

Gramatika a korpus II plin032

Středa: 7.30-9.00 G13

9. 3. Rozbor dú.

Efektivní vyhledávání homonym slovesných tvarů pomocí počítačových nástrojů I. (Koupil jsem si paštiku, ...)

Mějme větu, které se většina Čechů spontánně zasměje:

Koupil jsem si paštiku. Bylo na ní napsáno: zaječí. A neječí a neječí.

Proč je tato věta směšná? Inu proto, že se v ní vyskytuje jedno jediné slovo/ jeden jediný slovní tvar, který lze číst dvěma způsoby, tedy tvarové homonymum (více Petkevič 2015).

Mluvčí zvolil jedno z možných čtení a stěžuje si na důsledek své volby.¹

Formu, která je příznačná pro uvedené příklady jazykové komiky, můžeme gramaticky popsat:

- (a) *zaječí*= sloveso 3. osoby singuláru indikativu prezentu aktiva slovesa *zaječet*; negované sloveso (*ne*)*ječet* (=nevydávát výrazný zvuk) příbuzné se slovesem *zaječet* (=vydat výrazný zvuk).
- (b) *zaječí* = tvar adjektiva měkkého (*zaječí*=příslušející drobnému hlodavci *zající*) v několika pádech a rodech;

Jsou podobné případy pro češtinu typické, nebo jde jen o náhodu, kolik je podobných případů?

K tomuto účelu můžeme použít jazykový korpus. Jak? Nástroj *Morfio* umožní vyhledat v korpusu o rozsahu 100 milionů slovních tvarů² všechny tvary, které lze definovat a) na základě jejich formy a b) na základě jejich gramatických vlastností. Slovní formulace zadání by mohla vypadat následovně:

Chceme vyhledat dvojice (a) sloveso+ (b) adjektivum takové, že pro ně platí, že jsou identické v příslušných tvarech. V případě slovesa víme z gramatiky, že (a) slovesa, která ve 3. osobě singuláru indikativu prezentu aktiva končí na *í*, jsou slovesa, která v infinitivu končí na *[jeě]t*. V případě (b) adjektiva víme, že jde o měkké adjektivum, takže slovníkový tvar a řada dalších tvarů končí na *í*.

Formální zadání při práci s aplikací *Morfio* vypadá následně:

¹ Mluvčí byl zklamán ve svém očekávání a my se mu smějeme. Jak to? Věť: „*Koupil jsem si robot, bylo na něm napsáno: zaječí. A neječí a neječí.*“ se ovšem nesmějeme. Proč?

² Budeme pracovat s korpusem SYN2010 (<http://wiki.korpus.cz/doku.php/cnk:syn2010>).

<+ společný odlišný >+ Morf. specifikace:

vzor 1: + [ieět] slovesa V.*

vzor 2: + í přídavná jména A.*

Další vzor

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

Velikost písmen: ignorovat

Hledat zachovat záznam procesu

Odkaz na toto zadání: <http://morfio.korpus.cz/3V2CbhNa>

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	vlastnit (2438)	vlastní (58875)
2	zaječet (968)	zaječí (309)
3	plazit (781)	plazí (45)
4	svištět (288)	svišťi (63)
5	supět (157)	supí (45)
6	kosit (137)	kosí (48)
7	okounět (91)	okouní (4)
8	opičit (76)	opičí (334)
9	otročit (57)	otročí (14)
10	čertit (48)	čertí (14)
11	odvodnit (40)	odvodní (44)
12	rozvodnit (39)	rozvodní (3)
13	beranit (38)	beraní (52)
14	kurvit (21)	kurví (6)
15	jinačit (17)	jinačí (100)
16	mravenčit (14)	mravenčí (279)
17	rožnit (12)	rožní (4)
18	hutnit (10)	hutní (322)
19	zbožnit (6)	zbožní (68)
20	nervóznit (4)	nervózní (2577)
21	vlčet vlčit (4)	vlčí (575)
22	prasečit (3)	prasečí (888)
23	splavnit (3)	splavní (2)
24	letnit (2)	letní (8790)
25	štířit (2)	štíří (5)
26	vážnět (2)	vážní (13)
27	čapět (1)	čapí (56)
28	kohoutit (1)	kohoutí (148)
29	vdovit (1)	vdoví (5)
30	výřit (1)	výří (2)
31	zemnit (1)	zemní (1698)

Výsledkem hledání v korpusu je seznam 31 dvojic.

