**Gramatika a korpus II plin032**

**Středa: 7.30-9.00 G13**

20. 4. Rozbor dú.

Efektivní vyhledávání homonym slovesných tvarů pomocí počítačových nástrojů IV.

**Jak lze vyhledat v korpusu podstatná jména slovesná?**

Podstatné jména slovesná na *ní/tí*

***Deverbativa od slovesného kmene a jejich vyhledávání v korpusech***

**Motivační úvod**

Výzkum slovotvorby se orientuje jak na výzkum fungování slov utvořených, tak na vznik nových pojmenování různými slovotvornými postupy. V korpusech vidíme slova v přirozených kontextech jejich užití, korpusy zahrnují velké množství textů novějších, než jsou ty, ze kterých vyšly dostupné tištěné výkladové slovníky. Je tudíž užitečné vědět, jak lze korpusy použít pro vyhledávání a třídění dokladů pro výzkum slovotvorby. Slova utvořená nějakým slovotvorným prostředkem (prefix, sufix) lze v korpusech poměrně snadno vyhledávat. Už v kapitole věnované adjektivům jsme ukázali, jak lze využít morfologické vlastnosti slovesa při hledání kandidátů na účelová adjektiva. V této kapitole budeme používat analogické postupy.

**Nastínění problému**

O některých deverbativech se v mluvnicích tvrdí, že se tvoří paradigmaticky (lze je utvořit od každého slovesa bez omezení). Platí to o dějových jménech tvořených od tvaru shodného s potenciálním tvarem pasivního příčestí. Zatímco adjektiva z přechodníků mají ve značkování použitém pro anotace korpusů řady SYN zvláštní značky, substantiva na *-ní/-tí* jsou označena jako všechna ostatní substantiva. Tato praxe neodpovídá běžné praxi českých výkladových slovníků. Ty totiž pravidelně tvořená dějová jména nezaznamenávají jako samostatná heslová slova. Deverbální substantiva na *-ní/-tí* si ovšem uchovávají řadu gramatických vlastností fundujícího slovesa (vid, fakultativně zvratnost signalizovanou morfémem *se/si*, do určité míry valenční doplnění). Z tohoto hlediska představují slovotvorný typ, jehož zkoumání může být z mnoha aspektů zajímavé.

**Otázky**

Slovotvorným základem pro odvozování jmen (substantiv a adjektiv) od sloves může být buď kořen (základ) slovesa (např. ***mluv****-i-t>****mluv****-čí*), nebo jeden ze slovesných kmenů (kořen rozšířený o kmenotvornou příponu např. ***uč-i****-t>****uč-i****-tel*), nebo slovesný tvar (např. *kut-i-t>****kuti-i-l/Vp.\*****>****kut-i-l/N.\****). V mluvnicích se v oddílech věnovaných slovotvorbě hovoří o derivaci od kořene a od kmene.

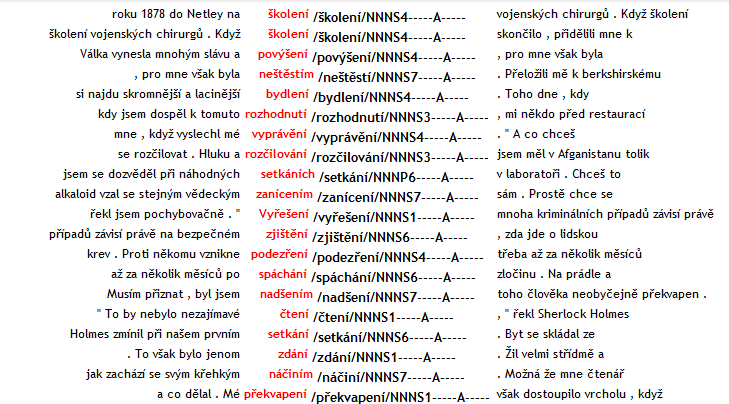
Které typy deverbativ jsou derivovány od slovesného tvaru, které od kmene a které od kořene? Většinou se v mluvnicích uvádí, že odvozovací kmen je kmen minulý. Pod tímto termínem se ale rozumí jistá abstrakce, protože tvary od kmene minulého (někdy též infinitivního) mohou být až tři[[1]](#footnote-1) (infinitiv – *brát*, l-ové příčestí + od něj tvořený přechodník minulý *bral/nabrav*, pasivní příčestí + od něj tvořené verbální adjektivum a substantivum *brán, braný, braní*). Je patrné, že kmenotvorná přípona (v našem případě *a/á*) alternuje, má více variant. Je vždy potřeba zvážit, která z variant je spoluformantem při odvozování od slovesného kmene.

