**1. Zjistěte, kterými slovními druhy mohou být slova *nahlas*, *stoje*, *prací* a *kout*.**

Postup: 1. Vyhledání „wordů“.

2. Frekvence vlastní: kombinace lemmatu a atributu „pos“



|  | **Filtr** | **lemma** | **pos [part of speech]** | **Freq** | **i.p.m.** | **srovnání slovnědruhové platnosti** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | p / n | kout | V | 34 | 0,28 |  substantivum : verbum 21:1 (verbum *kout* je knižní) |
| 2 | p / n | kout | N | 704 | 5,78 |
|  | p / n | Nahlas | D | 3868 | 31,75 |  |
| 3 | p / n | práce | N | 6 767 | 55,55 | substantivum : adjektivum 50:1(subst. 7. pád sg., 2. pád pl.adj. v různých tvarech) |
| 4 | p / n | prací | A | 135 | 1,11 |
| 5 | p / n | stát | V | 140 | 1,15 | verbum : substantivum 1,3:1(přechodník slovesa *stát*subst. 2. pád sg., 1. a 4. pád pl.) |
| 6 | p / n | stoj | N | 104 | 0,85 |

**2. S kterými pády se pojí předložky *na*, *za*, *naproti*?**

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="na|za|naproti"]

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

ve stejné pozici:



Abecední seznam:

|  | **Filtr** | **lemma** | **case [grammatical case]** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | na | 4 | 1 090 065 | 8 947,66 |
| 2 | p / n | na | 6 | 781 052 | 6 411,17 |
| 3 | p / n | na|co | 4 | 691 | 5,67 |
| 4 | p / n | na|copak | 4 | 3 | 0,03 |
| 5 | p / n | na|on | 4 | 50 | 0,41 |
| 6 | p / n | naproti | 2 | 22 | 0,18 |
| 7 | p / n | naproti | 3 | 4 241 | 34,81 |
| 8 | p / n | za | 4 | 259 164 | 2 127,32 |
| 9 | p / n | za | 7 | 86 867 | 713,04 |
| 10 | p / n | za | 2 | 33 439 | 274,48 |
| 11 | p / n | za|co | 4 | 1 279 | 10,5 |
| 12 | p / n | za|on | 4 | 2 | 0,02 |

na|co = nač, na|copak = načpak, na|on = naň

za|co = zač, za|on = zaň

Předložka *na* se pojí častěji se 4. pádem než se 7. pádem (v poměru cca 1,2:1)

Předložka *za* se pojí nejčastěji se 4. pádem, méně často se 7. pádem (v poměru 3:1) a nejméně často se 4. pádem (v poměru cca 8:3:1)

Předložka *naproti* se má pojit pouze se 3. pádem, spojení se 2. pádem (22 dokladů) je chybné.

**3. Které tvary mají ve 3. osobě sg. a ve 3. osobě pl. přítomného času slovesa *koupat*, *vykoupat*?**

**Singulár**

Dotaz:

[tag="V..S...3P.\*"&lemma="koupat|vykoupat"]

Frekvenční seznam podle slovních tvarů, uspořádaný abecedně:

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | koupá | 7 | 0,06 |
| 2 | p / n | koupe | 119 | 0,98 |
| 3 | p / n | nekoupe | 4 | 0,03 |
| 4 | p / n | nevykoupe | 1 | 0,01 |
| 5 | p / n | vykoupá | 1 | 0,01 |
| 6 | p / n | vykoupe | 26 | 0,21 |

U obou sloves se užívají dvojtvary, avšak tvary koupe, vykoupe výrazně převládají nad tvary koupá, vykoupá.

koupe : koupá 123 : 7 (18 : 1)

vykoupe : vykoupá 27 : 1

**Plurál**

Dotaz:

[tag="V..P...3P.\*"&lemma="koupat|vykoupat"]

Frekvenční seznam podle slovních tvarů, uspořádaný abecedně:

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | koupají | 16 | 0,13 |
| 2 | p / n | koupou | 56 | 0,46 |
| 3 | p / n | nekoupaj | 1 | 0,01 |
| 4 | p / n | nekoupají | 2 | 0,02 |
| 5 | p / n | nekoupou | 6 | 0,05 |
| 6 | p / n | vykoupají | 1 | 0,01 |
| 7 | p / n | vykoupou | 11 | 0,09 |

koupou : koupají 62 : 18 (3,5 : 1) vykoupou : vykoupají 11 : 1

Rozdíl mezi výskytem tvarou *koupou* a *koupají* není tak rozdílný jako mezi dubletními tvary v sg. a plurálovými tvary slovesa vykoupat.

Všechny tvary můžeme vyhledat jedním dotazem.

