



Distribuce grafických variant e/ě v české substantivní flexi

Klára Osolsobě

osolsobe@phil.muni.cz




MSoČ a korpusy

- Cílem přednášky je prakticky ukázat, že ačkoliv nová korpusová mluvnice češtiny (Cvrček, V. a kol., 2010) nezahrnula řadu informací, které lze z korpusů vyčíst, není jejím vydáním možnost využívat korpusy i nadále jako zdroje observací jazyka nikterak potlačena.



Distribuce -e/-ě v substantivní flexi

- Na s.174 se uvádí, že „Ke vzoru *duše* patří feminina s koncovkou -e někdy psanou -ě ...“,
- takřka stejná formulace se objeví na s. 188 a 189 (vzor *moře* a *kuře*).
- u vzoru *soudce* (s. 160n.)
- u vzoru *píseň* (s. 178n.)



Je psaní –e/-ě ve flektivních koncovkách popsateľné obecně platnými pravidly ?





















- Studenti bohemistiky (rodilí mluvčí) odpověď hledají s jistými obtížemi.
- Odpověď na tuto otázku komplikuje (protože jde opravdu o odpověď komplikovanou) řešení některých oblastí počítačového zpracování přirozeného jazyka (konkrétně češtiny).









Korpus SYN2010

- ▶ lc=„.*[eě]“
- ▶ P-filtr
- ▶ tag=„NN.*“
- ▶ konkordance – statistiky – frekvenční distribuce
- ▶ **Korpus SYN2010:**
Český národní korpus - SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <<http://www.korpus.cz>>.

[|c=".*[eě]" & tag="NN.*"]

		lemma	Frekvence	Freq [%]	
1.	p/n	rok	67 567	1.96	
2.	p/n	práce	43 269	1.25	
3.	p/n	doba	42 075	1.22	
4.	p/n	případ	38 629	1.12	
5.	p/n	země	28 039	0.81	
6.	p/n	peníze	26 824	0.78	
7.	p/n	ruka	26 064	0.76	
8.	p/n	dítě	25 096	0.73	
9.	p/n	strana	24 114	0.7	
10.	p/n	místo	23 922	0.69	
11.	p/n	svět	23 476	0.68	
12.	p/n	policie	23 159	0.67	
13.	p/n	den	22 048	0.64	
14.	p/n	Praha	20 750	0.6	
15.	p/n	dveře	20 491	0.59	
16.	p/n	informace	19 752	0.57	
17.	p/n	situace	19 681	0.57	
18.	p/n	muž	16 933	0.49	
19.	p/n	konec	15 979	0.46	
20.	p/n	srdce	15 018	0.44	

Závislost na okolí

		word	Frekvence	Freq [%]	
1.	p/n	roce	67 498	1.96	
2.	p/n	době	42 022	1.22	
3.	p/n	práce	41 318	1.2	
4.	p/n	případě	38 590	1.12	
5.	p/n	ruce	25 264	0.73	
6.	p/n	peníze	24 506	0.71	
7.	p/n	straně	24 029	0.7	
8.	p/n	místě	23 888	0.69	
9.	p/n	světě	23 115	0.67	
10.	p/n	země	22 917	0.66	
11.	p/n	Praze	20 622	0.6	
12.	p/n	dne	20 128	0.58	
13.	p/n	dveře	18 514	0.54	
14.	p/n	policie	18 334	0.53	
15.	p/n	dítě	18 296	0.53	
16.	p/n	informace	17 981	0.52	
17.	p/n	situace	17 821	0.52	
18.	p/n	muže	16 447	0.48	
19.	p/n	konce	15 842	0.46	
20.	p/n	životě	14 381	0.42	

[lɔ=".*[aeěiouyáéíóůúý][eě

ɚ"]

1.	p/ n	policie	18 334	6.92	
2.	p/ n	energie	10 614	4.01	
3.	p/ n	unie	7 198	2.72	
4.	p/ n	Marie	7 137	2.69	
5.	p/ n	historie	5 785	2.18	
6.	p/ n	fotografie	5 777	2.18	
7.	p/ n	teorie	5 759	2.17	
8.	p/ n	Francie	5 024	1.9	
9.	p/ n	Policie	4 714	1.78	
10.	p/ n	technologie	4 560	1.72	
11.	p/ n	kategorie	4 144	1.56	
12.	p/ n	demokracie	3 910	1.48	
13.	p/ n	studie	3 620	1.37	
14.	p/ n	série	3 365	1.27	
15.	p/ n	Británie	2 897	1.09	
16.	p/ n	Itálie	2 880	1.09	
17.	p/ n	Anglie	2 832	1.07	
18.	p/ n	strategie	2 602	0.98	
19.	p/ n	Lucie	2 596	0.98	
20.	p/ n	akcie	2 432	0.92	

