

**Varianty a dublety – korpusy jako zdroje
dat pro formulaci pravidel distribuce
variantních koncovek v české slovesné
flexi**

PLIN032

Středa: 12.00-13.40 G13

osolsobe@phil.muni.cz

O čem budeme mluvit

- Varianty a dublety – infinitiv, l-ové příčestí a substantivum slovesné
- Kolik má čeština infinitivních koncovek?
- Mohou mít tvary l-ového příčestí varianty?
- Mohou mít tvary pasivních příčestí a dějových jmen varianty?
- Jak lze varianty hledat v korpusu?

Infinitiv a variantnost

- -t (nést, brát, mazat, třít, krýt, kupovat, prosit, trpět, sázet, dělat **ALE** i péct, moct, říct)
- -ti (nésti, bráti, mazati, tříti, krýti, kupovati, prositi, trpěti, sázeti, dělati)
- c-i (péci, moci, říci)

Vyhledávání dvojic lišících se zakončením t/ti

Jazyk: čeština

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

vzor 1: .+ t vlastní tag > Vf.*

vzor 2: .+ ti vlastní tag > Vf.*|X.*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 0 Hledat: tvary Vyhodnotit: tvary

A = a

▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/xSaQx9MA> Nápověda

Dvojice podle frekvencí tvarů na ti

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	být (133765)	býti (699)
2	mít (52129)	míti (187)
3	stát (15125)	státi (92)
4	dát (13136)	dáti (90)
5	vidět (21638)	viděti (56)
6	žít (9483)	žíti (53)
7	mluvit (12458)	mluviti (50)
8	jít (16970)	jíti (46)
9	pít (2156)	píti (42)
10	učinit (1644)	učiniti (39)
11	vést (5043)	vésti (35)
12	vzít (7648)	vzíti (35)
13	hrát (13199)	hráti (30)
14	znát (4718)	znáti (29)
15	poznat (3704)	poznati (28)
16	psát (5474)	psáti (28)
17	hledat (6812)	hledati (26)
18	slyšet (7557)	slyšeti (26)
19	pracovat (9681)	pracovati (25)
20	číst (4573)	čísti (23)

Vyhledávání dvojic lišících se zakončením ct/ci

Jazyk: čeština

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

vzor 1: .+ ct vlastní tag > Vf.*|X.*

vzor 2: .+ ci vlastní tag > Vf.*|X.*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 0 Hledat: tvary Vyhodnotit: tvary

A = a ▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/BMUV94Xx> Nápověda

Dvojice podle frekvencí tvarů na ci

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	říct (22553)	říci (8245)
2	pomocť (3595)	pomoci (6520)
3	mocť (2009)	moci (5634)
4	napomocť (26)	napomoci (273)
5	utéct (1622)	utéci (230)
6	přemocť (63)	přemoci (183)
7	obléct (293)	obléci (152)
8	zříct (41)	zříci (107)
9	vypomocť (30)	vypomoci (85)
10	téct (245)	téci (76)
11	vymocť (7)	vymoci (72)
12	domocť (6)	domoci (67)
13	dopomocť (10)	dopomoci (57)
14	péct (352)	péci (47)
15	převléct (93)	převléci (43)
16	tloucť (132)	tlouci (37)
17	ict (4)	ici (25)
18	act (13)	aci (24)
19	svléct (29)	svléci (24)
20	neříct (83)	neříci (23)

IJP <https://prirucka.ujc.cas.cz/?id=510>

Slovesa vzoru „péci“ (infinitiv *-ci* – *-ct*)

[1 Infinitiv](#)

[2 První osoba jednotného čísla a třetí osoba množného čísla](#)

[2.1 Další tvary](#)

[3 Poznámky k některým slovesům](#)

1 Infinitiv

K typu „péci/péct“ náleží sedm sloves, tj. *péci/péct*, *síci/síct*, *téci/téct*, *tlouci/tlouct*, *vléci/vléct*, *moci/moct*, kniž. *stříci se*, a jejich předponové odvozeniny, např. *pomoci/pomoct*, *obléci/obléct*. (Některé předponové odvozeniny mají vedle tvarů podle vzoru „péci“ též tvary podle vzoru „tisknout“, viz [bod 3](#).)

S tvary infinitivu na *-ci* se setkáváme v dříve vydaných slovnících, ve starší literatuře nebo ve vyšším stylu. Infinitiv zakončený na *-ct* je novější, vznikl analogicky podle infinitivů ostatních sloves (*péct* jako např. *nést*), nabyl postupně ve spisovném jazyce platnosti neutrální a tvar na *-ci* je jím vytlačován do vrstvy prostředků knižních. U slovesa *moct/moci* je ještě stále častější variantou infinitivu podoba na *-ci*.

