**4. seminář**

**Úkoly na semináře 15., 16. a 17. října**

**Opakování**

1. Srovnejte frekvenci variant s krátkou a dlouhou samohláskou v kořenu slova: *umyvadlo* x *umývadlo*, *tematický* x *tématický*.
2. Vyhledejte v Českém národním korpusu SYN2020 **částice** užívané v češtině.

Dotaz: [tag="T.\*"]

1. Zjistěte, které tvary imperativu sloves *zabrzdit* a *přihlásit se* v korpusu SYN2020 vyskytují.
2. Vyhledejte substantiva tvořená sufixem *-yně*. Zaměřte se na 30 nejvíce frekventovaných: jaký význam mají tato jména nejčastěji?

**Úkol k zamyšlení**

1. V současné době znáte atributy **word**, **lemma**, **sublemma** a tag a umíte podle nich vyhledávat. Zamyslete se tedy nad následujícím: který atribut využijete **při vytváření frekvenčního seznamu**, abyste zjistili, kterým slovním druhem / kterými slovními druhy mohou být TVARY SLOV (= word): *mezi*, *počátkem*, *počínaje*?

1. Srovnejte frekvenci variant s krátkou a dlouhou samohláskou v kořenu slova: *umyvadlo* x *umývadlo*, *tematický* x *tématický*.

Řešení:

Můžeme hledat každé sublemma zvlášť, ale také všechna 4 sublemmata současně:

[sublemma="tematický|tématický|umyvadlo|umývadlo"]

A pak pomocí „Frekvence vlastní“ vytvořit frekvenční seznam sublemmat:



Výsledek:

 / 1

(celkem: 4 položky)

Sdílet tabulku:

|  | **Filtr** | **sublemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | tematický | 953 | 7,82 |
| 2 | p / n | tématický | 62 | 0,51 |
| 3 | p / n | umyvadlo | 1 434 | 11,77 |
| 4 | p / n | umývadlo | 121 | 0,99 |

tematický : tématický 15:1

umyvadlo : umývadlo 12:1

V obou případech převažují varianty s krátkou samohláskou.

2. Vyhledejte v Českém národním korpusu SYN2020 **částice** užívané v češtině.

Dotaz: [tag="T.\*"]

Výsledek: 186 položek

Sdílet tabulku:

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | až | 115 379 | 947,07 |
| 2 | p / n | ne | 100 984 | 828,92 |
| 3 | p / n | tedy | 84 258 | 691,62 |
| 4 | p / n | asi | 65 519 | 537,81 |
| 5 | p / n | například | 58 260 | 478,22 |
| 6 | p / n | možná | 52 651 | 432,18 |
| 7 | p / n | třeba | 52 412 | 430,22 |
| 8 | p / n | vůbec | 52 125 | 427,86 |
| 9 | p / n | to | 50 836 | 417,28 |
| 10 | p / n | totiž | 44 655 | 366,55 |
| 11 | p / n | dokonce | 43 874 | 360,13 |
| 12 | p / n | ano | 43 579 | 357,71 |
| 13 | p / n | nakonec | 41 943 | 344,28 |
| 14 | p / n | spíš | 36 524 | 299,8 |
| 15 | p / n | především | 35 699 | 293,03 |
| 16 | p / n | vlastně | 33 452 | 274,59 |
| 17 | p / n | samozřejmě | 33 363 | 273,86 |
| 18 | p / n | snad | 32 838 | 269,55 |
| 19 | p / n | přece | 32 603 | 267,62 |
| 20 | p / n | ovšem | 31 027 | 254,68 |
| 21 | p / n | prostě | 27 691 | 227,3 |
| 22 | p / n | zejména | 27 575 | 226,35 |
| 23 | p / n | no | 25 773 | 211,55 |
| 24 | p / n | naopak | 25 561 | 209,81 |
| 25 | p / n | určitě | 25 481 | 209,16 |
| 26 | p / n | prý | 21 045 | 172,75 |
| 27 | p / n | skutečně | 19 373 | 159,02 |
| 28 | p / n | zřejmě | 18 393 | 150,98 |
| 29 | p / n | jo | 18 114 | 148,69 |
| 30 | p / n | rozhodně | 16 249 | 133,38 |
| 31 | p / n | opravdu | 14 094 | 115,69 |
| 32 | p / n | nejspíš | 13 967 | 114,65 |
| 33 | p / n | bohužel | 12 737 | 104,55 |
| 34 | p / n | nikoli | 11 929 | 97,92 |
| 35 | p / n | ostatně | 8 163 | 67,01 |
| 36 | p / n | zvlášť | 7 665 | 62,92 |
| 37 | p / n | zkrátka | 7 615 | 62,51 |
| 38 | p / n | fakt | 5 649 | 46,37 |
| 39 | p / n | ať | 5 165 | 42,4 |
| 40 | p / n | co | 4 839 | 39,72 |
| 41 | p / n | zjevně | 4 517 | 37,08 |
| 42 | p / n | patrně | 4 327 | 35,52 |
| 43 | p / n | copak | 4 245 | 34,85 |
| 44 | p / n | přinejmenším | 4 026 | 33,05 |
| 45 | p / n | doopravdy | 3 468 | 28,47 |
| 46 | p / n | nepochybně | 3 414 | 28,02 |
| 47 | p / n | koneckonců | 2 872 | 23,57 |
| 48 | p / n | ba | 2 793 | 22,93 |
| 49 | p / n | pozor | 2 677 | 21,97 |
| 50 | p / n | akorát | 2 499 | 20,51 |