**Formulace dotazu pro získání dat z korpusů**

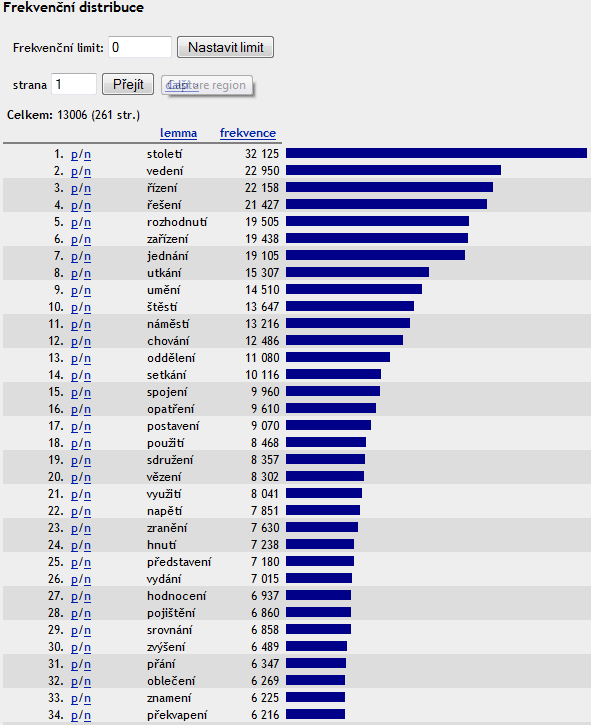
Pokusme se vyhledat v korpusu všechna substantiva slovesná na *-ní/-tí*. Uvedli jsme, že tato substantiva se tvoří od tvaru shodného s potenciálním tvarem příčestí trpného. Mohli bychom tedy vyjít ze znalosti morfematické stavby pasivního příčestí u jednotlivých slovesných tříd a formulovat dotaz pro korpusový manažer tak, abychom zachytili pouze ta substantiva, která by mohla být dějovými jmény. Můžeme ale také zvolit postup kombinovaný. Budeme si všímat korpusových dat a porovnávat výsledky jejich pozorování se znalostmi morfologie slovesného tvaru.

Dějové jméno se musí skloňovat podle vzoru *stavení* a musí končit na řetězec *[nt]í*. V menu **Konkordance** zvolíme **Typ dotazu cql** a do dotazovacího řádku zapíšeme *[lemma=“.\*[nt]í“ & tag=“NNN.\*“]* (hledáme všechna substantivní neutra, jejichž lemma končí na řetězec -*ní/tí*). Podívejme se na frekvenční distribuci lemmat.



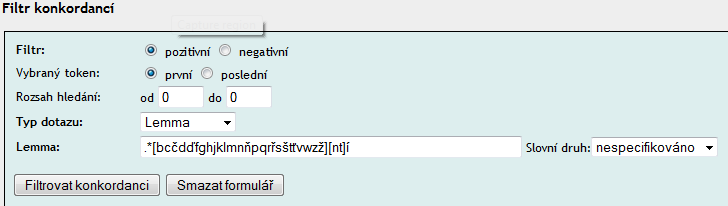


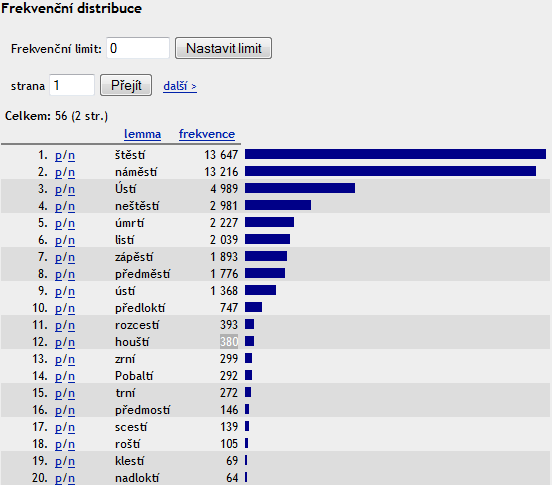
To, že lemma substantiva neutra končí na -*ní/tí*, je podmínka nutná, nikoli postačující pro to, aby slovo bylo dějovým jménem, jak o tom svědčí slova jako *století, štěstí, náměstí, ...* Podíváme-li se na tato tři slova blíže, vidíme, že dvě z nich mají společnou vlastnost, kterou nemá ani jedno ze slov ze zobrazené části seznamu, která za dějová pokládat lze. O jakou vlastnost se jedná? U substantiv *štěstí* a *náměstí* stojí před -*tí* souhláska. Všechny ostatní doklady (*rozhodnutí, užití, napětí, …*) mají před -*tí* samohlásku. Také všechna zobrazená neutra na -*ní*, která lze pokládat za dějová jména, mají před -*ní* samohlásku. Můžeme si položit další otázky. Může před *[nt]í*, má-li jít o dějové jméno, stát souhláska? Může před *[nt]í*, má-li jít o dějové jméno, stát libovolná samohláska? Odpovědi budeme hledat nejdříve v korpusu.