Může existovat hybridní tripleta?

- -t/-ti/-ct
- -ct/-ci/cti
- Pokuste se takové případy najít
- Pomocí nástroje *Morfio*
- Pomocí vyhledávače *KonText*
- Pozorujte výsledky a charakterizujte je

I-ové příčestí a variantnost

- Existují varianty?
- Pokud ano, lze definovat slovesa, která varianty mít mohou?
- Jak varianty vyhledat v korpusu?

Jak zadat dotaz na dvojice typu řekl/řek?

Morfio

Jazyk: čeština ▾

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

vzor 1: | ▾

vzor 2: ▾

Přidat vzor

Korpus: ▾ Frekvence vyšší než: Hledat: ▾ Vyhodnotit: ▾

A = a Alternace

Odkaz na toto zadání: <http://morfio.korpus.cz/6VZwbKje>

Výsledky seřazené podle frekvence 2. členu dvojice

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	předl (38)	před (141338)
2	početl (2)	počet (16979)
3	trhl (553)	trh (8151)
4	bodl (193)	bod (5487)
5	dechl (3)	dech (5275)
6	vznikl (4246)	vznik (4109)
7	podnikl (365)	podnik (3580)
8	dobl (2)	dob (3366)
9	krkl (11)	krk (3171)
10	poklesl (348)	pokles (2926)
11	řekl (70432)	řek (2775)
12	tipl (10)	tip (2496)
13	dopadl (1184)	dopad (2398)
14	jihl (1)	jih (2153)
15	bleskl (37)	blesk (2077)
16	vztekl (2)	vztek (2019)
17	tiskl (518)	tisk (1851)
18	stopl (11)	stop (1781)
19	pásl (92)	pás (1734)
20	hubl (8)	hub (1716)
21	křikl (500)	křik (1654)
22	odpadl (121)	odpad (1472)
23	odhadl (441)	odhad (1451)
24	chladi (19)	chladi (1354)
25	sloupl (12)	sloup (1165)

Všimněme si

- přegenerování – dějová jména homonymní s potenciálními tvary apokopovaných příčestí
- náhodná homonyma

vzor 1:	<input type="text" value=".+?[:Kons:]"/>	<input type="text" value="l"/>	vlastní tag >	<input type="text" value="Vp.*"/>
vzor 2:	<input type="text" value=".+?[:Kons:]"/>	<input type="text"/>	vlastní tag >	<input type="text" value="Vp.*"/>

[Přidat vzor](#)

Problém s nedostatkem AM

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	řekl (70432)	řek (1053)
2	mohl (43171)	moh (733)
3	nemohl (13045)	nemoh (349)
4	přinesl (3905)	přines (335)
5	neřekl (3716)	neřek (187)
6	vylezl (820)	vylez (106)
7	utekl (1626)	utek (100)
8	donesl (425)	dones (86)
9	četl (3489)	čet (68)
10	vytáhl (4774)	vytáh (65)
11	odnesl (1542)	odnes (60)

Příklady konkordancí pro "[word="(?)řek" & tag=".*"]" z korpusu syn2015

McInerney, Jay, Zářivá světla velkoměsta
Herling-Grudziński, Gustaw, Deník psaný v noci
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Velinský, Jaroslav, Hranice pomsty
Englund, Peter, Nepochopitá léta

Odysseus . Jasně , že **Řek** . " Kdyby aspoň Tad
" Poustevníkem byl Hélos , **Řek** z Egypta . Než požádal
Ty bys to taky hned **řek** na plnou hubu . .
Co kdybys mi teď třeba **řek** , jaký tě potkávají podivnosti
nevěděl , co mluví . **Řek** mi , abych se ti
. " " Tos jí **řek** ? " na to Eda
Kdyby mi býval tuhle Jindra **řek** , o co mu jde
nezadala . A co jsi **řek** ty ? " " Něco
" Ale ty jsi přece **řek** to samý . " "
? " " Jistě . **Řek** jsi , že je zboucháš
a jeden z nich mi **řek** , že by byla škoda
A ty jsi mu to **řek** , ty srabe ? "
a ty jsi na to **řek** : Leda kdyby mluvila jako
jsem Edu . . . **Řek** mi včerejší datum , a
o Vícku . Ouládnout úřad **řek** na německém pobřeží Baltu by

Variantní tvary pasivních příčestí

- n-ové a t-ové příčestí
- Jaká je distribuce variantních tvarových koncovek –n-/-t-?
- Existují slovesa s oběma variantami?
- Jak využít korpusové nástroje k jejich vyhledávání?

