

PLIN033

***Deriv* – nástroj pro testování
derivačních vztahů ve strojovém
slovníku a v korpusech**

Co dnes chceme?

- představit některé funkce *Deriv*
- ukázat na konkrétním příkladu postup práce při extrakci podkladů pro lingvistickou analýzu slovotvorných formací automaticky získaných prostřednictvím *Derivu*
- zadat úkoly na příště

Co je to *Deriv*?

- webové rozhraní
- schopnost pracovat s morfologickým slovníkem analyzátoru *(m)ajka*
- propojení s tištěnými slovníky
- propojení s korpusy

Oproti ajce jsou zde feminina rozdělena na životná a neživotná, chvíli jen přibližně, většina slov je správně, ale lze najít i mn

„Jazyk“ regulárních výrazů je v základních vlastnostech st

ajka – tagset

- <http://nlp.fi.muni.cz/projekty/ajka/tags.pdf>

ajka tagset

Radek Sedláček

March 1, 2006

1 k1 - Substantivum


x	Speciální vzor
P	půl
g	Rod
M	Rod mužský životný
I	Rod mužský neživotný
N	Rod střední
F	Rod ženský
R	Rodina (příjmení)

debdict

chrome://debdict - DEB Prohlížeč slovníků - Mozilla Firefox

všechny slovníky

Hledat hledat i v definicích



DEBII — Dictionary Editor and
DEBDict — novinky

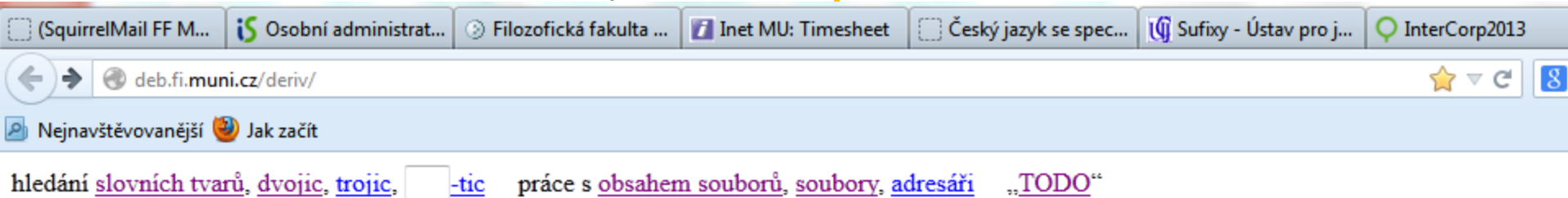
[úvod](#)
klienti
- [DEBDict](#)

Aktuální verze: 1.7.31 **Stav**

Postup pro instalaci nové verz

deb.fi.muni.cz/deriv

- HESLO: smerk@mail.muni.cz.
- Jméno: **PLIN033**, heslo: **plin**



Hledání slovních tvarů

- Lze hledat dle značky (brněnský systém - <http://nlp.fi.muni.cz/projekty/ajka/tags.pdf>)

hledání [slovních tvarů](#), [dvojic](#), [trojic](#), -[tic](#) práce s [obsahem souborů](#), [soubory](#), [adresáři](#)

Hledání

značka RE

- hledat i mezi hovorovými tvary (přesněji: tvary se značkou obsahující atribut w, v naprosté většině ale j)
- hledat i mezi vlastními jmény (přesněji: tvary obsahujícími i jiný znak, než jen malé písmeno, v naprosté většině ale j)

uložit do souboru

Co chceme?