**Třídění a pozorování dat získaných z korpusů**

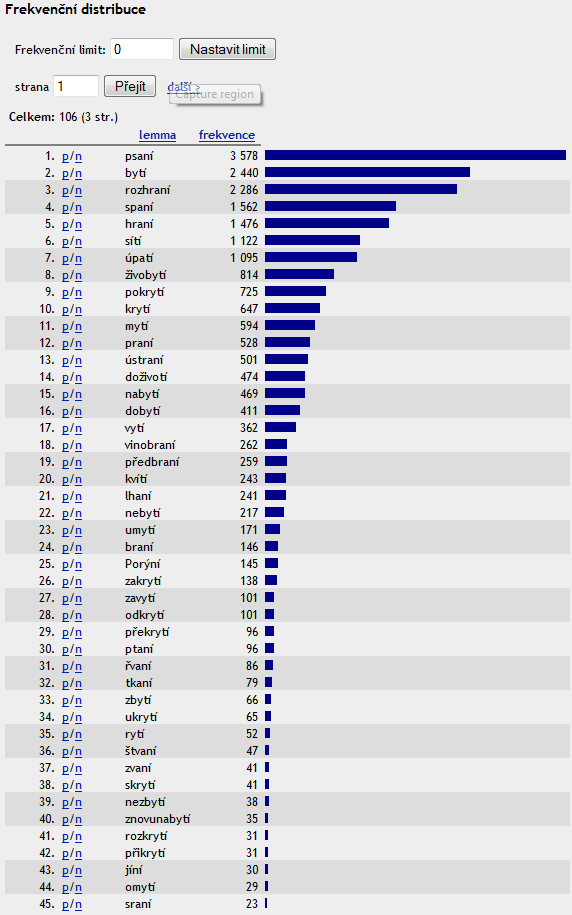
Zvolíme filtr (pozitivní), **Typ dotazu lemma** a do dotazovacího řádku zapíšeme *.\*[bcčdďfghjklmnňpqrřsštťvwxzž][nt]í*. Podívejme se na frekvenční distribuci lemmat.Projdeme-li seznam 56 lemmat, zjistíme, že dějová jména se nevyskytují. To znamená, že ve zvoleném korpusu není doloženo žádné dějové jméno sledovaných vlastností. Vrátíme se o krok zpět, zvolíme filtr (negativní) a do dotazovacího řádku zapíšeme *.\*[bcčdďfghjklmnňpqrřsštťvwxzž][nt]í*.



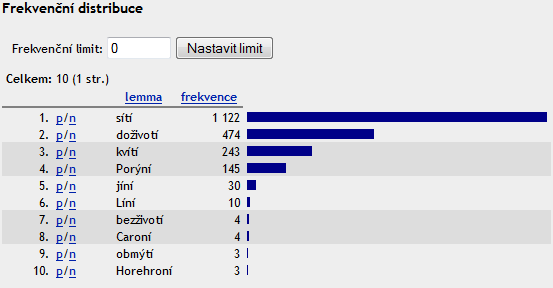


Dále budeme pozorovat, které samohlásky se vyskytují před *[nt]í* u substantiv, která jsou prokazatelně dějovými jmény. Mezi substantivy na první zobrazené stránce frekvenčního seznamu nacházíme taková, že před *[nt]í* předchází *[áeěiu]*. Všimneme si, že kromě *á* jde ve všech ostatních případech o krátký vokál. Nejprve se můžeme zaměřit na ty vokály, které chybějí, tj. *[aéíoóůyý]*.