Budeme postupovat stejně jako při tvorbě dvou předchozích dotazů, jen u čísla zatrhneme S a P (singulár i plurál) současně.



K tagu: [tag="V..[PS]...3.\*"]

Přidáme lemmata: [tag="V..[PS]...3.\*"&lemma="koupat|vykoupat"]

Frekvenční seznam tvarů (podle abecedy):

 / 1

(celkem: 13 položek)

Sdílet tabulku:

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | koupá | 7 | 0,06 |
| 2 | p / n | koupají | 16 | 0,13 |
| 3 | p / n | koupe | 119 | 0,98 |
| 4 | p / n | koupou | 56 | 0,46 |
| 5 | p / n | nekoupaj | 1 | 0,01 |
| 6 | p / n | nekoupají | 2 | 0,02 |
| 7 | p / n | nekoupe | 4 | 0,03 |
| 8 | p / n | nekoupou | 6 | 0,05 |
| 9 | p / n | nevykoupe | 1 | 0,01 |
| 10 | p / n | vykoupá | 1 | 0,01 |
| 11 | p / n | vykoupají | 1 | 0,01 |
| 12 | p / n | vykoupe | 26 | 0,21 |
| 13 | p / n | vykoupou | 11 | 0,09 |

**4. Zjistěte, s kterými substantiv v první pozici vpravo se pojí adjektivum *sluneční* a *slunečný*.**

**a) Vytvořte správně dotaz**.

[lemma="sluneční"][tag="N.\*"]

[lemma="slunečný"][tag="N.\*"]

**b) Kterých dvojic je více?**

**c) Zkopírujte si z tabulky vždy prvních 30 dvojic s nejvyšším výskytem a zjistěte, zda se některé ze substantiv pojí s oběma adjektivy:**

 (celkem: 308 položek)

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | sluneční paprsek | 957 | 7,86 |
| 2 | p / n | sluneční brýle | 725 | 5,95 |
| 3 | p / n | sluneční soustava | 560 | 4,6 |
| 4 | p / n | sluneční světlo | 539 | 4,42 |
| 5 | p / n | sluneční záření | 492 | 4,04 |
| 6 | p / n | sluneční svit | 384 | 3,15 |
| 7 | p / n | sluneční energie | 165 | 1,35 |
| 8 | p / n | *sluneční hodiny* | 165 | 1,35 |
| 9 | p / n | sluneční paseka | 109 | 0,9 |
| 10 | p / n | sluneční bůh | 102 | 0,84 |
| 11 | p / n | sluneční kotouč | 70 | 0,58 |
| 12 | p / n | sluneční aktivita | 64 | 0,53 |
| 13 | p / n | sluneční skvrna | 57 | 0,47 |
| 14 | p / n | sluneční vítr | 52 | 0,43 |
| 15 | p / n | sluneční záře | 51 | 0,42 |
| 16 | p / n | sluneční žár | 50 | 0,41 |
| 17 | p / n | sluneční systém | 44 | 0,36 |
| 18 | p / n | sluneční erupce | 34 | 0,28 |
| 19 | p / n | sluneční elektrárna | 32 | 0,26 |
| 20 | p / n | sluneční cyklus | 31 | 0,25 |
| 21 | p / n | sluneční den | 31 | 0,25 |
| 22 | p / n | sluneční clona | 29 | 0,24 |
| 23 | p / n | sluneční pavilon | 25 | 0,21 |
| 24 | p / n | sluneční projekt | 22 | 0,18 |
| 25 | p / n | sluneční panel | 21 | 0,17 |
| 26 | p / n | sluneční lázně | 20 | 0,16 |
| 27 | p / n | sluneční kolektor | 19 | 0,16 |
| 28 | p / n | sluneční jas | 19 | 0,16 |
| 29 | p / n | sluneční teplo | 17 | 0,14 |
| 30 | p / n | sluneční disk | 16 | 0,13 |

(celkem: 76 položek)