- Substantiva typu *náměstí*
- Jaké další známe?
- Jak je můžeme popsat?
- Jaké mají formální vlastnosti?

ná- -í

- Substantiva
- Neutra
- Lemma začíná na *ná* a končí na *í*

Vyplnění formuláře

Hledání

značka RE

- hledat i mezi hovorovými tvary (přesněji: tvary se značkou obsahující atribut v 2000 tvarů má jinou hodnotu)
- hledat i mezi vlastními jmény (přesněji: tvary obsahujícími i jiný znak, než jen velkým písmenem na začátku, byť v datech je třeba i *AIDS*, *PhDr.*, *SMSka*, *tie-breacích* atp)

uložit do souboru

lemmat přímo slovní tvary

Vyhledat a otevřít

- Seznam slov, která odpovídají zadání a která jsou uvedena ve slovníku morfologického analyzátoru *ajka*

Práce s obsahem souboru PLIN033/ná_í

Soubor **PLIN033/ná_í** byl úspěšně uložen (0 s).

Poznámkou je jednotlivý znak, jeden řádek může mít více poznámek.

Jako poznámku nelze použít dvojtečku, čárku a mezeru (budou-li zadány, budou ignorovány).

- | | | |
|----|--------------------------|----------------------------------|
| 1 | <input type="checkbox"/> | náboženství |
| 2 | <input type="checkbox"/> | nábožnůstkářství |
| 3 | <input type="checkbox"/> | nábřeží |
| 4 | <input type="checkbox"/> | nábytkářství |
| 5 | <input type="checkbox"/> | náčelnictví |
| 6 | <input type="checkbox"/> | náčiní |
| 7 | <input type="checkbox"/> | nádbí |
| 8 | <input type="checkbox"/> | nádenictví |
| 9 | <input type="checkbox"/> | nádeničení |
| 10 | <input type="checkbox"/> | nádní |

Co lze dále zjistit?

- Otevřít soubor
- Upravovat jeho obsah
- Prohlížet frekvence
- Nahlížet do slovníků
- Nahlížet do korpusů

Otevření s frekvencemi

neprázdné soubory v adresáři **Osolsobe/PLIN033** (přepnout do adresáře
)

u každého souboru je arita, počet řádků a případně seznam použitých poznámek (prázdná poznámka je znázor

• ná_í (1, 84)

obsah (jen pro n-tice) nebo otevřít pro editaci po

zobrazit seřazené retrográdně vybrat pouze slova neoznačená jako lze zadat

(jen pro soubory s aritou 2 nebo 3)

slovo (frekvence)

Práce s obsahem souboru PLIN033/ná_í

Poznámkou je jednotlivý znak, jeden řádek může mít více poznámek.
Jako poznámku nelze použít dvojtečku, čárku a mezeru (budou-li zadány, budou ignorovány).

První číslo v závorkách je počet výsledků korpusového dotazu [word="slovo"] v korpusu [SYN2000](#), druhé číslo [lemma="slovo"] (případně seznam takových výsledků pro různé slovní druhy, viz např. slovo dolomit) v tom korpusu, třetí číslo je počet výsledků dotazu [word="slovo"] v korpusu [CzTenTen](#) a čtvrté číslo je počet výsledků dotazu [lemma="slovo"] v korpusu [CzTenTen](#). Zobrazená čísla jsou předpočítaná, takže se mohou lišit od počtů udávaných korpusovým managerem (zpravidla v souboru [SYN2000](#) proto, že nemám k dispozici zdrojovou podobu verze, kterou odkazují, u CzTenTen proto, že korpusy se mohou měnit). Druhé a čtvrté číslo nemusejí mít dobrý smysl v případech, že v souboru nejsou lemmata, ale tvary, ovšem poskytuje neuchovávané informace, jakým způsobem soubor vznikl, takže nelze spolehlivě poznat, jestli jde o lemmata, nebo tvary. Zároveň fungují jako odkazy na konkordanční seznamy příslušných výskytů. Je vhodné mít na paměti, že zatímco první číslo odpovídá přímo korpusovým datům, druhé a čtvrté pracuje s výsledky automatické lemmatizace, která ovšem může obsahovat chyby.