Ke každému členu dvojice můžeme kliknutím zobrazit konkordanční seznam výskytů v kontextech. Seznam je seřazen podle frekvence (čísla v závorkách) slov v prvním sloupci. Frekvence je relevantní pro

představu o tom, zda jde o slovo v češtině obvyklé/neobvyklé. I to je zajímavá informace³. Zamysleme se třeba nad tím, jak bychom interpretovali syntagma „*kajmanka supí*“?

Viděli jsme, že počítačová aplikace *Morfio* užívala k vyhledávání znalosti o české gramatice (znalost slovních druhů). Ty se v korpusech češtiny nevzaly jen tak z ničeho nic. Byly do textů, z nichž je složen korpus, dodány (v procesu **automatické morfologické analýzy**, tedy závisejí na **použitém morfologickém slovníku a použité metodě/metodách disambiguace**).

Podívejme se nyní, jak si s případem dvojice (a) *zaječí/zaječet* a (b) *zaječí/zaječí* poradila automatická morfologická analýza. Na obrázku vidíme konkordanční seznam tvaru *zaječí* spolu s lemmaty a gramatickými tagy. První písmeno v tagu je *A* – adjektivum, nebo *N* – substantivum (název obce) a dále *V* – sloveso. Vyšlucené řádky na prvním seznamu jsou případy, kdy jde o sloveso, ovšem ve značce je chybně uvedeno, že jde o adjektivum a tudíž i lemma je chybně.

, ale stoprocentně do toho zažraný . Rance byl moje	zaječí /zaječí/AAFS1----1A----	pacička pro štěstí , a byl nejlepší v celé divizi
se mi narodil první kluk , a ač jsem měla	zaječí /zaječí/AAIP4----1A----	úmysly ještě před porodem , odjet do zahraničí s takhle
, Country Life nebo v drogerii DM . Zajíc načerno	Zaječí /zaječí/AAIP4----1A----	stehna prospíjeme asi třetinou slaniny , osolíme a opečeme je
ohyzná bábizně , popíjející bujón . " Kenny , "	zaječí /zaječí/AAFS1----1A----	ženština , " kde pořád vězíš ? Řekla jsem rybu
a zdálo se jim , že se svět zmenšuje a	zaječí /zaječí/AAFP4----1A----	uši prodlužují . " Jak je to možné , "
maso : všechna zobrazovaná zvířata se hodí k jídlu (zaječí /zaječí/AAFS1----1A----	, smřčí , výjimečně ryba) , a víme přece
sběr Vinařství U Kapličky , Morava , Velkopavlovická podoblast ,	Zaječí /zaječí/NNNS1----A----	, U Kapličky , 349 Kč (OCF , SON
osamělá . „ Mám v ruce jeho mozek , "	zaječí /zaječí/AAFS4----1A----	. A pak se postaví a vrávoravě vystoupí na sedadlo
, z nichž každého pojmenoval Tinker . První Tinker měl	zaječí /zaječí/AAIS4----1A----	pysk . Ten poslední provázel Davea všude a dokonce s
na hřbet myši a zamíří tam , kde dívku v	zaječí /zaječí/AAFS6----1A----	masce navštěvuje mladý muž . Zamíří na místo , které
. A další . Ze svatební hostiny . Pak někdo	zaječí /zaječí/AAFS1----1A----	. Určitě se nejedná ani o smích , ani o
navečer vybírat daň z pobytu v podobě dvou rousnic ,	zaječí /zaječí/AAMS1----1A----	mladík se neváhal posadit u špiček mých bot . Cítil
... Obsah oxidu uhličitého v kyselce činí 2300 mg/l .	Zaječí /zaječí/AAIS1----1A----	pramen a Devátá kyselka vyvěrají na zamokřených a zarostlých lukách
čelist , vyrval mu ji granát , a tak měl	zaječí /zaječí/AAIS4----1A----	pysk . Zpod řídkého kníru mu vyčníval z pootvěřených úst
17 hodin fotbalisty brněnské Zbrojovky . Tentokrát se představí v	Zaječí /zaječí/NNNS6----A----	proti slovenské Dunajské Středě . Zatím mají za sebou Briáné
co neexistuje . Nejčastěji se chodilo pro komáří sádlo ,	zaječí /zaječí/AAANS4----1A----	peří , obecní nůžky , hodinářskou motyku , červené hřebíky
proběhl v prostorách firmy Renishaw v Brně a poté v	Zaječí /zaječí/NNNS6----A----	u Brna odborný seminář společnosti technologysupport spojený s praktickými ukázkami
. Hodí přes bariéru primitivní kotvy s navázanými prostěradly .	Zaječí /zaječí/AAFS1----1A----	sířena a pohraničnicki zahájí palbu . Jedna z kulek zasáhne
1385) ; vzápětí pak zajistila , aby Postoupky i	Zaječí /zaječí/NNNS1----A----	případly jejím dětem (1385 , 1386) . Vic
vykřik . " Ty ses nejspíš úplně zbláznila , "	zaječí /zaječí/AAFS1----1A----	Beatrice . " Ve stodole se nekouří ! Chceš to
" Tak zůstaňe stát , nebo ne ? ! "	zaječí /zaječí/AAFP4----1A----	. Ale to mě nezastává . Osa našeho domu smějeje
je spisovatelka Most je uzavřený . Můžou za to trhliny	Zaječí /zaječí/AAANS1----1A----	- Nepříjemnou zprávu oznámili fiděům v Zaječí . Týká se
Bohoušův pobavený úsměv a pochopila , že přesně uhádí její	zaječí /zaječí/AAIP4----1A----	úmysly . Spiklenecky se na něj usklíbala a šla se
věty , nutí ho mluvit anglicky , při mluvení obrací	zaječí /zaječí/AAFS4----1A----	masku k modravé obrazovce monitoru a když skončí , pokaždě
záclony nejsou k jídlu , po čase proměnil obýváč v	zaječí /zaječí/AAFS4----1A----	noru , po které bez uzardění trouští bobky , čtrnáctiletý