Korpus SYN2010







C+[eě]

77.	p/n	vývoje	5 585	0.18	■
78.	p/n	prosince	5 575	0.17	■
79.	p/n	údaje	5 545	0.17	■
80.	p/n	dítěte	5 375	0.17	■
81.	p/n	komunikace	5 366	0.17	■
82.	p/n	naděje	5 300	0.17	■
83.	p/n	generace	5 291	0.17	■
84.	p/n	ředitele	5 243	0.16	■
85.	p/n	role	5 209	0.16	■
86.	p/n	šance	5 208	0.16	■
87.	p/n	potíže	5 192	0.16	■
88.	p/n	krve	5 177	0.16	■
89.	p/n	rodině	5 151	0.16	■
90.	p/n	kuchyně	5 095	0.16	■
91.	p/n	zdroje	5 019	0.16	■
92.	p/n	reakce	5 007	0.16	■
93.	p/n	jaře	5 003	0.16	■
94.	p/n	Země	4 978	0.16	■
95.	p/n	George	4 804	0.15	■
96.	p/n	týdne	4 650	0.15	■
97.	p/n	hřiště	4 622	0.15	■
98.	p/n	vesnice	4 568	0.14	■
99.	p/n	nebe	4 542	0.14	■
100.	p/n	pozice	4 499	0.14	■




soudce

[lemma=".*[eě]" & tag="NNM.*"]

		word	Frekvence	Freq [%]	
1.	p/n	zástupce	6 541	6.34	
2.	p/n	George	4 804	4.65	
3.	p/n	výrobce	4 243	4.11	
4.	p/n	soudce	3 444	3.34	
5.	p/n	obránce	2 960	2.87	
6.	p/n	vůdce	2 541	2.46	
7.	p/n	průvodce	2 005	1.94	
8.	p/n	zájemce	1 984	1.92	
9.	p/n	Mike	1 907	1.85	
10.	p/n	poradce	1 793	1.74	
11.	p/n	Jake	1 487	1.44	
12.	p/n	Rose	1 485	1.44	
13.	p/n	správce	1 402	1.36	
14.	p/n	tvůrce	1 333	1.29	
15.	p/n	hrabě	1 230	1.19	
16.	p/n	nástupce	1 215	1.18	
17.	p/n	předchůdce	1 137	1.1	
18.	p/n	obhájce	1 099	1.06	
19.	p/n	Malone	1 074	1.04	
20.	p/n	vládce	1 008	0.98	

růže

[lemma=".*[eě]" & tag="NNF.*"]

		word	Frekvence	Freq [%]	
1.	p/n	práce	41 318	4.69	
2.	p/n	země	22 917	2.6	
3.	p/n	dveře	18 514	2.1	
4.	p/n	informace	17 981	2.04	
5.	p/n	situace	17 821	2.02	
6.	p/n	akce	11 895	1.35	
7.	p/n	nemocnice	11 443	1.3	
8.	p/n	organizace	11 377	1.29	
9.	p/n	funkce	9 557	1.09	
10.	p/n	ulice	9 257	1.05	
11.	p/n	chvíle	8 551	0.97	
12.	p/n	televize	7 562	0.86	
13.	p/n	radnice	7 528	0.85	
14.	p/n	komise	7 454	0.85	
15.	p/n	hranice	7 065	0.8	
16.	p/n	silnice	6 938	0.79	
17.	p/n	péče	6 643	0.75	
18.	p/n	duše	6 421	0.73	
19.	p/n	stanice	6 146	0.7	
20.	p/n	krize	5 696	0.65	

moře/kuře

[lemma=".*[eě]" & tag="NNN.*"]