- 1 [náboženství](#) ([2533](#), [3070](#), [154411](#), [186107](#))
- 2 [nábožnůstkářství](#) ([0](#), [0](#), [0](#), [0](#))
- 3 [nábřeží](#) ([988](#), [1085](#), [25625](#), [28834](#))

Kliknutím na slovo získáme informace z tištěných slovníků uložených v databázi debdict

SSJC Slovník spisovného jazyka českého

nábřeží

-í s. *upravená a podezděná městská cesta podle řeky, jezera:* vltavské n.; procházet se p. kapitána Jaroše; --> expr. zdrob. **nábřežíčko**, -a s. (6. mn. -ách)

SSC Slovník spisovné češtiny

nábřeží

-í s. *stavebně upravený prostor podél řeky ap.* vltavské nábreží, chodit po nábreží, **nábřežní** příd.

Kliknutím na slovo získáme informace o výskytu v korpusech

- konkordance z korpusu (SYN2000, czeTenTen)

smilide	na vlhký obzor . Po nábřeží hopkaly veverky , které si
smilide	trocho anarchisticky vyhlížející mladík na nábřeží v lehké teplákové soupravě a
smilide	firmy Kodak . Jinak bylo nábřeží prázdné a mladíkovu opuštěnost umocňoval
smilide	. Nohy ho donesly na nábřeží a odtud do hospody na
smilide	, kam se dostal od nábřeží . Stále ho nikdo nesledoval
smilide	na chvíli se z okraje nábřeží zahleděl na doutnající tok řeky
smilide	. Šli spolu kousek po nábřeží , téměř až ke kříži
pi210	aby se postavil dům na nábřeží Vltavy , Hlahol . .
havel	2000 v Praze na původním nábřeží Palackého , později Rašínově ,
havel	Do nově zařízeného bytu na nábřeží č . 2000 přinesli z
havel	kolegiem . Doma , na nábřeží , mu ale chybějí .
havel	a potom zařízení bytu na nábřeží . Líbí se mi ,
ing	Dame zní zvon a po nábřeží promenují podezřelé krásky . Porci
brizy	ve čtvrtém poschodí , na nábřeží Marxe a Engelse . Mohl
svedeny	jakém smyslu ? Šli po nábřeží . Zastavil se a oběma
s_mloky	, kteří se trousí po nábřeží a po mostě , nemají
s_mloky	mostě cinkala tramvaj , po nábřeží putovaly chůvy s kočárky a
krakatit	Prokop si razí cestu po nábřeží . Mrazí ho a čelo
honzlova	jsem se dál směrem k nábřeží . Sfingy , co hlídají
honzlova	na ní rozhojnily , ale nábřeží bylo stejně mrtvé jako ráno

strana ze 52 [další](#) | [poslední](#)

Jak dále ?

- Méně obvyklá a méně frekventovaná slova

16	<input type="checkbox"/>	náhlaví (0 , 0 , 1 , 2)
17	<input type="checkbox"/>	náhlení (0 , 0 , 2 , 2)
18	<input type="checkbox"/>	náhradbí (0 , 0 , 3 , 4)
19	<input type="checkbox"/>	náhradí (0 , 0 , 151 , 208)

náhlaví

SSjC Slovník spisovného jazyka českého

náhlaví

-í s. *zeměd.* část předku pluhu spočívající mezi plužňaty; náhlavník

SSC Slovník spisovné češtiny

psjc Příruční slovník jazyka českého

náhlaví, -í n. *hosp.* část předku pluhu spočívající na drábcích mezi plužňaty, náhlavník

Výskyt v korpusu

- Odpovídá význam výskytů nalezených v korpusu významům, které uvádějí tištěné slovníky ?

Dotaz **náhlaví** 2 (0.0 v miliónu)

[doc#3306246](#) několik problémů s kompatibilitou s Bluetooth **náhlavími** . `</p><p>` Provádíme značné změny v knihovně
[doc#6747711](#) Setřídění seznamu se provede kliknutím na **náhlaví** sloupce, dle kterého je třídění požadováno

Pohled do zdrojů

- Kliknutím na url lze získat u velkých korpusů z webu přístup ke zdrojovým textům

⌵	
doc.url	http://www.suseportal.cz/kategorie/novinka/novinky-okolo-skype-pro-linux
doc.wordcount	559
p.heading	0
p.accent	yes

Celý text

- Co je to náhlaví?

