

STUDIE Z KORPUSOVÉ LINGVISTIKY

Redakční rada:

doc. RNDr. Vladimír Petkevič, CSc. (předseda)

doc. Mgr. Václav Cvrček, Ph.D.

prof. PhDr. František Čermák, DrSc.

Mgr. Anna Čermáková, Ph.D.

Mgr. Jan Kocek

PhDr. Marie Kopřivová, Ph.D.

Mgr. Michal Křen, Ph.D.

prof. PhDr. Karel Kučera, CSc.

Ing. Alexandr Rosen, Ph.D.

Mgr. Pavel Vondříčka, Ph.D.



Velmi moudré pojednání o jazyce

Šeherezáda Vágenknechtová

Tato kniha nikdy nevznikla. (Bohudíky!)

Recenzenti:

PhDr. První Chytrák, CSc.

Doc. Mgr. Druhej Chytrák, Ph.D.

Prof. Dr. Největší Chytrák, CSc.

Cover © Michaela Blažejová, 2019

© Filozofická fakulta Univerzity Karlovy, 2019

Všechna práva vyhrazena.

ISBN 978-80-XXXX-XXX-Y

Poděkování

Především bych chtěl poděkovat Honzovi Kockovi, že se o tu naši řadu tak pěkně stará. Ne každá ediční řada má takového Honzu, který by dbal na všechny grafické až typografické náležitosti sazby, jaké má každá správná knížka splňovat. A nezapomínejme taky na to, že hlavně je to přímá čupr nóbl človíček a supr férovej chlap se srdíčkem na správnym místě. Honzo, seš jednička!

OBSAH

1. Sazba pro Lidovky	7
1.1 Instalace a použití	7
1.2 Základní sazba	7
1.2.1 Patkový font – Charis SIL	8
1.2.2 Bezpatkový font – Myriad Pro	8
1.2.3 Strojopisný font – Cousine	8
1.3 Uvozovky a použití dalších jazyků v textu	8
1.4 CQL příkazy	9
1.5 Příklady a konkordance	9
1.5.1 Konkordanční seznamy	9
1.5.2 Příklady a doklady – petit	11
1.6 Tabulky	11
1.7 Obrázky	12
1.8 Matematické vzorce	12
2. Formátování dokumentu	13
Summary	14
Literatura	15

1. Sazba pro Lidovky

Tento dokument slouží jako šablona pro sazbu knih v ediční řadě *Studie z korpusové lingvistiky* pomocí \LaTeX . Jedná se o šablonu, která je platná od 20. svazku řady (etalonem pro nejasnosti budiž kniha T. Jelínka: *Skladební funkce a pád v korpusu*).

Pokud je už rukopis převeden do \TeX , nemělo by být potřeba do sazby nějak významněji zasahovat nad rámec stránkového a řádkového zlomu. Výjimku tvoří pouze tyto specifické oblasti:

1. tabulky – pro sazbu tabulek se využívá balík *booktabs* a aby byl dodržen grafický předpis z NLN je třeba běžné tabulky ručně upravit (viz níže 1.6)
2. konkordance – pro sazbu konkordancí je v této šabloně předdefinované specifické prostředí, do něhož je třeba konkordanční seznamy převést (viz níže 1.5.1)
3. příklady – grafický předpis předpokládá, že rozsáhlejší příklady, soupisy apod. budou sázeny speciálním způsobem, pro tyto účely je v této šabloně prostředí pro sazbu petitu (viz níže 1.5.2)
4. bibliografie – viz sekci Literatura

1.1 Instalace a použití

Soubor `cnkbook.cls` musí být ve stejném adresáři jako kompilovaný dokument, a stejně tak i podadresáře `supl` a `img`. Fonty z adresáře `fonty` je třeba nainstalovat do systému.

V dokumentu se pak použije `\documentclass{cnkbook}`, který naimportuje třídu `book` a řadu dalších potřebných a užitečných balíčků. Volba `textonly` zamezí sázení úvodních a závěrečných stran (titulní strana, tiráže, poděkování apod.) a vygeneruje navíc k textu už jen obsah. Jakékoliv další volby budou předány dál třídě `book`.

Pro sazbu kompletní knihy je také třeba nadefinovat zbylé parametry jako ve zdrojovém kódu tohoto příkladu (titul, jméno autora, rok vydání, ISBN, cenu atd.). Stránka s poděkováním se bude sázet jen tehdy, pokud je nějaké poděkování nadefinované a není prázdné. Blíže viz 2.