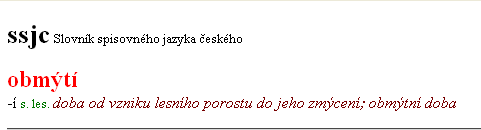
Zvolíme filtr (pozitivní), **Typ dotazu lemma** a do dotazovacího řádku zapíšeme *.\*[aéíoóůyý][nt]í*. Podívejme se na frekvenční distribuci lemmat.Pozorujeme, že předcházejí-li další krátké vokály *[aiy]*, pak jde až na jedinou výjimku (*úpatí*) o dějové jméno (nebo kompozitum, jehož druhým členem je dějové jméno). Naopak předchází-li dlouhý vokál, pak nejde v obou případech o dějové jméno (*sítí, kvítí*). Nemáme k dispozici doklady o jednom krátkém a dalších (kromě *i*) čtyřech dlouhých vokálech *[oéóůý]*.



Vrátíme se zpět. Zvolíme filtr (pozitivní), **Typ dotazu lemma** a do dotazovacího řádku zapíšeme *.\*[oéíóůý][nt]í*. Podívejme se na frekvenční distribuci lemmat.Pozorujeme, že mezi 10 nalezenými lemmaty není ani jedno dějové jméno. Některé hledané kombinace nejsou doloženy vůbec.



Jde v případě lemmatu *obmýtí* o překlep? Ve slovníku slovo *obmýtí* existuje*.*



Vrátíme se o krok zpět, zvolíme filtr (negativní) a do dotazovacího řádku zapíšeme *.\*[oéíóůý][nt]í*.

**Formulace závěrů**

Na základě pozorování dat získaných z korpusů můžeme tvrdit, že ve sledovaném korpusu jsou doložena pouze dějová jména taková, že před *[nt]í* předchází *[aeěiuyá]*. Jsou doložena neutra skloňovaná podle vzoru *stavení*, která nejsou dějovými jmény, u nichž před *[nt]í* předchází buď souhláska, nebo *[oíý].* Nejsou doložena žádná slova taková, u nichž před *[nt]í* předchází *[éóů]*.

Ve sledovaném korpusu jsou doložena lemmata neuter na *[ea][nt]í*, která nejsou dějovými jmény (*století, úpatí, …*).

Závěrem lze říci, že podmínkou nutnou, nikoli postačující k tomu, aby mohlo jít o dějové jméno, je zakončení lemmatu neutra na *[aeěiuyá][nt]í*.

Empirická pozorování korpusových dat můžeme doplnit na základě znalosti gramatiky. Kmenotvorná přípona, která je spoluformantem pro tvoření dějových jmen v češtině, může být pouze *[aá]*, a to u sloves I. třídy vzorů *brát, mazat* (srov. *braní, mazání*), III. třídy vzoru *kupovat* (*kupování*) a V. třídy vzoru *dělat* (*tkaní, dělání*), dále *[e]*, a to u sloves I. třídy vzoru *umřít* (srov. *tření*), *[eě]* II. třídy vzoru *začít* (srov. *početí, stětí*), *[eě]*, a to u sloves IV. třídy vzorů *trpět, sázet* (srov. *trpění, sázení*) a *[nm]u*, a to u sloves II. třídy vzorů *tisknout, minout* a *začít* (srov. *stisknutí, pominutí, vzedmutí*). Po KmV *0* následuje *[e]* u sloves I. třídy vzorů *nést, péci* (srov. *nesení, pečení*), *[eě]* u sloves II. třídy vzoru *tisknout* (srov. *tištění, nadšení*) a *[e]* u sloves IV. třídy vzoru *prosit* (srv. *vyprošení*)[[2]](#footnote-2). Zakončení *[nt]í* stojí bezprostředně za kořenovým vokálem a) u sloves III. třídy vzoru *krýt*, která v češtině, pokud nejde o KoV *[aá]* (např. *hraní, sání*), mají u těchto derivátů vždy KoV krátký – všechny krátké vokály kromě *o* lze doložit, neboť seznam sloves je úplný, protože jde o uzavřenou třídu, (srov. *spění, vypletí, bití, obutí, umytí*) a b) u atematických/nepravidelných sloves (*jedení, chtění, bytí, ...*).

Výsledky empirie jsou takto opřeny o pravidla gramatiky.