Zde jsou nějaké věci, kterými jsme se mezitím zabývali:

- Kompletně jsme přepsali knihovnu pro práci s audiem a věnovali značné úsilí přizpůsobování stávajícího kódu pro práci se zvukem do podoby lépe vyhovující pro Skype.
Také jsme opravili několik problémů s kompatibilitou s Bluetooth **náhlaví**mi.
- Provádíme značné změny v knihovně pro práci s videem. Pracujeme na tom, aby byla stabilnější a podporovala více webkamer a X video módů.
- Chystáme se adoptovat některá vylepšení uživatelského rozhraní ze Skype 4.0 pro Windows, například záložku s konverzacemi. Také rozmýšlíme nad přidáním jednookenního režimu.
Samozřejmě, toto je stále ve vývoji, ale náš cíl je vytvořit jednodušší uživatelské rozhraní vyžadující méně kliknutí myši pro provedení nejběžnějších operací.

Lingvistická analýza automaticky vyhledaných dat

- Které jednotky nepatří do seznamu substantiv, neuter tvořených cirkumfixem *ná-* *-í*?
- Projděme seznam (84 jednotek).
- Využijme funkce nástroje *Deriv* označovat nalezené jednotky (okénko mezi pořadovým číslem a slovem).

Ruční práce

- Co je to *nádbí* a *nádní*?

1	<input checked="" type="checkbox"/>	náboženství (2533 , 3070 , 154411 , 186107)
2	<input checked="" type="checkbox"/>	nábožnůstkářství (0 , 0 , 0 , 0)
3	<input type="checkbox"/>	nábřeží (988 , 1085 , 25625 , 28834)
4	<input checked="" type="checkbox"/>	nábytkářství (19 , 20 , 551 , 600)
5	<input checked="" type="checkbox"/>	náčelnictví (4 , 5 , 124 , 151)
6	<input type="checkbox"/>	náčiní (323 , 453 , 14873 , 17969)
7	<input type="checkbox"/>	nádbí (0 , 0 , 7 , 7)
8	<input checked="" type="checkbox"/>	nádenictví (0 , 0 , 12 , 18)
9	<input type="checkbox"/>	nádeničení (0 , 0 , 12 , 6)
0	<input type="checkbox"/>	nádní (0 , 0 , 30 , 42)
1	<input type="checkbox"/>	nádobí (1352 , 1492 , 69703 , 76543)

nádbí

- Odpovídá výklad ve slovníku významu korpusových vyhledávek?

psjc Příruční slovník jazyka českého

nádbí, -í *n. dial.* [naděje](#) [nádba](#) Tož honem nevěstu. Byla v nádbí, Bortková „pantlova vyhlédnuta. A. Mrš.

Dotaz **nádbí** 7 (0.0 v miliónu)

doc#29780	(Oh My Fucking God - jak se jednorázové nádbí tohoto typu vůbec do takové kuchyně dostane
doc#423144	mraznička, mikrovlnka, rychlovarná konvice, nádbí). V apartmá je připojení na internet, televize
doc#1497507	vysoké úrovni. Ať už je to keramika, dřevěné nádbí , košíky, hračky nebo dokonce i bylinky,
doc#1517984	když si spočítáte, jak dlouho vám takové nádbí vydrží, tak vám jistě dojde, co všechno
doc#1699136	dobře čistí nedovedeme (w nadby, snad = v nádbí). Pokus vyložití je by našel čtenář ve
doc#2060149	odmastí i organismus, protože ho tak úplně z nádbí nesmyješ, takže s dalším jídlem konzumuješ
doc#7072646	přepracovat dané receptury a mít dostatečně velké nádbí . <i></p><p></i> Obyvatelé a návštěvníci Domažlic

nádní

nádní I

příd. loď. n. prostor *jsoucí přímo nade dnem lodi*

nádní II

v. návni

SSC Slovník spisovné češtiny

psjc Příruční slovník jazyka českého

nádní adj. 1. *loď.* nádní prostor *který je přímo nade dnem lodi.*

nádní adj. 2. v. [návni](#)

nádní, -í n. *spodek lodi.* Tesal nádní loďky nové. **Hol.** Loď kde nádním smýká mžiká. F. S. Proch.

nádní

- Odpovídá značkování?