.

3. Zjistěte, které tvary imperativu sloves *zabrzdit* a *přihlásit se* v korpusu SYN2020 vyskytují.

Dotaz:

[tag="Vi.\*"&lemma="zabrzdit|přihlásit"]

Frekvence – slovní tvary (seřazeno abecedně):

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | přihlas | 11 | 0,09 |
| 2 | p / n | přihlasme | 1 | 0,01 |
| 3 | p / n | přihlaste | 120 | 0,99 |
| 4 | p / n | přihlaš | 1 | 0,01 |
| 5 | p / n | přihlašte | 8 | 0,07 |
| 6 | p / n | zabrzděte | 3 | 0,03 |
| 7 | p / n | zabrzdi | 1 | 0,01 |

Imperativní tvary slovesa *zabrzdit* jsou doloženy v minimálním počtu.

U slovesa *přihlásit* jsou doloženy vedle základních tvarů přihlas, přihlasme, přihlaste variantní tvary *přihlaš*, *přihlašte*, avšak ve výrazné menšině.

4. Vyhledejte substantiva tvořená sufixem *-yně*. Zaměřte se na 30 nejvíce frekventovaných: jaký význam mají tato jména nejčastěji?

Dotaz: [lemma=".\*yně"]

Z vyhledaných dokladů vytvoříme frekvenční seznam lemmat:

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | kuchyně | 15 631 | 128,31 |
| 2 | p / n | přítelkyně | 4 289 | 35,21 |
| 3 | p / n | jeskyně | 4 135 | 33,94 |
| 4 | p / n | kolegyně | 2 297 | 18,86 |
| 5 | p / n | ministryně | 2 036 | 16,71 |
| 6 | p / n | bohyně | 1 523 | 12,5 |
| 7 | p / n | předsedkyně | 1 267 | 10,4 |
| 8 | p / n | svatyně | 1 108 | 9,1 |
| 9 | p / n | tchyně | 954 | 7,83 |
| 10 | p / n | zástupkyně | 904 | 7,42 |
| 11 | p / n | soudkyně | 791 | 6,49 |
| 12 | p / n | hospodyně | 770 | 6,32 |
| 13 | p / n | mistryně | 679 | 5,57 |
| 14 | p / n | umělkyně | 644 | 5,29 |
| 15 | p / n | poradkyně | 614 | 5,04 |
| 16 | p / n | náměstkyně | 608 | 4,99 |
| 17 | p / n | poslankyně | 550 | 4,52 |
| 18 | p / n | zaměstnankyně | 418 | 3,43 |
| 19 | p / n | žákyně | 415 | 3,41 |
| 20 | p / n | místopředsedkyně | 409 | 3,36 |
| 21 | p / n | vévodkyně | 409 | 3,36 |
| 22 | p / n | průvodkyně | 378 | 3,1 |
| 23 | p / n | svědkyně | 290 | 2,38 |
| 24 | p / n | pěvkyně | 267 | 2,19 |
| 25 | p / n | Ruzyně | 254 | 2,09 |
| 26 | p / n | sportovkyně | 250 | 2,05 |
| 27 | p / n | běžkyně | 238 | 1,95 |
| 28 | p / n | otrokyně | 221 | 1,81 |
| 29 | p / n | důchodkyně | 220 | 1,81 |
| 30 | p / n | žalobkyně | 206 | 1,69 |

Kromě 4 položek jde vždy o názvy přechýlené.

**Úkol k zamyšlení**

5. V současné době znáte atributy **word**, **lemma**, **sublemma** a tag a umíte podle nich vyhledávat. Zamyslete se tedy nad následujícím: který atribut využijete **při vytváření frekvenčního seznamu**, abyste zjistili, kterým slovním druhem / kterými slovními druhy mohou být TVARY SLOV (= word): *mezi*, *počátkem*, *počínaje*?