Dokument se musí zkompileovat programem \XeLaTeX , nikoliv obyčejným \LaTeX em.

1.2 Základní sazba

Základní font je Charis SIL 9/11, nadpisy apod. jsou pomocí Myriad Pro, strojopisný font (zatím Cousine) se užívá jenom zřídka pro zachycení kusů kódu nebo CQL dotazů. Pokud není uvedeno jinak, měly by fungovat všechny standardní příkazy \LaTeX u.

- výčtové seznamy
- poznámky pod čarou

- nadpisy (až do úrovně paragraph)

■ 1.2.1 Patkový font – Charis SIL

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789.¹

■ 1.2.2 Bezpatkový font – Myriad Pro

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

■ 1.2.3 Strojopisný font – Cousine

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789

1.3 Uvozovky a použití dalších jazyků v textu

Uvozovky lze sázet pomocí příkazu `\enquote{...}`, který lze i vnořit, aby na druhé úrovni automaticky vytvořil uvození jednoduchými uvozovkami. Tento příkaz reaguje na aktuální nastavení jazyka a podle něj dosadí odpovídající uvozovky. Např. „Toto je text v ‚uvozovkách‘ cha!“. Ne všechny jazyky ovšem mají svůj styl uvozovek předdefinovaný (viz islandštinu v Summary) – ostatně, češtinu jsme si dodefinovali taky sami. U jiných jazyků lze naopak zase vybírat z více typů uvozovek. Vid' dokumentaci: <http://texdoc.net/texmf-dist/doc/latex/csquotes/csquotes.pdf>

Naše třída zavádí balík `polyglossia` (nástupce `babelu`) a automaticky nastaví češtinu jako výchozí jazyk. Aktuální nastavení jazyka lze měnit v textu odpovídajícím způsobem: buď přechodně pomocí příkazu `\text<jazyk>{...}`, tedy např. `\textenglish{F*ck!}`, nebo pro delší úseky textu jejich uzavřením do prostředí `\begin{jazyk} ... \end{jazyk}`. Nejprve je však potřeba

¹ V poznámkách pod čarou se používá menší velikost písma 7/8. příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789 *příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789* **příliš žluťoučký kůň úpěl ďábelské ódy.,!? PŘÍLIŠ ŽLUŤOUČKÝ KŮŇ ÚPĚL ĎÁBELSKÉ ÓDY 0123456789.**

všechny případné dodatečné jazyky nahrát v hlavičce dokumentu pomocí příkazu `\setotherlanguage{jazyk}` nebo (pro více jazyků) pomocí příkazu `\setotherlanguages{jazyk1,jazyk2,...}`. Pro podrobnosti pohlédniž na originální dokumentaci: <http://mirrors.ibiblio.org/CTAN/macros/latex/contrib/polyglossia/polyglossia.pdf>.

Makro `\uv{...}` pro sazbu českých uvozovek je předdefinováno také (pro „zpětnou“ kompatibilitu, resp. lenost autorů). To samozřejmě neumožňuje zanořování uvozovek, resp. na zanořování nijak nereaguje.

1.4 CQL příkazy

Korpusové dotazy, příp. i kusy programového kódu, se můžou sázet pomocí příkazu `\cql`, např. `[word="vo.*el"]` a `[tag="Vp---1--AA.*"]`. Kromě změny fontu příkaz `\cql` vypíná i \TeX ovské mapování a ligatury (dvojitě pomlčky, uvozovky apod.), takže není třeba dávat pozor na psaní několika pomlček za sebou v CQL dotazech.

1.5 Příklady a konkordance

1.5.1 Konkordanční seznamy

Pro konkordanční seznamy je v šabloně připraveno speciální prostředí s názvem `conclist`. Je k dispozici ve dvou variantách – defaultní varianta tiskne konkordanci ve tvaru *levý kontext* – KWIC – *pravý kontext*, volba s přepínačem `[ref]` pak předpokládá čtyři údaje v každém konkordančním řádku: *referenci* – *levý kontext* – KWIC – *pravý kontext*. Každý konkordanční řádek je navíc uvozen příkazem `\concline`, v němž se specifikují jednotlivé části konkordance.

```
\begin{conclist}[ref]
\concline{levý kontext}{KWIC}{pravý kontext} \% bez ref
\concline[reference]{levý k.}{KWIC}{pravý k.} \% s ref
...
\end{conclist}
```