Dotaz **nádní** 37 (0.0 v miliónu)

Strana ze 2 | [Poslední](#)

- [doc#14522](#) považuje sběrný prostor nade dnem strojovny (**nádní**). [7.09.2](#) Ke skladování použitého
- [doc#241654](#) zadržel na přídi, automatické čerpadlo **nádní** vody a koupací žebřík z nerezových trubek
- [doc#363919](#) patří automatická elektrická a ruční pumpa **nádní** vody, hasicí přístroj, poziční světla pro
- [doc#453386](#) Aby se loď nepotopila, byly odčerpávány **nádní** vody (asi 25 m³) s obsahem ropných látek
- [doc#584007](#) tvoří předové a zádové automatické čerpadlo **nádní** vody s kontrolkami chodu, čidla oxidu uhelnatého
- [doc#753811](#) drenážních zařízení zaolejovaných vod z **nádní** strojoven vnitrozemských plavidel. [7.09.2](#)
- [doc#753811](#) drenážního zařízení zaolejovaných vod z **nádní** strojoven týká českých plavidel plavících
- [doc#1252259](#) Kruhu přátel čes. jazyka), nadní (Morava) a **nádní** (dokládá PS. z Baara; je u Milevska [7.09.2](#)).
- [doc#2090565](#) sklolaminátu. [7.09.2](#) Příprava na pumpu na **nádní** vodu - k mému překvapení a potěšení jí
- [doc#2186467](#) 12 V, světlometem, elektrickým čerpadlem **nádní** vody, hasicím přístrojem a palivovým filtrem
- [doc#2208716](#) je 25l bojler. K eventuálnímu odčerpání **nádní** vody slouží stoková čerpadla, jedno elektrické
- [doc#2264427](#) vybaveny měřičem aktuálního stavu. Čerpadla **nádní** vody jsou celkem tři - ruční, elektrické

Chyby ve značkách

- I adjektivní výskyty jsou označovány jako substantiva

Dotaz **nádní** 37 (0.0 v miliónu)

Strana ze 2 | [Poslední](#)

doc#14522	považuje sběrný prostor nade dnem strojovny (nádní /k1gNnSc1/nádní).	nádní /k1gNnSc1/nádní	.
doc#241654	zabradlí na přídi, automatické čerpadlo	nádní /k1gNnSc2/nádní	vody a koupací žebřík z nerezových trubek
doc#363919	patří automatická elektrická a ruční pumpa	nádní /k1gNnSc2/nádní	vody, hasící přístroj, poziční světla pro
doc#453386	Aby se loď nepotopila, byly odčerpávány	nádní /k1gNnSc4/nádní	vody (asi 25 m 3) s obsahem ropných látek
doc#584007	tvoří před'ové a zád'ové automatické čerpadlo	nádní /k1gNnSc2/nádní	vody s kontrolkami chodu, čidla oxidu uhelnatého
doc#753811	drenážních zařízení zaolejovaných vod z	nádní /k1gNnSc2/nádní	strojoven vnitrozemských plavidel.
doc#753811	drenážního zařízení zaolejovaných vod z	nádní /k1gNnSc2/nádní	strojoven týká českých plavidel plavících
doc#1252259	Kruhu přátel čes. jazyka), nadní (Morava) a	nádní /k1gNnSc4/nádní	(dokládá PS. z Baara; je u Milevska<g />.
doc#2090565	sklolaminátu.	nádní /k1gNnSc6/nádní	vodu - k mému překvapení a potěšení jí
doc#2186467	12 V, světlometem, elektrickým čerpadlem	nádní /k1gNnSc2/nádní	vody, hasicím přístrojem a palivovým filtrem
doc#2208716	je 25l bojler. K eventuálnímu odčerpání	nádní /k1gNnSc2/nádní	vody slouží stoková čerpadla, jedno elektrické

Zvolme značky a vyznačme případy tzv. přegenerování

- XY?