**I. Nejjednodušší postup:**

1. Vyhledáme doklady pomocí atributu WORD: [word="mezi|počátkem|počínaje"]

2. Vytvoříme frekvenci lemmat: u slov ohebných mez, počátek, počínat je lemmatem tento základní tvar, u neohebných je lemma totožné se zadaným wordem. Tento jednoduchý postup však můžeme uplatnit tehdy, víme-li, kterým slovními druhy neohebnými a ohebnými zadaná slova bývají, a chceme jen srovnat četnost jejich výskytu.

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | mezi | 101 561 | 833,65 |
| 2 | p / n | počátkem | 1 102 | 9,05 |
| 3 | p / n | počínaje | 983 | 8,07 |
| 4 | p / n | počátek | 541 | 4,44 |
| 5 | p / n | počínat | 80 | 0,66 |
| 6 | p / n | mez | 60 | 0,49 |

II. Chceme-li BEZPEČNĚ zjistit, kterým slovním druhem tato slova jsou, musíme vytvořit frekvenční seznam „dvouúrovňový“ pomocí Frekvence vlastní.

1. Vyhledáme doklady pomocí atributu WORD: [word="mezi|počátkem|počínaje"]

2. Využijeme dvouúrovňovou Frekvenci vlastní, obě úrovně se budou týkat jedné pozice (Node). Jedna úroveň bude lemma, druhá úroveň tag.



**Výsledek:**

|  | **Filtr** | **lemma** | **tag** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | mezi | RR--7---------- | 86 035 | 706,21 |
| 2 | p / n | mezi | RR--4---------- | 15 526 | 127,44 |
| 3 | p / n | počátkem | RR--2---------- | 1 102 | 9,05 |
| 4 | p / n | počínaje | RR--7---------- | 983 | 8,07 |
| 5 | p / n | počátek | NNIS7-----A---- | 541 | 4,44 |
| 6 | p / n | počínat | VeMS------A-I-- | 80 | 0,66 |
| 7 | p / n | mez | NNFS6-----A---- | 48 | 0,39 |
| 8 | p / n | mez | NNFS3-----A---- | 11 | 0,09 |
| 9 | p / n | mez | NNFS5-----A---- | 1 | 0,01 |

Slovní druh určí **tag**. Nevýhodou je, že tag určuje i další gramatické vlastnosti, tj. substantivum *mez* je rozděleno do tří položek podle pádu tvaru *mezi*, předložka *mezi* se dělí do dvou položek podle toho, s kterým pádem se pojí.

Můžeme však využít atribut, který URČÍ POUZE SLOVNÍ DRUH: pos



**Nastavení frekvenčního seznamu – lemma + pos (part of speech = slovní druh)**



**Výsledek**

|  | **Filtr** | **lemma** | **pos [part of speech]** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | mezi | R | 101 561 | 833,65 |
| 2 | p / n | počátkem | R | 1 102 | 9,05 |
| 3 | p / n | počínaje | R | 983 | 8,07 |
| 4 | p / n | počátek | N | 541 | 4,44 |
| 5 | p / n | počínat | V | 80 | 0,66 |
| 6 | p / n | mez | N | 60 | 0,49 |

**Vyhledejte následující slova a určete, kterým slovním druhem mohou být:**

***chladno*, *středem*, *pomocí*, *nevyjímaje*.**

Postup:

1. Dotaz: [word="chladno|středem|pomocí|nevyjímaje"]
2. Vyhledání dokladu + vlastní frekvence dvouúrovňová:

**Frekvence vlastní – dvouúrovňová:**

**1. lemma**

**2. pos [part of speech]**



|  | **Filtr** | **lemma** | **pos [part of speech]** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | pomocí | R | 14 925 | 122,51 |
| 2 | p / n | pomoc | N | 4 281 | 35,14 |
| 3 | p / n | střed | N | 952 | 7,81 |
| 4 | p / n | chladno | D | 520 | 4,27 |
| 5 | p / n | vyjímat | V | 342 | 2,81 |
| 6 | p / n | chladno | N | 1 | 0,01 |
| 7 | p / n | středem | D | 1 | 0,01 |

Pouze slovo *pomocí* funguje jako tvar substantiva a jako předložka.

Podobně můžeme určit, kterým slovním druhem bývá slovo *vzhledem*.