Varianty dějových jmen

- Dějová jména na *ní/tí*
- Jak lze hledat varianty?

n/t



Jazyk: čeština

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

vzor 1: .+ n vlastní tag > Vs.*

vzor 2: .+ t vlastní tag > Vs.*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 0 Hledat: tvary Vyhodnotit: tvary

A = a ▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/VXX1r4dg> Nápověda

Vyzkoušejte i pro další korpusy, s nimiž pracuje nástroj *Morfio*

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	nasán (1)	nasát (1)
2	nesen (50)	neset (2)
3	vysán (2)	vysát (1)

ní/tí

Jazyk: čeština ▾

<+ ✖ společný ▾ ✖ odlišný ▾ +> Morf. specifikace:

vzor 1: .+ ▾ ní ▾ vlastní tag > ▾ NNN.*

vzor 2: .+ ▾ tí ▾ vlastní tag > ▾ NNN.*

Přidat vzor

Korpus: SYN2015 ▾ Frekvence vyšší než: 0 Hledat: lemmata ▾ Vyhodnotit: lemmata ▾

A = a

▶ Alternace

Hledat **Nové zadání** Odkaz na toto zadání: <http://morfio.korpus.cz/ChNFlaxz> Nápověda

Všimněte si přegenerování

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	dojení (184)	dojetí (600) ↓
2	klení (98)	kletí (1)
3	nasání (2)	nasátí (23)
4	odsání (4)	odsátí (30)
5	osení (30)	osetí (16)
6	pojení (1)	pojetí (4531) ↓
7	přísání (1)	přísátí (22)
8	roztání (3)	roztátí (12)
9	snění (667)	snětí (5) ↓
10	stolení (5)	století (31278) ↓
11	vání (32)	vátí (1)
12	vysání (1)	vysátí (19)
13	zaklení (46)	zakletí (42)

Zachycení variantnosti a automatická morfologická analýza

- Požadavek konzistentního zachycení variant
- Zlaté pravidlo morfologie a kodifikační přístup
- Historický aspekt vzniku variant
- Varianty a dialekty
- Varianty a další korpusové nástroje

Výzkum variant

SyD 

Diachronní

Synchronní

Nový dotaz

[1] péci

[2] péct

[3] píct

Porovnat varianty

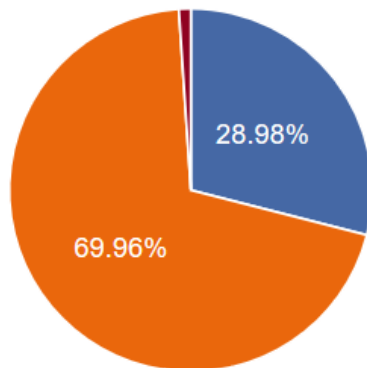
a=A

lemma

Zobrazit adresu dotazu

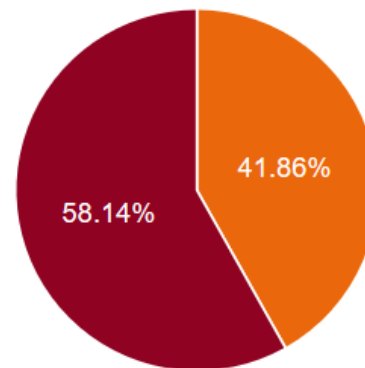
 Sdílet

Psaný jazyk



 [1] péci
 [2] péct
 [3] píct

Mluvený jazyk



 [1] péci
 [2] péct
 [3] píct

Domácí úklona 26. 9. 2022

- Varianty slovesných tvarů od kmene přítentního
- Které tvary od kmene přítentního mohou mít v češtině varianty?
- Jak je vyhledáme v korpusech?
- Jak je hodnotí kodifikace?
- Jak se variantnost odrazí ve slovotvorné variantnosti?

Ke čtení

- OSOLSOBĚ, Klára a Hana ŽIŽKOVÁ. Automatická morfologická analýza z hlediska pokrytí a nepokrytí morfologických variant. In *Korpusová lingvistika Praha 2016, Od mluvené češtiny k psané, 17. září 2016, Praha. 2016.*
- OSOLSOBĚ, Klára. Synonymní/variantní tvary infinitivu jako problém konzistentní lemmatizace. In Jana Bílková, Ivana Kolářová, Miloslav Vondráček. *Lingvistika - korpus – empirie*. 1. vyd. Praha: Ústav pro jazyk český AV ČR, v. v. i., 2021. s. 28-42. *Lingvistika - korpus – empirie*. ISBN 978-80-88211-13-6.