1	X	náboženství (2533 , 3070 , 154411 , 186107)
2	X	nábožnůstkářství (0 , 0 , 0 , 0)
3		nábřeží (988 , 1085 , 25625 , 28834)
4	X	nábytkářství (19 , 20 , 551 , 600)
5	X	náčelnictví (4 , 5 , 124 , 151)
6		náčiní (323 , 453 , 14873 , 17969)
7	?	nádbí (0 , 0 , 7 , 7)
8	X	nádenictví (0 , 0 , 12 , 18)
9	Y	nádeničení (0 , 0 , 12 , 6)
10		nádní (0 , 0 , 30 , 42)
11		nádobí (1352 , 1492 , 69703 , 76543)
12	?	nádrabí (0 , 0 , 0 , 0)
13		nádraží (4591 , 5283 , 208635 , 235714)
14	X	nádvornictví (0 , 0 , 0 , 0)
15		ná dvoří (1061 , 1168 , 50456 , 55587)
16		náhlaví (0 , 0 , 1 , 2)
17	Y	náhlení (0 , 0 , 2 , 2)

Přegenerování a podgenerování jako obecný problém automatické analýzy

- Podmínka v zadání je nutná, nikoli dostačující.
- Důsledek – přegenerování – výsledek obsahuje data, která jsme vyhledat nezamýšleli.
- Důsledek – podgenerování – výsledek neobsahuje data, která jsme vyhledat zamýšleli.

Přegenerované výsledky

- Substantiva (kolektiva) na [sc]tví – X.
- Dějová jména od sloves začínajících na *ná-* – Y.
- Další (kompozitum *názvosloví*) – Z.
- Podezřelé (chyby ve značkování) – !
- Podezřelé (překlepy) – ?

Uložení do nového souboru

- **Název nového souboru se liší od starého**

76	<input checked="" type="checkbox"/>	nástupnictví (174 , 194 , 3171 , 3381)
77	<input type="checkbox"/>	návěstí (41 , 43 , 1198 , 454)
78	<input type="checkbox"/>	návěští (22 , 32 , 1188 , 1555)
79	<input type="checkbox"/>	návětrí (1 , 1 , 167 , 185)
80	<input type="checkbox"/>	návidlí (0 , 0 , 0 , 1)
81	<input checked="" type="checkbox"/>	návladnictví (0 , 0 , 28 , 30)
82	<input checked="" type="checkbox"/>	návrhářství (40 , 45 , 1911 , 2062)
83	<input type="checkbox"/>	návrší (220 , 253 , 9651 , 5533)
84	<input checked="" type="checkbox"/>	názvosloví (156 , 177 , 8957 , 10216)

do souboru

Označovaný soubor

- Funkce vybrat pouze slova označená jako

neprázdné soubory v adresáři **Osolsobe/PLIN033** (přepnout do adresáře

)

u každého souboru je arita, počet řádků a případně seznam použitých jednoznakových poznámek (prázdná poznámka)

- ná_í (1, 84)
- ná_í_KO (1, 84, :XYZ)
- ná_í_KO1 (1, 84, :!XYZ)
- ná_í_KO1_?! (1, 36, :!?)
- ná_í_KO1_x?! (1, 39, :!x)
- ná_í_KO_bezX (1, 55, :YZ)
- ná_í_KO_bezXY (1, 41, :Z)
- ná_í_KO_bezXYZ (1, 40)

obsah (jen pro n-tice) nebo otevřít pro editaci poznámek
zobrazit seřazené retrográdně vybrat pouze slova neoznačená jako (více poznámek)

Vybrat slova bez poznámek

- Je možné ručně upravit – odstranit přegenerovaná data.