**Opakování s prezentací**

**1. Ověřte v korpusu SYN2020 existenci pravopisných variant. Pokud najdete obě, napište přesné počty a vyjádřete číselně poměr mezi jejich frekvencí (např. 6:1, 8:1 apod.)**

a) citron – citrón

b) kurziva – kurzíva

c) desinfekce – dezinfekce

Dotaz:

[sublemma="citron|citrón|kurziva|kurzíva|desinfekce|dezinfekce"]

Vytvoření frekvenčního seznamu **sublemmat** pomocí „Frekvence – vlastní“:



**Abecední seznam:**

|  | **Filtr** | **sublemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | citron | 1 498 | 12,3 |
| 2 | p / n | citrón | 174 | 1,43 |
| 3 | p / n | desinfekce | 26 | 0,21 |
| 4 | p / n | dezinfekce | 619 | 5,08 |
| 5 | p / n | kurziva | 88 | 0,72 |
| 6 | p / n | kurzíva | 67 | 0,55 |

citron : citrón 9:1

kurziva : kurzíva 1,3:1

desinfekce : dezinfekce 1:24

**2. Jaké se užívají tvary v lokálu singuláru substantiva *kořen*? Najdete-li dvojtvary, napište, jaký je poměr jejich frekvencí. Využijte korpus SYN2020.**

Dotaz: [tag="N..S6.\*"&lemma="kořen"]

| **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- |
| p / n | kořeni | 63 | 0,52 |
| p / n | kořenu | 10 | 0,08 |

kořenu : kořeni 1 : 6

**3. Vytvořte frekvenční seznam slov s kořenem *želez*. Kolik položek jste vyhledali? ? Našli jste nějaké chybné nebo nevyhovující požadavku zadání?**

**Dotazy:**

[lemma=".\*želez.\*"]

[lemma=".\*želíz.\*"]

Vyhledali jsme 75 položek, v .. případech šlo o chyby: tvary *želez, železníc, železničím, železnú, železničné, železničného* hodnocené jako lemma, chyby *zeželezem, zajistilželezniční, železaren, železaje, železito, železitáazemitávůnězbrojířádnýchčlenůalahodnéaroma.*

Po vyškrtání 12 chyb zbude 63 položek. U složených slov je rozkolísaný pravopis.

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | železniční | 4 152 | 34,08 |
| 2 | p / n | železný | 3 569 | 29,3 |
| 3 | p / n | železo | 2 211 | 18,15 |
| 4 | p / n | železnice | 1 977 | 16,23 |
| 5 | p / n | železobetonový | 745 | 6,12 |
| 6 | p / n | železárna | 241 | 1,98 |
| 7 | p / n | železničář | 209 | 1,72 |
| 8 | p / n | železářství | 136 | 1,12 |
| 9 | p / n | železobeton | 120 | 0,99 |
| 10 | p / n | železitý | 114 | 0,94 |
| 11 | p / n | železářský | 66 | 0,54 |
| 12 | p / n | železničářský | 35 | 0,29 |
| 13 | p / n | železnatý | 33 | 0,27 |
| 14 | p / n | železa | 30 | 0,25 |
| 15 | p / n | železnobrodský | 22 | 0,18 |
| 16 | p / n | železnorudný | 20 | 0,16 |
| 17 | p / n | železnorudský | 18 | 0,15 |
| 18 | p / n | železnička | 16 | 0,13 |
| 19 | p / n | lamželezo | 15 | 0,12 |
| 20 | p / n | ~~želez~~ | 9 | 0,07 |
| 21 | p / n | železář | 6 | 0,05 |
| 22 | p / n | železně | 6 | 0,05 |
| 23 | p / n | železník | 5 | 0,04 |
| 24 | p / n | železnohorský | 5 | 0,04 |
| 25 | p / n | železo-niklový | 5 | 0,04 |
| 26 | p / n | železárenský | 4 | 0,03 |
| 27 | p / n | železňák | 4 | 0,03 |
| 28 | p / n | železnický | 4 | 0,03 |
| 29 | p / n | železitoamonný | 3 | 0,03 |
| 30 | p / n | železivec | 3 | 0,03 |
| 31 | p / n | železový | 3 | 0,03 |
| 32 | p / n | železo-betonový | 2 | 0,02 |
| 33 | p / n | železorudný | 2 | 0,02 |
| 34 | p / n | železnica | 2 | 0,02 |
| 35 | p / n | železo-fosfátový | 2 | 0,02 |
| 36 | p / n | železogalových | 2 | 0,02 |
| 37 | p / n | miniželeznice | 1 | 0,01 |
| 38 | p / n | sirnoželezitý | 1 | 0,01 |
| 39 | p / n | superželezo | 1 | 0,01 |
| 40 | p / n | skleněno-železobetonový | 1 | 0,01 |
| 41 | p / n | železo-ocelový | 1 | 0,01 |
| 42 | p / n | železnodorožniků | 1 | 0,01 |
| 43 | p / n | ~~zajistilželezniční~~ | 1 | 0,01 |
| 44 | p / n | ~~železničím~~ | 1 | 0,01 |
| 45 | p / n | železnovodský | 1 | 0,01 |
| 46 | p / n | hexakyanoželezitanu | 1 | 0,01 |
| 47 | p / n | železničářův | 1 | 0,01 |
| 48 | p / n | železogalovým | 1 | 0,01 |
| 49 | p / n | železobetovové | 1 | 0,01 |
| 50 | p / n | kamenoželezný | 1 | 0,01 |
| 51 | p / n | ~~zeželezem~~ | 1 | 0,01 |
| 52 | p / n | ~~železitáazemitávůnězbrojířádnýchčlenůalahodnéaroma~~ | 1 | 0,01 |
| 53 | p / n | ~~železníc~~ | 1 | 0,01 |
| 54 | p / n | ~~železaje~~ | 1 | 0,01 |
| 55 | p / n | železnatohlinitý | 1 | 0,01 |
| 56 | p / n | hořečnato-železnatý | 1 | 0,01 |
| 57 | p / n | ~~železaren~~ | 1 | 0,01 |
| 58 | p / n | sklo-železo-betonový | 1 | 0,01 |
| 59 | p / n | ~~železito~~ | 1 | 0,01 |
| 60 | p / n | železničníukončila | 1 | 0,01 |
| 61 | p / n | sirno-železitý | 1 | 0,01 |
| 62 | p / n | ~~železničné~~ | 1 | 0,01 |
| 63 | p / n | odželezněný | 1 | 0,01 |
| 64 | p / n | železnatoduběnkové | 1 | 0,01 |
| 65 | p / n | hořečnatoželeznatý | 1 | 0,01 |
| 66 | p / n | železnato-železitý | 1 | 0,01 |
| 67 | p / n | ~~železnú~~ | 1 | 0,01 |
| 68 | p / n | železitoamonného | 1 | 0,01 |
| 69 | p / n | ~~železničného~~ | 1 | 0,01 |
| 70 | p / n | železobetopnovým | 1 | 0,01 |
| 71 | p / n | železogalového | 1 | 0,01 |
| 72 | p / n | železogalové | 1 | 0,01 |
| 73 | p / n | železopilka | 1 | 0,01 |
| 74 | p / n | železnicovým | 1 | 0,01 |
| 75 | p / n | železogalovými | 1 | 0,01 |

