

Jak se adaptují slova v češtině a jak lze používat korpus a další nástroje pro popis slovo tvorby

Naším cílem je na praktickém příkladu ukázat postup práce s korpusem a diskutovat problémy. Mějme následující úkol: chceme zjistit, jak se v češtině (reprezentované jaz. korpusy) uplatňuje přejatá slovní zásoba, přičemž jde o slova, která chápeme jako popisná a jsme schopni na základě principu analogie a opakování (Slavičková, 1967) nalézt hranici mezi morfy, přestože slovo bylo přejato jako celek (nevzniklo v rámci češtiny). Zajímá nás rovněž, zda takto vydělitelné (cizorodé) stavební prvky mají nějaký vliv na češtinu, a sice, zda a do jaké míry vedou ke vzniku hybridních derivátů.

Mějme substantiva jako *rafinérie, pizzerie, drogérie, oranžérie, ...*

1. Ačkoliv nejde o slova utvořená v rámci češtiny, jsme schopni vydělit stavební prvek, a to na základě principu (1) opakování (a) stavebního prvku [e]rie a (b) základu vzniklého po jeho odtržení (*rafin-ovaný, pizz-a, drog-ista*), a (2) analogie (*oranž-ový* ale *pomeranč*).
2. Základy jsou evidentně základy přejatých slov. Je tomu tak vždy, nebo je možné příslušný stavební prvek kombinovat s českými základy?
3. Kolik takových jednotek se v češtině (reprezentované korpusy) vyskytuje a jaké mají tato slova významy (jde jenom o názvy míst, nebo lze vysledovat i další významy).
4. Jak postupujeme ve sporných případech (např. nevíme-li, zda je slovo opravdu utvářené tak, jak lze předpokládat na základě jeho formy)?
5. Všimáme si, že stavební prvek má dvě varianty, a to s krátkým *e* a dlouhým *é*. Jakými pravidly se užití *e/é* řídí?

V prvním kroku použijeme nástroj *Morfio* a budeme v korpusu SYN2010 hledat dvojice slov se společným řetězcem libovolným, a řetězci *erie/érie* odlišnými (obrázek 1).

Získáme 33 dvojic (obr. 2), které budeme dále pozorovat.

Obrázek 1

Odkaz na toto zadání: <http://morfio.korpus.cz/qwef4Mcr>

Obrázek 2

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	archebakterie (2)	archebaktérie (1)
2	bakterie (2451)	baktérie (61)
3	bižuterie (220)	bižutérie (10)
4	cafeterie (5)	cafetérie (1)
5	draperie (48)	drapérie (41)
6	drogerie (93)	drogérie (238)
7	dyzenterie (9)	dyzentérie (3)
8	euroserie (1)	eurosérie (1)
9	galanterie (59)	galantérie (2)
10	galerie (6574)	galérie (44)
11	haterie (2)	hatérie (3)
12	cherie (19)	chérie (21)
13	infanterie (17)	infantérie (1)
14	kafeterie (15)	kafetérie (9)
15	kafilerie (8)	kafilérie (22)
16	koketerie (9)	koketérie (53)
17	kyberie (1)	kybérie (2)
18	latimerie (19)	latimérie (15)
19	mašinerie (24)	mašinérie (323)
20	materie (194)	matérie (46)
21	menažerie (17)	menažérie (2)
22	mizerie (46)	mizérie (174)
23	mysterie (17)	mystérie (25)
24	oranžerie (27)	oranžérie (17)
25	palmerie (1)	palmérie (5)
26	parfumerie (95)	parfumérie (3)
27	periferie (571)	periférie (144)
28	pikanterie (8)	pikantérie (18)
29	pizzerie (221)	pizzérie (1)
30	pruderie (17)	prudérie (9)
31	rafinerie (327)	rafinérie (122)
32	scenerie (313)	scenérie (326)
33	tapiserie (101)	tapisérie (53)

Které jednotky obsahují hledaný stavební prvek a kdy jde o případy přegenerování?

Určitě ano: 3, 4, 6, 9, 14, 16, 19, 26, 28, 29, 31, 32

Ověříme: 5, 10, 13, 21, 24, 25, 33

1, 2, 7, 8, 11, 13, 15, 17, 18, 27, 30

12, 20, 22, 23

Jaká jsou pravidla pro užití e/é?

Psaní samohlásek v zakončení přejatých slov ([Skrýt](#))

1. [Dublety](#)
2. [Slova s dlouhou samohláskou](#)
3. [Slova s krátkou samohláskou](#)

Délka vyslovované samohlásky v zakončení slov není v mnoha případech zcela jednoznačná. Kvantita je pevná pouze u některých slov (např. *guvernér, bazén, akvamarín, gobelín, angína, vegetarián, vitrína*, viz [bod 2](#)), u jiných kolísá, popř. se vyslovuje polodélka. Důsledné zkrácení samohlásky v grafickém zápisu vybraných typů zakončení (viz [bod 1](#)) navržené v úvodních kapitolách PČP z roku 1993 (Psaní slov přejatých) a zachycené ve slovníkové části PČP neznámá změnu ve výslovnosti – lze vyslovovat krátce i dlouze, což bývá uvedeno u příslušných slov v hranatých závorkách.

