

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	



`[(lemma!=„.*[aáí]" & lc=".*[eě]" & tag="NNF.*") | (lemma="[eě]" & lc=".*[eě]" & tag="NN.*")]`

Celkem: 8270 (166 str.)

	Filtr	lemma	Frekvence
1.	p/n	práce	42 182
2.	p/n	země	25 637
3.	p/n	dveře	21 003
4.	p/n	situace	18 735
5.	p/n	informace	17 031
6.	p/n	energie	15 303
7.	p/n	policie	13 793
8.	p/n	akce	11 877
9.	p/n	funkce	10 744
10.	p/n	nemocnice	10 206
11.	p/n	organizace	9 048
12.	p/n	péče	8 847
13.	p/n	chvíle	8 800
14.	p/n	obec	8 529
15.	p/n	soutěž	8 381



Před ě nemůže předcházet

- Vokál
- Některé konsonanty

[lɔ=".*[aeěiouyáéíóůúý][eě

ɯ]

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 SYN2015

C+[eě]

Celkem: 6751 (136 str.)

	Filter	lemma	lc [lowercase word]	Frekvence
1.	p/n	práce	<u>práce</u>	42 182
2.	p/n	země	<u>země</u>	25 636
3.	p/n	dveře	<u>dveře</u>	20 991
4.	p/n	situace	<u>situace</u>	18 735
5.	p/n	informace	<u>informace</u>	17 031
6.	p/n	akce	<u>akce</u>	11 877
7.	p/n	funkce	<u>funkce</u>	10 744
8.	p/n	nemocnice	<u>nemocnice</u>	10 206
9.	p/n	organizace	<u>organizace</u>	9 026
10.	p/n	péče	<u>péče</u>	8 847
11.	p/n	chvíle	<u>chvíle</u>	8 800
12.	p/n	obec	<u>obec</u>	8 529
13.	p/n	soutěž	<u>soutěže</u>	8 381
14.	p/n	tvář	<u>tváře</u>	8 297
15.	p/n	ulice	<u>ulice</u>	8 148
16.	p/n	síť	<u>sítě</u>	7 560
17.	p/n	hranice	<u>hranice</u>	7 388
18.	p/n	konstrukce	<u>konstrukce</u>	7 072
19.	p/n	generace	<u>generace</u>	6 928
20.	p/n	postel	<u>postele</u>	6 793

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 mnou na podlaze v bezvědomí a kolem hlavy měla louži

církv

i civilních uskupení vytváří pevný základ pro spolup

krve

. " Horatio ohromeně škulbl hlavou . " V bezvědomí

Ted' 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 , nejprv

celé toto vyprávění se stalo (alespoň pro členy pozdější

církv

) věroučnou legendou . Ať už se tyto události odehra

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

Církv

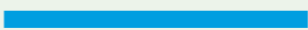

















československé husitské a nakonec o Mistru Husov

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 zvidavými 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/>