–125.
- MARTÍNEK, František – OCELÁK, Radek: „Tvorba korpusu k lingvistické analýze humanistické češtiny“, in: *Varia XIX. Zborník plných příspěvků z XIX. kolokvia mladých jazykovědců (Trnava – Modra-Harmónia 18.–20. 11. 2009)*, Trnava: PeDF TU – SJS pri SAV, 2010, s. 219–227. (elektronická publikace)
- MARTÍNEK, František: „Analytická verbonominální spojení – důkaz ovlivnění starší češtiny latinou a němčinou?“, in: J. Valková (ed.), *Kulturní translace – Kulturelle Translationen – Translacje kulturove: příspěvky z mezinárodní studentské konference interFaces VII v červnu 2009 v Praze*, Praha: Akropolis a FF UK, 2010, s. 37–47.
- MARTÍNEK, František: „K valenci u analytických verbonominálních spojení (s příklady ze starší i nové češtiny)“, příspěvek na konferenci 11. mezinárodní setkání mladých lingvistů (Olomouc 10.–12. 5. 2010), v tisku.
- MARTÍNEK, František: „Pravopis starších textů versus textový korpus k lingvistickým analýzám“, příspěvek na konferenci *Dějiny českého pravopisu*, Brno 23.–25. 9. 2010, v tisku.
- MARTÍNEK, František – RYSOVÁ, Kateřina: „A Corpus of Older Czech Texts for Linguistic Purposes“, příspěvek na konferenci SLAVICORP, Varšava 22.–23. 11. 2010, v tisku.

## Jaká bude budoucnost?

(To závisí především na přízni grantových agentur...)

- Dva hlavní úkoly:
- vyhodnocovat dostupný jazykový materiál
  - rozšiřovat, integrovat, propojovat ... zdroje
- V plánu jsou detailní analýzy jazykových jevů, pro něž poskytují korpus dostatečný materiál, tedy např. sledování postupu hláskových změn probíhajících v 16. století (resp. výsledovatelých z tehdejších písemných záznamů jazyka) a vytvoření frekvenčního slovníku tvarů slov, na jehož základě budou vyhodnocena slova jinde nedoložená. Obojí ještě v tomto roce!

**Korpus bude volně přístupný a uvítáme, pokud se stane materiálem i pro Vaše analýzy!**

## Řešitelský tým

František Martínek (řešitel), Karel Kučera (školitel), Jiří Milička, Radek Ocelák, Jiří Pergler, Kateřina Rysová, Kateřina Smyčková, Andrea Svobodová, Eva Záhořová, Anna Zitová (studenti magisterského a doktorského studia), Robert Dittmann, Marek Janosik-Bielski, Pavel Kosek, Michal Křen, Petr Nejedlý, Martin Sluka, Šárka Zikánová (odborní konzultanti).