		word	Frekvence	Freq [%]	
1.	p/n	dítě	18 296	10.31	
2.	p/n	srdce	13 861	7.81	
3.	p/n	slunce	8 606	4.85	
4.	p/n	moře	8 541	4.82	
5.	p/n	pole	7 145	4.03	
6.	p/n	finále	5 682	3.2	
7.	p/n	dítěte	5 375	3.03	
8.	p/n	hřiště	4 622	2.61	
9.	p/n	nebe	4 542	2.56	
10.	p/n	letišťe	3 877	2.19	
11.	p/n	zvíře	3 358	1.89	
12.	p/n	ovoce	3 220	1.82	
13.	p/n	Slunce	3 181	1.79	
14.	p/n	vejce	3 175	1.79	
15.	p/n	odpoledne	2 858	1.61	
16.	p/n	poledne	2 393	1.35	
17.	p/n	schodiště	2 375	1.34	
18.	p/n	parkoviště	2 356	1.33	
19.	p/n	semifinále	2 058	1.16	
20.	p/n	děvče	1 979	1.12	

píseň/kost

[lemma!=".*[aáeěéííoóuúůý]" &
tag="NNF.*"]

		word	Frekvence	Freq [%]	
1.	p/ n	soutěže	10 010	6.02	
2.	p/ n	obce	8 843	5.32	
3.	p/ n	tváře	7 082	4.26	
4.	p/ n	kanceláře	6 636	3.99	
5.	p/ n	sítě	6 550	3.94	
6.	p/ n	daně	6 101	3.67	
7.	p/ n	postele	5 739	3.45	
8.	p/ n	potíže	5 192	3.12	
9.	p/ n	krve	5 177	3.12	
10.	p/ n	zbraně	3 673	2.21	
11.	p/ n	církvě	3 420	2.06	
12.	p/ n	věže	2 979	1.79	
13.	p/ n	úrovně	2 893	1.74	
14.	p/ n	mládeže	2 764	1.66	
15.	p/ n	písně	2 712	1.63	
16.	p/ n	větve	2 428	1.46	
17.	p/ n	Paříže	2 288	1.38	
18.	p/ n	lodě	2 010	1.21	
19.	p/ n	dlaně	1 906	1.15	
20.	p/ n	Plzně	1 892	1.14	

Analýza

- Existují grafémy, za kterými se v češtině nepíše v koncůvkách (zakončeních) substantiv ani –e, ani –ě. Jsou jimi *d', t', ň*.
- Existují grafémy, za kterými se v češtině píše v koncůvkách (zakončeních) substantiv vždy pouze –e. Jsou jimi *c, č, g, h, j, k, l, r, ř, s, š, z, ž*.
- Existují grafémy, za kterými se v češtině píše v koncůvkách (zakončeních) substantiv buď –e nebo –ě. Jsou jimi *b, d, f, m, n, p, t, v*.
- Existují grafémy, za kterými se v češtině píše v koncůvkách (zakončeních) substantiv buď –e nebo –ě, a to u téhož lemmatu. Plyne to z toho, že počet všech lemmat není vždy totožný se součtem lemmat, u nichž je buď jedna, nebo druhá varianta. Dle sledovaného korpusu jsou jimi *b, d, m, n, t, v*.

Tabulka

	soudce	duše	píseň	moře	kuře
b[eě]	<u>Vosolsobě</u>	0	0	nebe	hrabě
d[eě]	-	hýždě	lodě	?rande	hádě
f[eě]	-	0	0	kafe	0
m[eě]	-	země	země	sémě	0
n[eě]	Bechyně	kuchyně	daně	poledne	štěně
p[eě]	-	koupě	0	kanape	doupě
t[eě]	-	kleště	sítě	letišťě/?karate	dítě
v[eě]	-	0	církvě	0	0

Analýza

- ▶ Substantiva skloňovaná podle vzorů *soudce*, *růže*, *kuře* mají (na základě korpusových dokladů) po grafémech [bd(f)mnpt(v)] koncovku –e vždy realizovanou jako grafické –ě.
- ▶ Substantiva skloňovaná podle vzoru *píseň* mají (na základě korpusových dokladů) po grafémech dtň [d'ň] koncovku –e vždy realizovanou jako grafické –ě.
- ▶ Substantiva skloňovaná podle vzoru *moře* mají (na základě korpusových dokladů) po grafému t[t'] koncovku –e vždy realizovanou jako grafické –ě, přičemž jde vždy o sufix –iště.