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | želízko | 186 | 1,53 |
| 2 | p / n | želízka | 52 | 0,43 |
| 3 | p / n | ~~Rucehobolíodželízek~~ | 1 | 0,01 |

**1. Ověřte v korpusu SYN2020 existenci pravopisných variant. Pokud najdete obě, napište přesné počty a vyjádřete číselně poměr mezi jejich frekvencí (např. 6:1 apod.)**

a) benzin – benzín

b) galanterie - galantérie

c) exkurs – exkurz

Dotaz:[sublemma="benzin|benzín|galanterie|galantérie|exkurs|exkurz"]

Vytvoření frekvenčního seznamu **sublemmat** pomocí „Frekvence – vlastní“:



**Abecední seznam:**

|  | **Filtr** | **sublemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | benzin | 1 397 | 11,47 |
| 2 | p / n | benzín | 528 | 4,33 |
| 3 | p / n | exkurs | 32 | 0,26 |
| 4 | p / n | exkurz | 99 | 0,81 |
| 5 | p / n | galanterie | 87 | 0,71 |

benzin : benzín 2,6:1

galanterie pouze 1 varianta

exkurs : exkurz 1:3

1. **Jaké se užívají tvary v lokálu singuláru substantiva *kabát*? Najdete-li dvojtvary, napište, jaký je poměr jejich frekvencí. Využijte korpus SYN2020.**

Dotaz: [tag=“N..S6.\*“&lemma=“kabát“]

| **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- |
| p / n | kabátě | 421 | 3,46 |
| p / n | kabátu | 144 | 1,18 |

kabátu : kabátě 1 : 3

**3. Vytvořte frekvenční seznam slov s kořenem *kořen*. Kolik položek jste vyhledali? Našli jste nějaké chybné nebo nevyhovující požadavku zadání?**

**Dotaz:** [lemma=".\*kořen.\*"] [lemma=".\*kořín.\*"]

Vyhledali jsme 100 položek, z nichž 4 nevyhovovaly (*pokořený, pokořeně, sýkořenka, kořenie*) a v 6 případech šlo o chyby/překlepy: *kořenůmdnešní, skleničkakoření, dřevitěkořennou, kořenů1, kořenu14, Zkořenů.* V 1 případě není tvar slova spojen se svým lemmatem (*neukořeněností*).