Podívejme se nyní, jak vypadají konkordance případů, které byly označovány jako slovesa (V).

³ Slova s frekvencí 1 jsou velmi neobvyklá, o to mohou být pro poznání jazyka zajímavější. Např. sledujeme-li kontext slovesa *výřít*: ... *jak se to milostně zubí! S jakou by rozkoší zatkly mi asi do těla zuby. Hahaha to jsi ty drahoušku milý, a máš tak samotou výřít Počekj mne trochu začneme spolu chvíli. To budou na nás výřít!* ..., zjistíme, že jde o slovní hříčku, neboť v bezprostředním kontextu se nachází sloveso *vřít* = „dělat *vír*, rychle se krouživě pohybovat“, zatímco sloveso *výřít* není překlep, ale okazionálně utvořené sloveso napodobující kolokviální sloveso *vejrat* = „dívat se udiveně jako *výř*“. Jde o homofonii (v češtině slova lišící se *i/y* a *í/ý* znějí stejně), která je rovněž častým zdrojem českého humoru.

dozví, že její matka je obyčejná – JUDITA (zaječí /zaječet/VB-S---3P-AA---P	a otočí se na SIMONA) Mlč! Neříkej to
vykoukne . JACKIE spatří, jak se dveře pohnou a	zaječí /zaječet/VB-P---3P-AA---P	. SANDY vejde dovnitř . JACKIE Á , to jste
„ „ Co ? “ „ Fetus ! “	zaječí /zaječet/VB-S---3P-AA---P	a mrští miskou o stěnu . Ta se od ní
: Stella : Děvenko , dovol profesionální radu . Iris	zaječí /zaječet/VB-S---3P-AA---P	a uteče . Stella vypadá nemocně a unaveně . Volá
ani teď neodpoví , Yula se na něj podívá a	zaječí /zaječet/VB-S---3P-AA---P	mu do ucha : „ Stů tu , co ?
zpět na její židli . Ale ona se vytrhne ,	zaječí /zaječet/VB-S---3P-AA---P	, vrhne se zase na Violet , popadne ji za
má bývalá nevěrná žena , se ve mně pohne a	zaječí /zaječet/VB-S---3P-AA---P	vyšokým hlasem : „ Já jsem ti sice kalhoty nikdy

Zjistíme, že se chyby nevyskytují, nicméně konstatujeme, že jako sloveso je tvar označkován pouze 7krát, přičemž v 25 výskytech na předešlém seznamu jsme našli 6 chyb. Je tedy vidět, že disambiguace homonymií sledovaného typu je neradostná.⁴

V případě homonym bývá dodání jednoznačné interpretace „tvrdým oříškem“ a mnohdy „kamenem úrazu“.