.*[bpfvmdtn]

Striktní respektování těchto zásadních principů ze strany **církve** i civilních uskupení vytváří pevný základ pro spolupráci.
mnou na podlaze v bezvědomí a kolem hlavy měla louži **krve** . " Horatio ohromeně škulbl hlavou . " V bezvědomí
Teď jim budou říkat : " Ne , nenoste ty **lahve** do kontejneru , co máte u domu , musíte je
propojení silnic I/17 a I/37 , tedy silnic vedoucích z **Chrudimě** na Pardubice a na Prahu . Trasa plánované přeložky
chodbě a díval se , jak se rozednívá , nejbližší **větve** buků v ústavní zahradě už vystoupily ze tmy , nejprve
celé toto vyprávění se stalo (alespoň pro členy pozdější **církve**) věroučnou legendou . Ať už se tyto události odehraly
n nabušení bodyguardi v šedých tričkách sebrali u vstupu **lahve** s vodou , která uvnitř stála čtyři eura . A
ná zajímavou historii , spjatou s prvními pokusy o převod **krve** . Na objevení krevních skupin má významný podíl če
Jiří Tichánek , poté jej vystřídala farářka Marta Silná z **Církve** československé husitské a nakonec o Mistru Husov
všechno akorát . To byla kurva . Vypil jsem půl **lahve** piva a přistoupil k ní . " Ty víš ,

[lemma=".*[bpfvm][eě]" &
tag="NNN *"]

	word	Frekvence	Freq [%]
1.	p/ n nebe	4 542	42.77
2.	p/ n Labe	1 382	13.01
3.	p/ n kafe	1 193	11.23
4.	p/ n Nebe	507	4.77
5.	p/ n břímě	263	2.48
6.	p/ n Adobe	194	1.83
7.	p/ n doupě	162	1.53
8.	p/ n rámě	159	1.5
9.	p/ n doupěte	151	1.42
10.	p/ n Dame	136	1.28
11.	p/ n hříbě	121	1.14
12.	p/ n SME	98	0.92
13.	p/ n konkláve	90	0.85
14.	p/ n sémě	88	0.83
15.	p/ n Luftwaffe	81	0.76
16.	p/ n Sme	79	0.74
17.	p/ n poupě	73	0.69
18.	p/ n markraběte	68	0.64
19.	p/ n Kafe	68	0.64
20.	p/ n AVE	64	0.6

Závěr

- Ke vzoru *píseň* patří skupina substantiv zakončených na *-ev*, u nichž se koncovka *-e* vždy realizuje jako grafické *e*.
- Ke vzoru *píseň* patří několik málo substantiv zakončených na *-m* u nichž se koncovka *-e* vždy realizuje jako grafické *ě*.
- Substantiva zakončená na [bfmpvdnt] patřící ke vzoru *moře* mají s výjimkou derivátů na *-iště* a skupiny substantiv *sémě*, *plémě*, *símě* koncovku *-e* realizovanou jako grafické *-e*.
- Jde o poměrně malý počet substantiv. Nicméně se většinou jedná o substantiva poměrně frekventovaná.

Závěr

- ▶ Můžeme je tudíž definovat výčtem, přičemž s ohledem na rozsah korpusu můžeme předpokládat relativní úplnost výčtu frekventovaných jednotek.
- ▶ Vzhledem k tomu, že distribuce variant je alespoň u vzorů *píseň* a *moře* vázána nikoliv na distribuci danou grafickým okolím, ale na jednotlivé skupiny lexému, je třeba připustit, že v češtině existují u některých vzorů dvě varianty koncovek *-e/-ě* a že tyto varianty nejsou grafickými variantami v témže smyslu, jako jsou jimi varianty *-e/-ě* u jiných vzorů.



Závěr

- ▶ Na základě pozorování dat získaných z korpusů je možné odpovídat na otázky, které před zvědavými čtenáři MSoČ mohou vyvstat.
- ▶ Dopátrat se žádoucích odpovědí není vždy snadné, je ale žádoucí.



Bibliografie

- Cvrček, V. a kol.: *Mluvnice současné češtiny*. Praha : Nakladatelství Karolinum. 2010.
- Kosek, P. – Křístek, M . – Osolsobě, K. – Vojtová, J. – Žiková, M.: Recenze: Cvrček, V. a kol.: *Mluvnice současné češtiny*. Praha : Nakladatelství Karolinum. 2010. *Naše řeč* 3/94, 2011, s. 144-155.



Odkazy na „online“ zdroje

- ▶ Český národní korpus - SYN2000/SYN2005/SYN2006PUB/SYN. Ústav Českého národního korpusu FF UK, Praha 2000. K vyhledání on-line <<http://ucnk.ff.cuni.cz>>. (<http://ucnk.ff.cuni.cz/bonito/>)
- ▶ Rychlý, P.: Bonito – grafické uživatelské rozhraní systému Manatee, Verze 1.49. 1998-2003. K vyhledání on-line <http://ucnk.ff.cuni.cz/bonito/>