

Gramatika a korpus II plin032

Středa: 7.30-9.00 G13

23. 3. Rozbor dú.

Efektivní vyhledávání homonym slovesných tvarů pomocí počítačových nástrojů II. (I když pořád spěj, pořád někam spěj ...)

Dluž a splácej!¹

V korpusech (jakkoliv jsou ty, které dnes máme k dispozici, rozsáhlé) není zachycení tvarů imperativu takové, jaké bychom si přáli. Důvody vidíme jednak v omezení výskytu tvarů imperativu, které plyne přirozeně z významů jednotlivých sloves (rozkazovat k činnosti pojmenované příslušným slovesem je mnohdy podivné až nesmyslné), jednak v omezeních pragmalingvistických (imperativ je vnímán jako tvar, kterému se v řadě komunikačních situací snažíme spíše vyhnout). Problematické je ovšem pro automatickou morfologickou analýzu i to, že v řadě případů je tvar imperativu (2. os. sg.) homonymní s nějakým často zcela významově naprosto vzdáleným (krátkým) slovem, či zkratkou (*kop, oř, kos, jez, nos, kup, div, děj, lež, boj, pas, per, tel, ...*), které bývá frekventovanější, než potenciální imperativní tvar. S tím souvisí, že morfologické analyzátoři s potenciálními tvary imperativu pracují mnohdy selektivně (technické omezení rozgenerování tvarů, které opakovaně způsobují chyby v lemmatizaci – srov. např. výskyt lemmatu *telit* v korpusu SYN2000). Tyto skutečnosti je třeba mít na zřeteli, chceme-li se věnovat konzistentnímu zachycení variantních tvarů imperativu.

V tomto semináři se pokusíme najít pomocí nástroje Morfio podklady pro uvedený typ homonym. Všimneme si, že nemusí vždy jít jen o homonyma a podíváme se, jak vypadá disambiguace.

Dluž a splácej!

Jaké dvojice obecně představují slovesa podobná těm, která vidíme v uvedeném případě?

- 1) slovesa s kořeným *ou*, jejichž imperativ má alternaci (*ou/u*: **dloužit/dluž**) a jsou tedy homonymní se slovesy s kořeným *u* (**dlužit**).
- 2) Slovesa IV. třídy (sázet: **splácejt**) a slovesa V. třídy (dělat: **splácejt**) mohou mít homonymní tvary imperativu (**sázej/dělej**: **splácej**).

Použijeme nástroj Morfio: (Odkaz na toto zadání: <http://morfio.korpus.cz/Od5waQ4X>)

¹ Hezkým dokladem homonymních imperativů od různých sloves je lichvářský slogan, na který upozornila J. Hlaváčová: *Dluž a splácej*.

<+ ~~společný~~ ~~odlišný~~ ~~společný~~ +> Morf. specifikace:

vzor 1: .[*o] ou .*t slovesa V.*

vzor 2: .[*o] u .*t slovesa V.*

Další vzor

Korpus: SYN2010 Frekvence vyšší než: 0 Hledají se: lemmata Vyhodnocují se: lemmata

Velikost písmen: ignorovat Alternace

Hledat zachovat záznam procesu

Odkaz na toto zadání: <http://morfio.korpus.cz/0d5waQ4X>

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	koupit (16811)	kupit (276)
2	toužit (6859)	tužit (46)
3	koukat (5170)	kukat (22)
4	nakoupit (2022)	nakupit (147)
5	poučit (2015)	pučit (60)
6	proudit (1814)	prudit (42)
7	poučovat (560)	pučovat (7)
8	ploužit (292)	plužit (13)
9	bouchat (270)	buchat (3)
10	loupat (202)	lupat (9)
11	šoupat (102)	šupat (1)
12	zakoukat (91)	zakukat (18)
13	louskat (84)	luskat (25)
14	chroupat (78)	chrupat (33)
15	vypoulit (40)	vypulit (1)
16	vyboulit (38)	vybulit (2)
17	dloužit (31)	dlužit (1649)
18	vylouhovat (29)	vyluhovat (62)
19	čoudit (26)	čudit (1)
20	doploužit (13)	doplužit (1)
21	nakoukat (9)	nakukat (78)
22	boulit (7)	bulit (37)
23	křoupat (7)	křupat (93)
24	čmoudit (5)	čmudit (1)
25	pomoučit (4)	pomučit (2)
26	příploužit (3)	příplužit (1)
27	chřoupat (2)	chřupat (6)
28	natoužit (2)	natužit (4)
29	houlit (1)	hulit (81)
30	vykoulit (1)	vykulit (278)
31	vyšoupat (1)	vyšupat (3)

31 dvojic. Všimněte si, že některé dvojice by bylo možné klasifikovat jako synonyma. Které?

<+ společný odlišný +> Morf. specifikace:

vzor 1:

vzor 2:

Další vzor

Korpus: Frekvence vyšší než: Hledají se: Vyhodnocují se:

Velikost písmen: ignorovat

zachovat záznam procesu

Odkaz na toto zadání: <http://morflo.korpus.cz/QsiWT4Jv>

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	svět svet (81911)	svat (7)
2	pet pět (39603)	pat (903)
3	opet opět (31809)	opat (797)
4	deset (29721)	desat (1)
5	sedět (28898)	sedat (896)
6	držet (25027)	držat (1)
7	smět smet (19992)	smat (1)
8	ret (8832)	rat (44)
9	hledět (7744)	hledat (25254)
10	čtyřicet (6630)	čtyřicat (1)
11	příspět (6154)	příspat (68)
12	uspět (5469)	uspat (299)
13	dospět (5160)	dospat (60)
14	bolet (5110)	bolat (1)
15	zavrtět (5040)	zavrtat (218)
16	kráčet (4448)	kráčat (2)
17	doprovázet (4145)	doprovázat (1)
18	provázet (3778)	provázat (156)
19	loket (2719)	lokat (28)
20	street (2574)	streat (1)
21	zacházet (2572)	zacházat (1)
22	chvět (2119)	chvat (262)
23	zahledět (1761)	zahledat (2)
24	prospět (1662)	prospat (481)
25	vyplácet (1616)	vyplácet (8)
26	set (1497)	sat (39)
27	splácet (1199)	splácet (17)
28	vrtět (1149)	vrtat (744)
29	zet (778)	zat (26)

Celkem nalezeno 192 dvojic

Problémy ručního třídění a pozorování vlastností kandidátů na uvedené případy homonymie:

- a) sloveso na [eě]t nemusí mít imperativ na [eě]j (*vrtět/vrtat*)

- b) sloveso na at nemají alternace (jotaci) labiálních finál v kontextu imperativního –ej (supět, supat, zasípět/zasípat)
- c) sloveso na –at není slovesem V. třídy (nepravidelné –*spat*: *prospět/prospat*)

Vyber slovesa, kterých se možné homonymie týkají a sleduj desambiguaci potenciálně homonymních tvarů v korpusu SYN.

Problémy: Více větších korpusů.

Vyzkoušejme řadu Araneum, CzTenTen a internet.

DÚ:

1. Existují i jiné případy krácení kořenného vokálu v imperativu, které mohou zapříčinit vznik tvarové homonymie?
2. Kdy se kořenný vokál krátí? Lze popsat pravidlem, nebo je krácení podmíněno lexikálně?
3. Homonymní čtení má i slogan: ***Nevaž se, odvaž se!*** (Chcete hubnout, nebo trojčit?). Podívejte se, jak je v případě sloves, která mají homonymní tvary, vyřešena disambiguace v korpusu SYN.