Sledujme problémy značkování plynoucích z prokazatelné homonymie a všimněme si možností vylepšení.

Příklady konkordancí pro "[lemma="(?)supí" & tag="A.*"]" z korpusu syn2010

Le Carré, John,Ruská sekce	mnou vznáší málem dvoumetrová fousatá	supí	postava , v ruce flašku
Chase, James Hadley,Položte ji mezi lillemi	jsem , jak za mnou	supí	Kermani . Nastrčil jsem hlavu
Goodwin, Jason,Strom janičárů	zná lépe ; je to	supí	hnízdo , potřísněné špinou a
Goodwin, Jason,Strom janičárů	silnější : slepá sebranka a	supí	hnízdo , nebo velení ,
King, Stephen,Nezbytné věci	držící mikrofon , připomínala	supí	pařát . " Až po
Meyrink, Gustav,Golem	a sahal vysušenými prsty jako	supími	pařáty do strun harfy .
Exner, Milan,Svatoušek	svině , vy svině ,	supí	Sváta a všechno přikloň ,
Exner, Milan,Svatoušek	za níž v těsném závěsu	supí	láska ! A vidím -
Abagnale, Frank W. - ...:	měl přerostlé a zahnuté jako	supí	pařáty . Vůbec hodně připomínal
García Márquez, Gabri...:	kerá se dala vyléčit jediné	supí	žlučí , poněvadž byla nejen
García Márquez, Gabri...:	kůži zelenou od morfia a	supím	proflem : každým coulem představoval
McKinneyová, Meagan,Maskovaná kráska	dostatečně užila nákupů . "	Supí	pohled , jenž na ni
Márquez, Gabriel García,Všechna špina světa	natahovaly k nám své dlouhé	supí	krky . Za nimi .
Greene, Graham,Ministerstvo strachu	pod sněhobílými vlasy měl trochu	supí	, ušlechtilý a poněkud teatrální
Rendell, Ruth,Osudný zvrst	hitá benzin , když tak	supí	krajinou a dělá strašlivý kravál
Olson, Neil,Věčná ikona	s černými pažemi roztaženými jako	supí	křídla . Fakt , že
Woolf, Virginia,Deniky	k moři ty kopce se	supími	šijemi . Ale první noc
Watsonová, Caroli,Moje první encyklopedie	se k hostině . Roztažená	supí	křídla měří přes 2,5 metru
Kašpar, Oldřich - Mán...:	vejde , k tomu přidala	supí	dech , medvědí pot ,
Jungmann, Milan,Literárky můj osud	zní , když kolem člověka	supí	stádo budoucích feditelů . Tenkrát
Y,Lidé a země, č. 4/2002	při rozpravě historického Buddha na	Supím	vrchu , kde Buddha obklopený
Y,Lidé a země, č. 7/2003	klidu stopovat . Proti mně	supí	kolona obrněných vozidel SFOR .
Kantůrková, Eva - Cim...:	v horských kotlinách . Kolem	supí	nákladáky a jejich brzdomé obložení
Y,Lidé a země, č. 9/2006	do tradiční iránské domácnosti .	Supí	hory Abyaneh je ideální základnou
Y,Lidé a země, č. 9/2006	. Karkas znamená v překladu	Supí	hory . Geologové se nemohou

Zamysleme se, zda existují/mohou existovat v češtině další dvojice tohoto typu. Pokud ano, proč jsme je nenašli a jak můžeme na základě pozorování korpusových dat ověřit správnost/nesprávnost našich předpokladů?