Konkordanční seznam bez referenčních údajů po levé straně se vysází následujícím způsobem:

jezera se pod *mými* kroky **rozezněla** tuberkulózním kašlem a čím více
a čím více jsem dupotem **zdůrazňoval** svoje *vítězství* , tím více

Konkordanční seznam s referenčními údaji po levé straně vypadá následovně:

abc	jezera se pod mými kroky rozezněla tuberkulózním kašlem a čím více
xyz	a čím více jsem dupotem zdůrazňoval svoje vítězství , tím více
uvw	Síla , která mě tak neodolatelně hnala vstříc nebeským jatkám , se už
Abcd	vítězství , tím více krve se zachycovalo na kapesíkovitém dnu .

Pro rozsáhlejší konkordance je třeba využít prostředí *sidewaystable* a *sidewaysfigure* z balíku *rotating*.

■ 1.5.2 Příklady a doklady – petit

Pro poznámky drobnějším písmem a hlavně pro příklady a doklady je určeno prostředí `petit`.

Písmo petit: Dokládá doklady, příkládá příklady a poznamenává poznámky.

Podle grafického předpisu z NLN má každý řádek v `petitu` začínat bez odsazení a celý text má být v kurzivě (obojí je už přednastaveno v definici prostředí). Blok je pak odsazen z obou stran prázdným řádkem. Vzhledem k tomu, že vertikální mezera mezi řádky (vzdálenost účař?) je stejná jako u normálního písma, působí `petit` trošku vzdušnějším dojmem.

1.6 Tabulky

Pro tabulky se používá balík `booktabs`, příp. spolu s balíkem `longtable`. Sázejí se do plovoucího prostředí `table` a měly by mít popisek (caption).

```
\begin{table}[ht]
\begin{tabular}{@{}llr@{}}
\toprule
\multicolumn{2}{c}{Item} \\
\cmidrule(r){1-2}
Animal & Description & Price (\$) \\
\midrule
Gnat & per gram & 13.65 \\
& & each & 0.01 \\
Gnu & stuffed & 92.50 \\
Emu & stuffed & 33.33 \\
\addlinespace[3pt]
Armadillo & frozen & 8.99 \\
\bottomrule
\end{tabular}
\caption{Toto je popisek tabulky...}
\label{abc}
\end{table}
```

Výše uvedený kód vysází následující tabulku 1.

Item		
Animal	Description	Price (\$)
Gnat	per gram	13.65
	each	0.01
Gnu	stuffed	92.50
Emu	stuffed	33.33
Armadillo	frozen	8.99

Tab. 1. Toto je popisek tabulky. Tento popisek je poněkud delší, takže nám dá-li Bůh přeteče na další řádek

U tabulek **nepoužíváme svislé čáry**! Při definici sloupců (v prvním řádku bloku `\tabular`) je dobré před první a za poslední sloupec vložit `@{}`, čímž se eliminuje vertikální odsazení tabulky zleva i zprava. Horizontální čáry v tabulce jsou rozdělené podle funkce na úvodní (`\toprule`), prostřední (`\midrule`) a závěrečnou (`\bottomrule`). U každé je možné definovat šířku zvlášť. Příkazy `\hline` či `\cline` bychom v zájmu snadných úprav neměli používat.

1.7 Obrázky

Obrázky by měly být v tiskové kvalitě (ideálně vektorové; podrobnosti doplní Honza). Sázejí se do plovoucího prostředí `figure` a měly by mít popisek (`caption`).



Obr. 1. Toto je popisek obrázku

Obrázky se sází na střed (viz obr. 1).

1.8 Matematické vzorce

Příklad textu se vzorci. Otázkou je, zda je matematický font vyhovující. Zjevně to totiž není Charis SIL.

Analýzou se zjišťuje, jak k celkovému rozptylu ve všech sloučených datech (s_{celk}^2) přispěl rozptyl mezi jednotlivými skupinami (s_{mezi}^2) a jaký díl odráží rozptyl uvnitř jednotlivých skupin (s_{vnit}^2): $s_{celk}^2 = s_{vnit}^2 + s_{mezi}^2$. Poměr rozptylů má přitom F rozdělení: $F = s_{mezi}^2 / s_{vnit}^2$.