|  | **Filtr** | **lemma** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | slunečný den | 183 | 1,5 |
| 2 | p / n | slunečný počasí | 91 | 0,75 |
| 3 | p / n | slunečný odpoledne | 23 | 0,19 |
| 4 | p / n | slunečný ráno | 22 | 0,18 |
| 5 | p / n | slunečný místo | 14 | 0,12 |
| 6 | p / n | slunečný pobřeží | 11 | 0,09 |
| 7 | p / n | slunečný stanoviště | 10 | 0,08 |
| 8 | p / n | slunečný dopoledne | 7 | 0,06 |
| 9 | p / n | slunečný neděle | 5 | 0,04 |
| 10 | p / n | slunečný léto | 5 | 0,04 |
| 11 | p / n | slunečný sobota | 4 | 0,03 |
| 12 | p / n | slunečný strana | 4 | 0,03 |
| 13 | p / n | slunečný víkend | 4 | 0,03 |
| 14 | p / n | slunečný ráj | 3 | 0,03 |
| 15 | p / n | slunečný jas | 3 | 0,03 |
| 16 | p / n | slunečný obloha | 3 | 0,03 |
| 17 | p / n | *slunečný hodina* | 3 | 0,03 |
| 18 | p / n | slunečný jih | 3 | 0,03 |
| 19 | p / n | slunečný terasa | 3 | 0,03 |
| 20 | p / n | slunečný týden | 2 | 0,02 |
| 21 | p / n | slunečný svit | 2 | 0,02 |
| 22 | p / n | slunečný byt | 2 | 0,02 |
| 23 | p / n | slunečný pokoj | 2 | 0,02 |
| 24 | p / n | slunečný podvečer | 2 | 0,02 |
| 25 | p / n | slunečný poloha | 2 | 0,02 |
| 26 | p / n | slunečný paprsek | 2 | 0,02 |
| 27 | p / n | slunečný nebe | 2 | 0,02 |
| 28 | p / n | slunečný pláž | 2 | 0,02 |
| 29 | p / n | slunečný země | 2 | 0,02 |
| 30 | p / n | slunečný poledne | 2 | 0,02 |

S oběma adjektivy se pojí pouze slova *den* a *svit*.

Více je dvojic s adjektivem *sluneční*.

**Konkurence pasivních tvarů**

Vyhledejte pasivní tvary sloves *přihlásit*, *nahlásit* a *odsouhlasit*.

1) Dotaz spojující tag pro pasivum a řadu lemmat:

[tag="Vs.\*"&lemma="přihlásit|nahlásit|odsouhlasit"]

2) U vyhledané konkordance vytvoříme frekvenční seznam slovních tvarů, které v něm uspořádáme abecedně:

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | nahlášen | 18 | 0,15 |
| 2 | p / n | nahlášena | 13 | 0,11 |
| 3 | p / n | nahlášeni | 4 | 0,03 |
| 4 | p / n | nahlášeno | 51 | 0,42 |
| 5 | p / n | nahlášenu | 3 | 0,03 |
| 6 | p / n | nahlášeny | 7 | 0,06 |
| 7 | p / n | nepřihlášeno | 1 | 0,01 |
| 8 | p / n | nepřihlášeny | 1 | 0,01 |
| 9 | p / n | odsouhlasen | 15 | 0,12 |
| 10 | p / n | odsouhlasena | 22 | 0,18 |
| 11 | p / n | odsouhlaseno | 25 | 0,21 |
| 12 | p / n | odsouhlaseny | 14 | 0,12 |
| 13 | p / n | přihlášen | 35 | 0,29 |
| 14 | p / n | přihlášena | 15 | 0,12 |
| 15 | p / n | přihlášeni | 29 | 0,24 |
| 16 | p / n | přihlášeno | 110 | 0,9 |
| 17 | p / n | přihlášeny | 25 | 0,21 |

3) Pokud bychom chtěli přesné počty, museli bychom tvary jednotlivých sloves sečíst. Není to potřeba, protože v pasivních tvarech ani jednoho slovesa nejsou dvojtvary.

Pozor na rozdíl v alternaci souhlásky na konci kmene imperativu a souhlásky před příponou příčestí trpného u sloves odvozených od slovesa *hlásit*:

|  |  |  |  |
| --- | --- | --- | --- |
| 3. osoba sg. préz. | imperativ – není alternace s:š |  infinitiv | příčestí trpné – je alternace s:š |
| přihlás-í | přihlas (se), přihlaste (se) | přihlás-i-t | přihlášen, -a, -opřihlášeni, -y, -a |
| nahlás-í | nahlas, nahlaste | nahlás-i-t | nahlášen, -a, -onahlášeni, -y, -a |
|  | poznámka: alternace délky samohlásky v základu: á:a |  | poznámka: není alternace délky samohlásky v základu: á:a |

**Vyhledejte v korpusu SYN2020 spojení:**

město Most

město Brno

město Jihlava

město Pardubice

Po vyhledání zjistěte frekvenci **slovních tvarů**.

1) Vyřaďte z frekvenčního seznamu položky, v nichž má město základní tvar nebo tvar množného čísla

2) U zbývajících položek zjistěte, zda jsou častější případy:

* ve kterých se tvar slova *město* a tvar vlastního jména shodují v pádě;
* ve kterých se slovo *město* skloňuje a tvar vlastního jména se nemění.