Lingvisticky vzato je hledané slovesné lemma (.*[ieč]t) **lemmatem produktivní slovesné třídy (IV. tř. vzory prosit, trpět, sázet)**. Dále nám **lingvistická intuice** napoví, že produktivita třídy se mimo jiné opírá o tvoření **sloves od adjektiv s významem činit/stávat se takovým jako motivující adjektivum** (viz dvojice **vážnět/vážní** bez ohledu na doklady, které nejsou pro naši intuici nijak průkazné).

Lingvisticky vzato nám ovšem pozorování této dvojice napoví, že bylo **chybou vyhodnocovat pouze lemmata**. Proč? Inu proto, že **homonymní tvar nemusí mít pouze adjektiva měkká** (jako je **zaječí**), ale i

⁴ Výše uvedený seznam vygenerovaný aplikací *Morfio* je také omezen chybami v disambiguaci, protože pracuje na základě zadání, v němž je užito morfologické značkování (informace o slovních druzích hledaných lemmat).

tvrdá, u nichž je homonymní **nominativ plurálu maskulin životných** (tedy **vážní** jakožto tvar adjektiva **vážný**).

V aplikaci **Morfio** budeme tedy vyhodnocovat tvary a získáme 119 dvojic, přičemž vidíme, že tvrdých adjektiv je mezi nimi podstatné množství.

The screenshot shows the Morfio application interface with the following settings:

- společný (checked), odlišný (unchecked)
- Morf. specifikace: slovesa (dropdown), V+ (dropdown)
- vzor 1: + (dropdown), [ie]t (dropdown)
- vzor 2: + (dropdown), i (dropdown)
- Další vzor: (empty)
- Korpus: SYN2010 (dropdown), Frekvence vyšší než: 0 (input), Hledají se: tvary (dropdown), Vyhodnocují se: tvary (dropdown)
- Velikost písmen: ignorovat, zachovat záznam procesu
- Alternace (dropdown)
- Buttons: Hledat, Nové zadání, Nápověda
- Footer: Odkaz na toto zadání: <http://morfio.korpus.cz/Wi8V2PQt>

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	tvrdit (1981)	tvrdí (108)
2	plnit (1675)	plní (100)
3	stanovit (1627)	stanoví (1)
4	divit (1493)	diví (65)
5	živit (695)	živí (395)
6	vlastnit (482)	vlastní (39107)
7	čistit (402)	čistí (89)
8	pyšnit (243)	pyšní (292)
9	plazit (145)	plazí (35)
10	spořit (133)	spoří (7)
11	sušit (127)	suší (9)
12	mírnit (108)	mírní (25)
13	hladovět (93)	hladoví (186)
14	plavit (90)	plaví (6)
15	nepustit (79)	nepustí (1)
16	ranit (78)	raní (18)
17	slepit (77)	slepí (194)
18	klanět (70)	klaní (2)
19	zdravit (70)	zdraví (110)
20	marnit (68)	marní (4)
21	řečnit (53)	řeční (9)
22	plašit (51)	plaší (90)
23	šedivět (41)	šediví (37)
24	jistit (38)	jistí (500)
25	kalit (36)	kalí (3)

SSJČ

raný příld. (2. st. ranější)

- brzy ráno n. vůbec brzy nastávající, přicházející, vyskytující se; časný 1 (op. pozdní): v r-ch hodinách; být r. časné vstávat, r. podzim; zeměd., zahr. brzy zralý, dospělý: r-é plodiny, brambory, třešně; r-á sklizeň; r-é plemeno vepřů; r-á drůbež
- Jsoucí v počátcích něj. období, doby, vývoje; počáteční: r. středověk; r-á gotika, r-á díla Alšova první; r-é mládí
- zř. zast. a bás. ranní, jitřní: rosa r-á (Čech); r-é slunce červené (Ner.);
→ přísl. k 1, 2 raně: r. vstávat časné, brzy, r. kvetoucí sorty tulipánů; - r. gotická malba;
→ podst. ranost v. t.; — rozl. od ranný