Výpočet vnitroskupinového a meziskupinového rozptylu přitom není úplně přímočarý.

$$s_{vnit}^2 = \frac{\sum_{j=1}^s \sum_{i=1}^{n_j} (\bar{x}_j - x_{ji})^2}{n - s}$$

kde s je počet skupin, n_j je počet prvků v dané skupině.

$$s_{mezi}^2 = \frac{\sum_{j=1}^s n_j (\bar{\bar{x}} - \bar{x}_j)^2}{s - 1}$$

kde $\bar{\bar{x}}$ je celkový průměr, který je poměřován se skupinovými průměry (\bar{x}_j) tolikrát, kolik je prvků v jednotlivých skupinách (n_j).

2. Formátování dokumentu

Uživatel šablony musí vložit pouze svůj text a to nejlépe pomocí příkazu `\include`. Není-li použita volba `textonly` (viz kap. 1.1), automaticky se k dokumentu do-generuje (nikoli degeneruje!) stránka s redakční radou a titulní stránka, následuje strana s dedikací, recenzenty a ISBN, a poté poděkování a obsah. Na závěr dokumentu se vloží sekce „V NLN už vyšlo“ a tiráž.

Tyto části se generují podle údajů, které je nutné vyplnit v hlavičce (viz zdrojový kód tohoto příkladu). Jedná se o příkazy:

```
\title \author \date \isbn \redakce \sazba \dedikace  
\recenzenti \podekovani
```

Automaticky by se měl vygenerovat i zmíněný obsah, a to s formátováním, které bylo zvoleno jako optimální (nesouhlasí zcela s předchozími svazky, ale zdálo se nám s Honzou nejlepší).

Summary

I pravil \TeX : „Uvozovky budou ,obyčejně‘ použity české, neboť je to výchozí jazyk. Pokud se však jazyk změní, uvozovky se přizpůsobí, jak vidno níže.“

“Something in ‘English’...” You should definitely use `\begin{english}` ...`\end{english}` environment in summary.

Something in “English”...You may also use `\textenglish`, but `\verb` will not work inside.

...og så litt på norsk, også. La oss prøve «anførelsestegn og enda «andre» i tillegg».

?Mun það einnig fara ?vel? á íslensku, ef til vill?? Nei! Því miður er íslenska ekki afmörkuð í csquotes!

„Auf Deutsch geht es ,hoffentlich‘ auch.“

« Mais je ne parle pas “français”, moi. »

«Yo no hablo “español”.»

Literatura

Literatura se sází ve speciálním výčtovém prostředí *biblio*. Každá položka v bibliografii začíná příkazem `\item`. Ostatní požadavky na formátování bibliografie (kurziva apod.) si musí autor ohlídat sám.

- Adolphs, S. (2008): *Corpus and Context. Investigating Pragmatic Functions in Spoken Discourse*. John Benjamins. Amsterdam/Philadelphia.
- Anděl, J. (2005): *Základy matematické statistiky*. MatfyzPress. Praha.
- Baayen, H. R. (2001): *Word Frequency Distributions*. Kluwer Academic Publishers. Dordrecht/Boston/London.
- Baayen H. R. (2008): *Analyzing Linguistic Data*. Cambridge University Press. Cambridge.
- Baker, P. (2010): *Sociolinguistics and Corpus Linguistics*. Edinburgh UP. Edinburgh.
- Bartoň, T. – Cvrček, V. – Čermák, F. – Jelínek, T. – Petkevič, V. (2009): *Statistiky češtiny*. Nakladatelství Lidové noviny. Praha.
- Biber, D. – Conrad, S. – Cortes, V. (2004): If you look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25(3), (s. 371–405).

Velmi moudré pojednání o jazyce

Šeherezáda Vágenknechtová

Vydalo NLN, s. r. o.,
Nakladatelství Lidové noviny,
Dykova 15, 101 00 Praha 10,
v roce 2019
Vydání první

Jazyková a odborná redakce Cenzor Cenzorovič
Obálku navrhla Michaela Blažejová
Knihu graficky upravila Marie Tvrdá
Technická redakce Josef Kupka
Sazba v systému \TeX písmem Charis SIL Kuňha Kuňhatá
Tisk Art D – Grafický ateliér Černý s. r. o., Praha

Doporučená cena včetně DPH 129 Kč

ISBN 978-80-XXXX-XXX-Y