Zbývá 89

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | kořen | 5 379 | 44,15 |
| 2 | p / n | koření | 2 301 | 18,89 |
| 3 | p / n | kořenový | 767 | 6,3 |
| 4 | p / n | zakořeněný | 435 | 3,57 |
| 5 | p / n | kořeněný | 392 | 3,22 |
| 6 | p / n | okořenit | 358 | 2,94 |
| 7 | p / n | zakořenit | 284 | 2,33 |
| 8 | p / n | kořenice | 148 | 1,22 |
| 9 | p / n | okořeněný | 138 | 1,13 |
| 10 | p / n | pokoření | 135 | 1,11 |
| 11 | p / n | kořenka | 69 | 0,57 |
| 12 | p / n | kořenitý | 68 | 0,56 |
| 13 | p / n | vykořenit | 65 | 0,53 |
| 14 | p / n | ~~pokořený~~ | 64 | 0,53 |
| 15 | p / n | vykořenění | 63 | 0,52 |
| 16 | p / n | kořenit | 61 | 0,5 |
| 17 | p / n | vykořeněný | 59 | 0,48 |
| 18 | p / n | kořenářka | 57 | 0,47 |
| 19 | p / n | kořenící | 56 | 0,46 |
| 20 | p / n | kořenný | 50 | 0,41 |
| 21 | p / n | zakořenění | 35 | 0,29 |
| 22 | p / n | kořenář | 22 | 0,18 |
| 23 | p / n | kořenáč | 22 | 0,18 |
| 24 | p / n | vykořeněnost | 21 | 0,17 |
| 25 | p / n | kořenující | 20 | 0,16 |
| 26 | p / n | prostokořenný | 19 | 0,16 |
| 27 | p / n | prokořenění | 14 | 0,12 |
| 28 | p / n | kořeninový | 11 | 0,09 |
| 29 | p / n | kořenění | 11 | 0,09 |
| 30 | p / n | dokořenit | 8 | 0,07 |
| 31 | p / n | prokořeněný | 8 | 0,07 |
| 32 | p / n | ~~pokořeně~~ | 8 | 0,07 |
| 33 | p / n | okořenění | 8 | 0,07 |
| 34 | p / n | kořenovník | 7 | 0,06 |
| 35 | p / n | kořenovec | 6 | 0,05 |
| 36 | p / n | kořenovitě | 6 | 0,05 |
| 37 | p / n | překořeněný | 6 | 0,05 |
| 38 | p / n | prokořenit | 5 | 0,04 |
| 39 | p / n | mimokořenový | 5 | 0,04 |
| 40 | p / n | nevykořenitelný | 5 | 0,04 |
| 41 | p / n | kořenovský | 5 | 0,04 |
| 42 | p / n | kořenomorka | 5 | 0,04 |
| 43 | p / n | prokořenělý | 4 | 0,03 |
| 44 | p / n | kořenitost | 4 | 0,03 |
| 45 | p / n | zakořeněnost | 4 | 0,03 |
| 46 | p / n | kořenkový | 4 | 0,03 |
| 47 | p / n | jednokořenové | 3 | 0,03 |
| 48 | p / n | kořenářství | 3 | 0,03 |
| 49 | p / n | překořenit | 3 | 0,03 |
| 50 | p / n | kořenoústka | 3 | 0,03 |
| 51 | p / n | přikořenit | 3 | 0,03 |
| 52 | p / n | kořenně | 3 | 0,03 |
| 53 | p / n | kořenářský | 3 | 0,03 |
| 54 | p / n | protikořenovou | 2 | 0,02 |
| 55 | p / n | zakořenitelný | 2 | 0,02 |
| 56 | p / n | kořenovitý | 2 | 0,02 |
| 57 | p / n | kořeněně | 2 | 0,02 |
| 58 | p / n | kořeniště | 2 | 0,02 |
| 59 | p / n | dokořenění | 2 | 0,02 |
| 60 | p / n | biokoření | 2 | 0,02 |
| 61 | p / n | vykořeněně | 2 | 0,02 |
| 62 | p / n | kořenito-citrusový | 1 | 0,01 |
| 63 | p / n | vkořeněný | 1 | 0,01 |
| 64 | p / n | megakořeněný | 1 | 0,01 |
| 65 | p / n | překořenění | 1 | 0,01 |
| 66 | p / n | kořenářův | 1 | 0,01 |
| 67 | p / n | retrokoření | 1 | 0,01 |
| 68 | p / n | ~~kořenůmdnešní~~ | 1 | 0,01 |
| 69 | p / n | zakořené | 1 | 0,01 |
| 70 | p / n | nepokoření | 1 | 0,01 |
| 71 | p / n | ~~sýkořenka~~ | 1 | 0,01 |
| 72 | p / n | ~~kořenísklenička~~ | 1 | 0,01 |
| 73 | p / n | vlastnokořenné | 1 | 0,01 |
| 74 | p / n | ~~dřevitěkořeněnou~~ | 1 | 0,01 |
| 75 | p / n | vícekořenový | 1 | 0,01 |
| 76 | p / n | jednokořenového | 1 | 0,01 |
| 77 | p / n | mělkokořenných | 1 | 0,01 |
| 78 | p / n | přikořenění | 1 | 0,01 |
| 79 | p / n | kořeněnost | 1 | 0,01 |
| 80 | p / n | vykořeněnec | 1 | 0,01 |
| 81 | p / n | mimokořenově | 1 | 0,01 |
| 82 | p / n | přezakořenění | 1 | 0,01 |
| 83 | p / n | pravokořenný | 1 | 0,01 |
| 84 | p / n | bezkořenný | 1 | 0,01 |
| 85 | p / n | zakořeněně | 1 | 0,01 |
| 86 | p / n | hlínokořen | 1 | 0,01 |
| 87 | p / n | sedmikoření | 1 | 0,01 |
| 88 | p / n | prakořen | 1 | 0,01 |
| 89 | p / n | prokořenitelný | 1 | 0,01 |
| 90 | p / n | ~~kořenu14~~ | 1 | 0,01 |
| 91 | p / n | ~~Zkořenů~~ | 1 | 0,01 |
| 92 | p / n | Pětikoření | 1 | 0,01 |
| 93 | p / n | kořenokvětka | 1 | 0,01 |
| 94 | p / n | ~~kořenie~~ | 1 | 0,01 |
| 95 | p / n | ~~kořenů1~~ | 1 | 0,01 |
| 96 | p / n | ~~Neukořeněností~~ | 1 | 0,01 |
| 97 | p / n | pětikoření | 1 | 0,01 |
| 98 | p / n | zvěrokořenářky | 1 | 0,01 |
| 99 | p / n | nevykořenitelnost | 1 | 0,01 |
| 100 | p / n | vkořenit | 1 | 0,01 |

