

21. 10. Příslušné spřežky, slovnědruhové přesahy tvary na *o*

PLIN032, G13, 10.00-11.30

Jak jsou lemmatizovány a slovnědruhově interpretovány tvary adverbii, které končí na *-o*?

```
[lc=".*o" & tag="D.*"]
```

Frekvenční distribuce x

[Běžná](#) [Podle typů textů](#) [Závislost dvou atributů](#)

Frekvenční limit:

Úroveň	Atribut	Nerozlišovat velikost	Pozice	(Node) začít od
1.	<input type="text" value="lc [lowercase word]"/>	<input type="checkbox"/>	Node ▾	slova KWIC nejvíce vlevo ▾ ▲ x
2.	<input type="text" value="lemma_lc [lowercase lemma]"/>	<input type="checkbox"/>	Node ▾	slova KWIC nejvíce vlevo ▾ ▲ x
3.	<input type="text" value="tag"/>	<input type="checkbox"/>	Node ▾	slova KWIC nejvíce vlevo ▾ ▲ x

[+](#)

[Vytvořit frekvenční seznam](#)

Minimální frekvence:

Celkem: 373 položek (8 stránek)

	Filter	lc [lowercase word]	lemma_lc [lowercase lemma]	tag	Freq	
1	p / n	proto	proto	Db-----	64168	
2	p / n	často	často	Dg----1A---	43788	
3	p / n	co	co	Db-----	34447	
4	p / n	přímo	přímo	Dg----1A---	30191	
5	p / n	dlouho	dlouho	Dg----1A---	29615	
6	p / n	přesto	přesto	Db-----	24591	
7	p / n	skoro	skoro	Db-----	22441	
8	p / n	daleko	daleko	Dg----1A---	16710	
9	p / n	málo	málo	Dg----1A---	15980	
10	p / n	ráno	ráno	Db-----	15388	
11	p / n	naprosto	naprosto	Db-----	14021	
12	p / n	snadno	snadno	Dg----1A---	11728	
13	p / n	dávno	dávno	Dg----1A---	11161	
14	p / n	takto	takto	Db-----	11146	
15	p / n	těžko	těžko	Dg----1A---	9222	
16	p / n	nedávno	nedávno	Db-----	8585	
17	p / n	najevo	najevo	Db-----	5140	
18	p / n	vlevo	vlevo	Db-----	4157	
19	p / n	vpravo	vpravo	Db-----	3957	
20	p / n	líto	líto	Db-----	3949	
21	p / n	blízko	blízko	Dg----1A---	3671	
22	p / n	vysoko	vysoko	Dg----1A---	3654	
23	p / n	hluboko	hluboko	Dg----1A---	3653	
24	p / n	nato	nato	Db-----	3285	

Proč jsou některá označována jako stupňovatelná a jiná nikoli?

Je pozitivem od *snadněji* vždy tvar *snadno*?

Proč je lemmatem tvaru *brzo* tvar *brzy*? Existuje více takových případů?

```
P-filtr [lc=".*o" & lemma!=".*o"]
```

Frekvenční seznam

1 / 1

Minimální frekvence:

Celkem: 7 položek (1 stránka)

	Filter	lc [lowercase word]	lemma_lc [lowercase lemma]	tag	Freq
1	p / n	brzo	brzy	Dg-----1A--1-	2203
2	p / n	blizoučko	blizouče	Db-----	48
3	p / n	teďko	teďka	Db-----6-	21
4	p / n	stopro	stoprocentně	Dg-----1A---	19
5	p / n	ko	ko	Db-----8-	3
6	p / n	al-ko	al-ko	Db-----8-	1
7	p / n	dto	dto	Db-----8-	1

Morfio

<+ společný odlišný >+ Morf. specifikace:

vzor 1:

vzor 2:

Další vzor

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

Velikost písmen: ignorovat

Odkaz na toto zadání: <http://morfio.korpus.cz/SFr2qjdl>

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	brzy (16127)	brzo (2468)
2	teďky (3)	teďko (59)

Jak lematizovat a interpretovat tvar *such*, *sucha*, *sucho*, *suchy* ale i *suši*?

- Bylo velké **sucho**.* (substantivum → POS=N)
- Bylo velmi **sucho**.* (adverbium → POS=D);
- Bylo sucho.*** (adverbium → POS=D)
- Dítě bylo **sucho**.* (adjektivum – jmenný tvar → POS=A);
- Chlapi seděli v hospodě **na sucho**.* (adverbium → POS=D)
- Chlapi si v hospodě stěžovali **na sucho**.* (substantivum → POS=N)
- Do sucha**, které ničilo úrodu v roce 1947, jsem hroznější počasí nezažil.* (substantivum → POS=N)
- Utři se **do sucha**.* (adverbium → POS=D)
- Jsem už zcela **such**, manželka je **sucha** i dítě jsou již **sucha**, ale šaty nejsou dosud zcela **suchy**.* (adjektivum – jmenný tvar → POS=A)

j. Tolik **such** nebylo ani v minulém století. (substantivum → POS=N)

k. Tak to by bylo v **suchu**. (adverbium → POS=D)

l. V takovém dlouhotrvajícím **suchu** dojde snadno k požáru. (substantivum → POS=N)

m. Nevěděli jsme si rady s velkými **suchy**. (substantivum → POS=N)

Podívejte se do korpusu na tvary předložkových spojení a spřežek.

Vyhledejte v korpusu dvojice adverbii od stejného základu lišící se slovnědruhovou charakteristikou (např. *smutně/smutno*).

- - - - -

<+ společný odlišný +> Morf. specifikace:

vzor 1:

vzor 2:

Další vzor

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

Velikost písmen: ignorovat

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

Pozorujte vyhledané dvojice a pokuste se je klasifikovat.

Zopakujte vyhledávání tak, aby typ na *o* nebyl rozpoznán automatickou analýzou.

<+ společný odlišný +> Morf. specifikace:

vzor 1:

vzor 2:

Další vzor

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

Velikost písmen: ignorovat

Odkaz na toto zadání: <http://morfio.korpus.cz/WKc9fFMJ>

Výsledky seřad'te podle frekvence vzoru2.

Definujte nedostatky slovníku použitého analyzátoru.

Podívejte se na typ adverbíí na y. Jak jsou lematizovány a značkovány tvary typu *po* .*/sc/ku?

Závěry

Jmenné tvary adjektiv

Adverbializace jmenných tvarů

Pravidla pro zaznamenávání adverbializovaných jmenných tvarů ve slovnících automatických analyzátorů

Pravidla pro desambiguaci homonymní tvarů

Úkol na 28.10. (25.11.)

V návrhu na nové značkování české morfologie (NovaMorf) se předpokládá, že tvary pasivních příčestí budou lematizovány několikanásobným lemmatem, a sice dlouhým tvarem adjektiva tvořeného z krátkého tvaru (příčestí) a infinitivem.

Jednou z námitek proti tomuto řešení byla údajná nedoloženost dlouhých tvarů.

V synv7 jsou doloženy tvary pasiv bez odpovídajících složených tvarů: *důvěřováno, křivděno, reptáno, hřešeno, poshověno, hýbáno, naléháno, přáno, zacházeno, pohlíženo, jednáno, vyhrožováno, hovořeno, přihlíženo, hleděno, zabraňováno, vycházeno, pracováno, referováno, reagováno, dohlíženo, pátráno...* a další.

Projděte korpusy (webové – řada Araneum, czTenTen) a shromážděte doklady odporující uvedené námitce. Pokuste se najít důvody pro výskyt doložených dlouhých tvarů.