• ná_í (1, 84)

• ná_í_KO (1, 84, :XYZ)

obsah (jen pro n-tice) nebo otevřít pro editaci po
zobrazit seřazené retrográdně vybrat pouze slova neoznačená jako lze zadat

Postup odstranění/výběru ručně označovaných jednotek

- Chceme-li zvolit více poznámek (ručně zvolených značek), pak je třeba zadat volbu pod menu „více poznámek spojit spojkou“.

O1 (1, 84, :! ?XYZ)

O1_?! (1, 36, :! ?)

O1_x?! (1, 39, :! ?x)

O_bezX (1, 55, :YZ)

O_bezXY (1, 41, :Z)

O_bezXYZ (1, 40)

obsah (jen pro n-tice) nebo otevřít pro editaci poznámek

řazené retrográdně vybrat pouze slova neoznačená jako <YZ> (více poznámek spojit sp

Soubor bez ručně označených dat

- neodstranili jsme podezřelé jednotky

2 [náčiní](#) ([323](#), [453](#), [14873](#), [17969](#))

3 ? [nádbí](#) ([0](#), [0](#), [7](#), [7](#))

4 ! [nádní](#) ([0](#), [0](#), [30](#), [42](#))

5 [nádobí](#) ([1352](#), [1492](#), [69703](#), [76543](#))

6 [nádrabí](#) ([0](#), [0](#), [0](#), [0](#))

7 [nádraží](#) ([4591](#), [5283](#), [208635](#), [235714](#))

Soubor uložíme

- Nový název souboru je např. ná_í_KO1_?!

35 [návěstí](#) ([41](#), [43](#), [1198](#), [454](#))

36 [návěstí](#) ([22](#), [32](#), [1188](#), [1555](#))

37 [návětrí](#) ([1](#), [1](#), [167](#), [185](#))

38 [návidlí](#) ([0](#), [0](#), [0](#), [1](#))

39 [návrší](#) ([220](#), [253](#), [9651](#), [5533](#))

do souboru

Další zpracování

- můžeme např. odstranit „podezřelé“ jednotky stejně, jak bylo uvedeno výše
- můžeme práci odložit
- můžeme dále kontrolovat, zda jsme někde neudělali chybu
- můžeme se věnovat lingvistickému popisu, např. vytvořit následující slovníkové heslo

slovníkové heslo

- návrh

há- -i

Stavba: ná- -(i)

Cirkumfixem ná- -i se tvoří okrajově **substantiva** od **I. substantiv**. Označují názvy **(1a) míst**, a to jednak z předložkových pádů jmen užitých v odpovědi na otázku *kde?* (v úzu běžná a frekventovaná *náměstí*, *nádraží* i okazionálně tvořená *námoří*), jednak **(1b) názvy míst s působením nějakého povětrnostního vlivu** (*náledí*, *návětrí*), jednak názvy **(1c) dle polohy** (*náruči*, archaický termín *nákončí*). Synchronně většinou neprůhledné jsou názvy **(2) souborů prostředků** (*nádobí*, *náčini*, *nářadí*). Význam **(3) uplatnění motivujícího substantiva** má *násilí*. V době obrozenecké (Dobrovský) bylo přijato utvořené desubstantivum označující **(4) dialekt** (*nářečí*). Právnícký termín je **(5) návěti** = premisa.

Od **II. slovesa** je utvořeno substantivum s významem **(6) prostředku** (*návěsti*, *návěšti*).

nebo se na data podívat z hlediska frekvenční analýzy

- vybrat slova, která uvadějí slovníky, ale v korpusech se nevyskytují (nebo jde o překlepy)
- vybrat slova, která mají velmi nízké frekvence (hapaxy)

Úkol na 2. 10. 2013

- Prostudovat tagset **ajky**
- Zajistit si **přístupová práva ke slovníkům pod debdictem**
- Zajistit si **přístupová práva ke Sketch engine**
- **Podle návodu zpracujte** substantiva typu sou- - í (jako **souhvězdí**).
- **Popište problémy**, na které jste při práci narazili a připravte si **dotazy k technickým problémům**.