U druhého dotazu jsou jen 3 správné doklady.

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | kořínek | 508 | 4,17 |
| ~~2~~ | ~~p / n~~ | ~~Kokořínsko~~ | ~~95~~ | ~~0,78~~ |
| ~~3~~ | ~~p / n~~ | ~~kokořínský~~ | ~~35~~ | ~~0,29~~ |
| ~~4~~ | ~~p / n~~ | ~~Kokořín~~ | ~~13~~ | ~~0,11~~ |
| ~~5~~ | ~~p / n~~ | ~~Oskořínek~~ | ~~4~~ | ~~0,03~~ |
| 6 | p / n | kořínkový | 3 | 0,03 |
| 7 | ~~p / n~~ | ~~kořínkovitým~~ | ~~2~~ | ~~0,02~~ |
| 8 | ~~p / n~~ | ~~kořínkovitých~~ | ~~1~~ | ~~0,01~~ |
| 9 | ~~p / n~~ | ~~kořínci~~ | ~~1~~ | ~~0,01~~ |
| 10 | p / n | kořínkovitý | 1 | 0,01 |

**1. Ověřte v korpusu SYN2020 existenci pravopisných variant. Pokud najdete obě, napište přesné počty a vyjádřete číselně poměr mezi jejich frekvencí (např. 6:1, 8:1 apod.)**

a) faraon – faraón

b) disciplina - disciplína

c) konkurs – konkurz

**Dotaz:**

[sublemma="faraon|faraón|disciplina|disciplína|konkurs|konkurz"]

Vytvoření frekvenčního seznamu **sublemmat** pomocí „Frekvence – vlastní“:



**Abecední seznam:**

|  | **Filtr** | **sublemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | disciplina | 120 | 0,99 |
| 2 | p / n | disciplína | 3 853 | 31,63 |
| 3 | p / n | faraon | 276 | 2,27 |
| 4 | p / n | faraón | 105 | 0,86 |
| 5 | p / n | konkurs | 156 | 1,28 |
| 6 | p / n | konkurz | 835 | 6,85 |

faraon : faraón 2,6:1

disciplina : disciplín 32:1

konkurs : konzurz 5:1

**2. Jaké se užívají tvary v lokálu singuláru substantiva *záhon*? Najdete-li dvojtvary, napište, jaký je poměr jejich frekvencí. Využijte korpus SYN2020.**

Dotaz: [tag="N..S6.\*"&lemma="záhon"]

| **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- |
| p / n | záhoně | 55 | 0,45 |
| p / n | záhonu | 141 | 1,16 |

záhonu : záhoně 3:1

**3. Vytvořte frekvenční seznam slov s kořenem *kabát*. Kolik položek jste vyhledali? Našli jste nějaké chybné nebo nevyhovující požadavku zadání?**

Našli jsme 27 slov, z kterých 1 nevyhovovalo (*kabátu*, *Dalskabáty*) a v 5 případech šlo o překlepy (*Daskabát, rozepnulkabát, kabátutvář, kabátemnebyl*).