Příklady konkordancí pro "[word="(?)raní" & tag="A.*"]" z korpusu syn2010

<p>Orwell, George,Uvnitř velryby Kracauer, Siegfried,Ornament masy Y,Austrálie Hauková, Jiřina,Básně Zahradníček, Jan,Pokoušení smrti Y,Zničilas všechno, m...: okcitánská milostná ... Vacková, Jarmila,Odpo...: Toman, Rolf (ed.),Gotika Toman, Rolf (ed.),Gotika Švecová, Milada (et a...: Darwin, Charles,O pohlavním výběru Vacková, Jarmila,Odpo...: Keene, Michael,Světová náboženství Keene, Michael,Světová náboženství Keene, Michael,Světová náboženství Y,Právo, 17. 5. 2007 Y,Lidové noviny, 5. 3. 2008 Y,Právo, 9. 4. 2009</p>	<p>blíž pravdě než Wellsův . Rani , pak to jsou Weberovi . rani . První stromy zde vysadili . rani , pleskne zobák vodu , rani srdce nastavovat krásám jež rani rani , avšak odpor vždy Amorovi rani příbuzenství se starým mistrem vyhlásili rani přispělo k tomu , že rani obr. str. 391) . Rani odřezávali maso od kostí . Rani se domnívat , že jejich rani příbuzenství se starým mistrem vyhlásili rani jsou zpravidla shrnovány pod označením Rani Syn Boha a Syn člověka Rani Jan a Petr i jiní rani května po 7 . hodině rani malého vzrůstu do Mississippi všichni rani de siecle s terorem . Rani</p>	<p>bolševici byli třeba anděly či kapitalisté . Oblíbeným terčem posměchu evropských přistěhovalců ; některé z voda kámen hladí Záměstí které a poraněné smysly jak smečku duši , a tak se impresionisté , anebo kdy Camille Nizozemci jsou většinou - i Nizozemci tento " obraz v zástupci rodu Homo První zástupci předkové byli také jasně zbarvení impresionisté , anebo kdy Camille proroci a jsou to knihy křesťané nazývali často Ježíše , vůdcové církve . x Kniha projížděl mezi obcí Velešín a primáti . Srovnatelnou velikost má expresionisté stejně jako futuristé toužili</p>
--	---	---

SSJČ

ranní *příd. vztahující se k ránu; jitřní*: v r-ch hodinách; r. vysílání; r. noviny; r. mrazíky; r. vlak; r. směna *na kt. se nastupuje ráno*;

♦ r. ptáče dál doskáče (přísloví); — zpodst. **ranní**, -í ž.

1. **ranní mše**: nechodí na hrubou, ale na r. (Jir.)

2. ob. **ranní směna**: mít r. (J. Mar.);

→ přisl. ***ranně**: r. svěží zahrada (Maj.)

SSJČ

ranný *příd. zast. k rána; týkající se rány*: r-á zimnice (Jir.); med. r. šok; řidč. r-á tkáň **kalus**;

— rozl. od raný

Chyba: **klaný**, ale i **kalý**.

Y,Romano hangos, č. 12/2005
Y,Romano hangos, č. 12/2005
Y,Romano hangos, č. 12/2005

nabízejí postřibřené řetízky s medailonkem legendě . Podle ní byla jezdcí na koních - budou

Kalí Sáry . A když zaplatíte
Kalí Sára otrokyní a sloužila u
Kalí Sárů provázet až do moře

Argumenty proti uvádění adjektiv typu **kalý** ve slovníku.

Filtr	word	Frekven
1.	p/ n kalých	15
2.	p/ n Kalá	14
3.	p/ n Kalou	10
4.	p/ n kalé	10
5.	p/ n kalou	10
6.	p/ n kalého	10
7.	p/ n kalým	5
8.	p/ n Kalé	5
9.	p/ n kalá	4
10.	p/ n Kalí	4
11.	p/ n Kalym	3
12.	p/ n kalýho	3
13.	p/ n kalými	3
14.	p/ n kalym	2
15.	p/ n kalejší	2
16.	p/ n kalý	2
17.	p/ n Kalej	2
18.	p/ n kalému	1
19.	p/ n KALYM	1
20.	p/ n kalí	1
21.	p/ n nejnekalejší	1
22.	p/ n Kalý	1

Celkem: 202 (5 str.)

	Filtr	word	Frekvence
1.	p/ n	nekalé	3 977
2.	p/ n	nekalých	1 819
3.	p/ n	nekalou	1 644
4.	p/ n	nekalého	1 557

16. 3. Dů: Analýza zadaných dvojic homonymních tvarů z hlediska disambiguace.