**Formulace dalších otázek vyplynuvších ze zkoumání daného jevu**

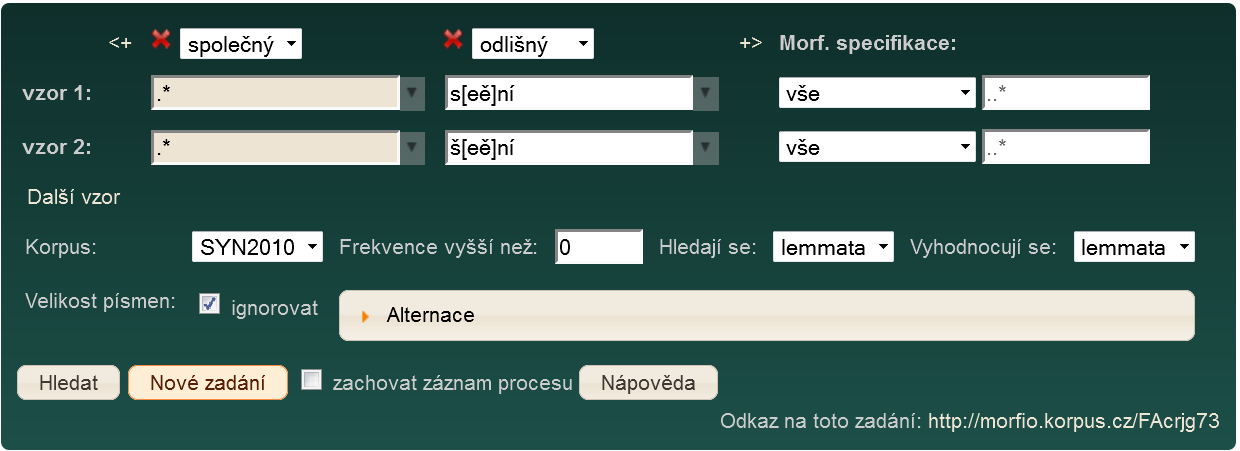
Najděte v korpusu lemma *obmýtí*.

Lze jednoduše (jedním dotazem) vybrat dějová jména, která jsou zaručeně odvozena od sloves určité třídy? Lze totéž udělat pro určitý vzor?

Ručně vyberte frekventovaná slova, která splňují formální podmínky zakončení dějových jmen, ale dějovými jmény nejsou (*století, úpatí, …*). Je všechna z nich třeba vyhledávat ručně? Lze najít nějaký postup, který by pomohl při ručním vyhledávání?