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | kabát | 4 852 | 39,83 |
| 2 | p / n | kabátek | 482 | 3,96 |
| 3 | p / n | kabátec | 139 | 1,14 |
| 4 | p / n | červenokabátník | 30 | 0,25 |
| 5 | p / n | ~~Daskabát~~ | 6 | 0,05 |
| 6 | p / n | kabátový | 5 | 0,04 |
| 7 | p / n | černokabátník | 3 | 0,03 |
| 8 | p / n | překabátit | 3 | 0,03 |
| 9 | p / n | kabátovský | 2 | 0,02 |
| 10 | p / n | ~~Dalskabáty~~ | 2 | 0,02 |
| 11 | p / n | kabátnický | 1 | 0,01 |
| 12 | p / n | kabátuprostý | 1 | 0,01 |
| 13 | p / n | modrokabátník | 1 | 0,01 |
| 14 | p / n | prekabátené | 1 | 0,01 |
| 15 | p / n | ~~rozepnulkabát~~ | 1 | 0,01 |
| 16 | p / n | kabátutvář | 1 | 0,01 |
| 17 | p / n | černokabátnický | 1 | 0,01 |
| 18 | p / n | zelenokabátník | 1 | 0,01 |
| 19 | p / n | ~~kabáta~~ | 1 | 0,01 |
| 20 | p / n | černokabátníkův | 1 | 0,01 |
| 21 | p / n | maxikabát | 1 | 0,01 |
| 22 | p / n | rudokabátník | 1 | 0,01 |
| 23 | p / n | převlečkabátem | 1 | 0,01 |
| 24 | p / n | ~~kabátemnebyl~~ | 1 | 0,01 |
| 25 | p / n | kabátus | 1 | 0,01 |

**Pády předložek**

Zjistěte, s kterým pádem se pojí předložky *před*, *nad*, *skrz*.

Postup:

1) Vyhledáme předložky jako lemmata v kombinaci s tagem pro slovní druh (někdy jsou tyto výrazy označkovány jako „zkratky“, nebo dokonce jako adverbia).

[tag="R.\*"&lemma="nad|před|skrz"]

2) Zadáme „Frekvenci – vlastní“ jako kombinaci lemmatu (1. úroveň) a tagu (2. úroveň)

ve stejné pozici:



Abecední seznam podle lemmatu

|  | **Filtr** | **lemma** | **tag** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | nad | RR--4---------- | 8 878 | 72,87 |
| 2 | p / n | nad | RV--7---------- | 1 036 | 8,5 |
| 3 | p / n | nad | RV--4---------- | 374 | 3,07 |
| 4 | p / n | nad | RR--7---------- | 57 610 | 472,88 |
| 5 | p / n | před | RR--4---------- | 7 278 | 59,74 |
| 6 | p / n | před | RR--7---------- | 134 934 | 1 107,59 |
| 7 | p / n | před | RV--7---------- | 3 332 | 27,35 |
| 8 | p / n | před | RV--4---------- | 236 | 1,94 |
| 9 | p / n | skrz | RV--4---------- | 3 422 | 28,09 |
| 10 | p / n | skrz | RR--4---------- | 2 516 | 20,65 |

RR = základní podoba, RV = vokalizovaná podoba

Rozlišení základní a vokalizované podoby nám sice dává některé zajímavé informace:

podoba skrze převládá nad podobou skrz v poměru cca 1,4:1.

Předložky nad a před se častěji pojí se 7. pádem než se 4. pádem, ale:

* základní podoba předložky *nad* se se 7. pádem pojí častěji než se 4. pádem cca 6,5x (v poměru 6,5:1), vokalizovaná podoba pouze 3x (v poměru 3:1)

zatímco

* základní podoba předložky *před* se pojí se 7. pádem častěji než se 4. pádem cca 18,5x (v poměru 18,5:1), vokalizovaná podoba cca 14x (v poměru 14:1).

Mezi výskytem předložky nad/nade ve spojení se 7. a se 4. pádem je tedy výrazně nižší rozdíl, než je tom u předložky před/přede.

To nám ukáže frekvenční seznam s využitím jiného atributu než tag: case = pád, který nerozliší základní a vokalizovanou podobu.



Abecední seznam podle lemmatu

|  | **Filtr** | **lemma** | **case [grammatical case]** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | nad | 4 | 9 252 | 75,94 |
| 2 | p / n | nad | 7 | 58 646 | 481,39 |
| 3 | p / n | před | 7 | 138 266 | 1 134,94 |
| 4 | p / n | před | 4 | 7 514 | 61,68 |
| 5 | p / n | skrz | 4 | 5 938 | 48,74 |

nad +7. pád : nad + 4. pád 6:1

před +7. pád : před + 4. pád 18:1