Dublety

Platná PČP doporučují psát mnohá slova se zakončením *-en, -in, -iv, -iva, -ivum, -ivní, -emie, -erie, -on, -onek, -ona, -onka, -ped* jen krátce. Podle Dodatku k PČP ve slovech, kde PČP 1993 uvádějí jen psaní se samohláskou krátkou a předchozí PČP jen se samohláskou dlouhou, je psaní se samohláskou dlouhou rovněž správné. Dodatek tedy vrací více možností zápisu, ale bohužel za cenu drobných nejasností. Úvodní věta (5. bod Dodatku) není zcela v souladu s některými uvedenými příklady. Příklady se netýkají pouze slov psaných původně jen se samohláskou dlouhou, ale jsou zde zahrnuty i původní dublety (*benzín/benzín, penicilin/penicilín, archiv/archív, pasivní/pasívní, citron/citrón, bižuterie/bižutérie* apod.). Proto se respektuje dvojitá možnost zápisu jak u slov psaných podle kodifikace z roku 1957 jen se samohláskou dlouhou (*balón, faraón*), tak psaných dlouze i krátce (*benzín/benzín, vitamín/vitamin, acetylen/acetylen, vagón/vagon, kamión/kamion*).

Slova s dlouhou samohláskou

Některá slova se zakončením uvedeným v [bodě 1](#) mají natolik ustálenou délku samohlásky, že je píšeme v tomto zakončení pouze s dlouhou samohláskou. Jsou to např. slova:

-én: *bazén, drén, fenomén, gabardén, holocén, kretén, paleocén, pleistocén, refrén, suterén, suverén,*

-ín: skupina slov označujících textilní látky – *cibelín, etamín, krepdešín, mušelín, popelín, stramín, vatelín*; bytosti – *beduín, fílištín, cherubín, jakobín, kapucín, mandarín, mouřenín, rabín*; barvy – *akvamarín, karmín* a dále např. slova *cukerín, gobelín, hermelín, magazin, mokasín, naftalín, parafín, tramín, zepelín,*

-érie: *sibérie,*

-ón: *archón, fanfarón, kujón,*

-óna: *koróna.*

S dlouhou samohláskou píšeme rovněž většinu přejatých slov se zakončením **-ína:** *angína, balerína, balzamína, cesmína, disciplína, doktrína, drezína, figurína, kolombína, konkubína, krinolína, latrína, limuzína, mandolína, medicína, okarína, pelerína, plastelína, primabalerína, tamburína, trampolína, turbína, vazelína, vitrína.*

Pozn.: Krátce však píšeme např. slova *cukina, gilotina, pagina, serpentina, vagina, želatina*.

Slova s krátkou samohláskou

Jen krátce píšeme zakončení slov na *-ura*, např. *agentura, akupresura, aparatura, bordura, bravura, brožura, cenzura, expozitura, faktura, fraktura, inventura, karikatura, klauzura, klaviatura, korektura, kubatura, kultura, kvadratura, literatura, miniatura, nomenklatura, oponentura, ouvertura, partitura, purpura, receptura, struktura, tonzura*, a *-urka* (např. *brožurka*), dále *-urní*, např. *agenturní, bravurní, inventurní, kulturní, skulpturní, strukturní, texturní*, a *-urový*, např. *armaturový, fakturový, partiturový*.

Pouze krátce píšeme rovněž zakončení slov na *-una* (*harpuna, kalafuna, laguna, tribuna*).

Podívejme se též na existenci slov, pro která nebyly nalezeny obě varianty způsobu psané *e/é*.

(Vzor 1 a Vzor 2 nevyžlucená slova, viz obrázek 3 a obrázek 4 seřazení podle frekvence).

Která slova patří do námi hledané skupiny slov a která nikoli?

Jaké další významy se objevují ve sledované skupině slov se stavebním prvkem *erie/érie*.

Jaké morfologické vlastnosti postrádají jednotky, které jsou přegenerované?

Jak posuzovat slova jako *bat(t)erie, artilerie* na pozadí slov jako *kavalerie, infanterie, banderie*?

Co je to *musarderie* ?

Je možné automaticky vyloučit slova v cizojazyčných textech ponechaných v jednojazyčném korpusu ?

Co jsou to paronyma a jaký je jejich dopad na automatickou morfologickou analýzu? Víte, co je to *nezoberie*?

Je možné, že stavební prvek *erie/érie* má své místo v české slovo tvorbě skrze módní názvy gastro-zařízení? Jak nám mohou pomoci korpusy?

Víte, co je to *gregerie*?

Obrázek 3

abc/cba ▲▼	fq ▲▼
galerie	6574
bakterie	2451
baterie	2111
periferie	571
hysterie	538
loterie	529
rafinerie	327
scenerie	313
fotogalerie	250
pizzeria	221
bižuterie	220
materie	194
kavalerie	145
tapiserie	101
parfumerie	95
jerie	94
drogerie	93
autobaterie	59
galanterie	59
draperie	48
mizerie	46
musarderie	30
artilerie	28
oranžerie	27
mašinerie	24
listerie	23
cherie	19
latimerie	19
santerie	19
pedanterie	18

Obrázek 4

abc/cba ▲▼	fq ▲▼
série	7348
karosérie	794
nigérie	373
scénérie	326
mašinérie	323
drogérie	238
valérie	223
mizérie	174
prérie	152
periférie	144
rafinérie	122
artérie	101
libérie	101
baktérie	61
koketérie	53
tapisérie	53
minisérie	52
matérie	46
galérie	44
drapérie	41
mystérie	25
kafilérie	22
férie	21
chérie	21
pikantérie	18
oranžérie	17
latimérie	15
bižutérie	10
kafetérie	9
prudérie	9
sibérie	7
palmérie	5
krimisérie	4
dyzentérie	3
fantasy/sci-fisérie	3
feérie	3
hatérie	3
kotérie	3
nevkusérie	3

Zdroje:

Slavičková, E.: Některé problémy morfémové analýzy češtiny. *Slovo a slovesnost* 28 (1967), 1, 7-17.

Rejzek, J: Český etymologický slovník, Praha 2001.

IJP: <http://prirucka.ujc.cas.cz/?slovo=drogerie#bref2>