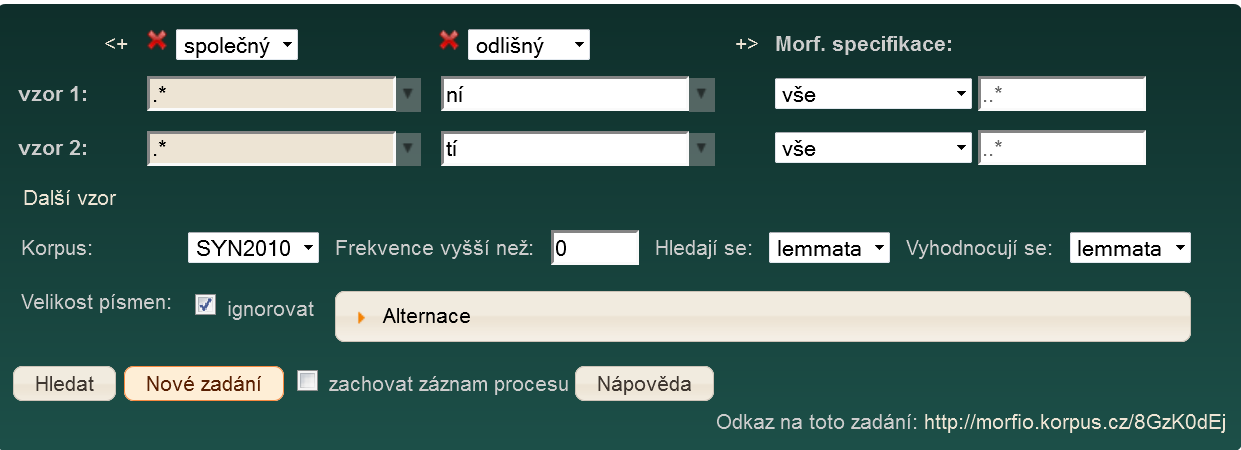
**Vyhledejte dvojice, které se liší:**

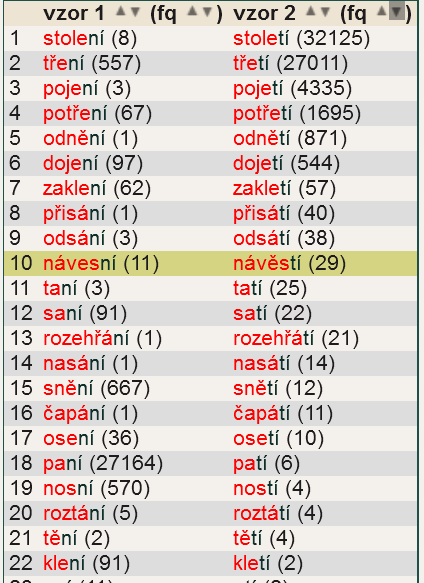
1. alternací finály kořene (*mísení/míšení*)
2. –ní/-tí (*kletí, klení*)





b)





Dú

Kdy jde v případě dvojice dějových jmen o synonyma a kdy jde o případy významově odlišné?

|  |  |
| --- | --- |
| Babincová, Adriana | z-ž (1.-12.) |
| Dvořáková, Jitka | z-ž (13-24) |
| Machura, Jakub | z-ž (25-36) |
| Ryšavá, Daniela | zd-žd |
| Stará, Marie | d-z |
| Volková, Nikol | t-c |

1. Více Osolsobě 20111. [↑](#footnote-ref-1)
2. Toto rozdělení vychází z Komárkových prací a podrobně je popsáno ve druhém svazku *Mluvnice češtiny* (Komárek a kol. 1987 : 427–495). Odpověď na otázku, proč je *e/ě* u sloves některých tříd a vzorů interpretováno jako kmenotvorná přípona a u jiných jako část tvarotvorné přípony pasivního participia, je možné synchronně zdůvodnit na základě a) analogie v rámci tvarů od kmene minulého a b) shod a neshod na úrovni hláskových alternací finálního konsonantu kořene (před kmenotvornou příponou, jíž je *0* nebo *e/ě*). Proč je *e/ě* u sloves I. třídy vzorů *nést, péci*, II. třídy vzoru *tisknout* a IV. třídy vzoru *prosit* součástí tvarotvorné přípony a u sloves I. třídy vzoru *umřít* a IV. třídy vzorů *trpět, sázet* je chápáno jako kmenotvorná přípona? V případě sloves I. třídy vzorů *nést, péci* a II. třídy vzoru *tisknout* jde o systém tvarů od kmene minulého. Všechny tvary od kmene minulého mají u sloves I. třídy vzorů *nést* a *péci* kmenotvornou příponu *0*. Tvary od kmene minulého mohou mít u sloves vzoru *tisknout* kmenotvornou příponu buď *nu/nou*, nebo *0*. V případě sloves IV. třídy je důvod jiný. Kmenotvorná přípona *0* u sloves IV. třídy vzoru *prosit* nezpůsobuje alternaci finál *[bmpv]* na švu kořen/kmenotvorná přípona (*rozzloben, vykrmen, potupen, zastaven, pohřben, …*), ale fakultativně způsobuje alternace jiných finál (např. *s>š prošen, z>ž zaražen, ť>c zaplacen, ď>z zrazen, …*). Kmenotvorná přípona *e/ě* u sloves 4. třídy vzorů *trpět* a *sázet* způsobuje alternaci finál *[bmpv]* na švu kořen/kmenotvorná přípona (*vyráběn, srozuměn, protrpěn, zastavěn, …*), ale nezpůsobuje jiné alternace, a to zejména u sloves vzoru *sázet* (např. *zasazen, obletěn, předváděn, …*). Pokud by se v obou případech počítalo s kmenotvornou příponou *e/ě*, museli bychom předpokládat dvě různá *e* v kmenotvorné příponě sloves téže třídy. Jisté anomálie nacháyíme ovšem u dějových jmen na *[eě]ní*. Tak např. dějová jména jako *utrpení* a *uvědomění* jsou s výše uvedeným pravidlem o chování labiál v rozporu. Důvodem u substantiva *utrpení* je tzv. pokles jotace ve stč. (substantivum *utrpění* je doloženo i v synchr. korpusech). Rovněž substantivum *uvědomení* je v korpusu doloženo. Slovníky uvádějí dubletu u pas. příč., přičemž pravidelné *uvědomen* hodnotí jako řidč. kniž. [↑](#footnote-ref-2)