**Postup:**

1. Dotaz:
* [lemma="město"][lemma="Most"]
* [lemma="město"][lemma="Brno"]
* [lemma="město"][lemma="Jihlava"]
* [lemma="město"][lemma="Pardubice"]
1. Frekvenční seznamy slovních tvarů, vyřazení výše uvedených položek.
2. „Ruční“ sečtení počtu dokladů se shodnými tvary (např. *města Prahy*) a počtu dokladů, v nichž mění tvar pouze slovo město (např. *městem Praha*).

|  | **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- |
| 1 | p / n | města mostu | 16 | 0,13 |
| ~~2~~ | ~~p / n~~ | ~~město most~~ | ~~14~~ | ~~0,12~~ |
| 3 | p / n | města most | 8 | 0,07 |
| ~~4~~ | ~~p / n~~ | ~~měst mostu~~ | ~~6~~ | ~~0,05~~ |
| ~~5~~ | ~~p / n~~ | ~~měst most~~ | ~~3~~ | ~~0,03~~ |
| 6 | p / n | městem most | 2 | 0,02 |
| 7 | p / n | městě most | 1 | 0,01 |
| ~~8~~ | ~~p / n~~ | ~~městy most~~ | ~~1~~ | ~~0,01~~ |
| 9 | p / n | městem mostem | 1 | 0,01 |

shodné tvary: 17

tvary s neskloňovaným vlastním jménem: 11

| **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- |
| 1 | p / n | města jihlavy | 40 | 0,33 |
| ~~2~~ | ~~p / n~~ | ~~město jihlava~~ | ~~10~~ | ~~0,08~~ |
| 3 | p / n | městem jihlava | 4 | 0,03 |
| 4 | p / n | města jihlava | 4 | 0,03 |
| 5 | p / n | městě jihlava | 1 | 0,01 |
| 6 | p / n | město jihlavu | 1 | 0,01 |
| shodné tvary 41 (většinou genitivi)tvary s neskloňovaným vlastním jménem: 9 |  |
|  |
|  |  | Města brna | 288 | 2,36 |
| ~~2~~ | ~~p / n~~ | ~~město brno~~ | ~~89~~ | ~~0,73~~ |
| 3 | p / n | městem brnem | 17 | 0,14 |
| 4 | p / n | městě brně | 7 | 0,06 |
| 5 | p / n | městu brnu | 7 | 0,06 |
| 6 | p / n | městem brno | 2 | 0,02 |
| ~~7~~ | ~~p / n~~ | ~~měst brna~~ | ~~1~~ | ~~0,01~~ |
| 8 | p / n | města brno | 1 | 0,01 |
| 9 | p / n | městě brno | 1 | 0,01 |

shodné tvary: 319

tvary s neskloňovaným vlastním jménem: 4

| **Filtr** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- |
| 1 | p / n | **města pardubic** | **54** | 0,44 |
| ~~2~~ | ~~p / n~~ | ~~město pardubice~~ | ~~43~~ | ~~0,35~~ |
| 3 | p / n | města pardubice | 23 | 0,19 |
| 4 | p / n | městu pardubice | 3 | 0,03 |
| 5 | p / n | městem pardubice | 2 | 0,02 |
| ~~6~~ | ~~p / n~~ | ~~města pardubicím~~ | ~~1~~ | ~~0,01~~ |
| ~~7~~ | ~~p / n~~ | ~~měst pardubice~~ | ~~1~~ | ~~0,01~~ |

shodné tvary: 55

tvary s neskloňovaným vlastním jménem: 27

Ve všech spojeních převažují tvary vlastního jména shodující se v pádě se slovem *město*.

Vyhledejte spojení *hora Sněžka*, *hora Milešovka*, *hora Praděd*, *hora Klínovec*, *hora Říp*.

Pokusíme se vyloučit z dokladu doklady s nom. sg. a s plurálovými tvary slova *město* UŽ V DOTAZU:



Frekvenční seznam vytvoříme na základě Frekvence vlastní, abychom něm mohli „manipulovat“ s tvary vlastních jmen: hora vlastní jméno



Frekvenční seznam uspořádaný abecedně:

(celkem: 10 položek)

Sdílet tabulku:

|  | **Filtr** | **word** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| 1 | p / n | horu | Milešovku | 1 | 0,01 |
| 2 | p / n | hory | Praděd | 1 | 0,01 |
| 3 | p / n | horou | Pradědem | 1 | 0,01 |
| 4 | p / n | hoře | Říp | 1 | 0,01 |
| 5 | p / n | horou | Říp | 2 | 0,02 |
| 6 | p / n | hory | Říp | 6 | 0,05 |
| ~~7~~ | ~~p / n~~ | ~~horu~~ | ~~Říp~~ | ~~17~~ | ~~0,14~~ |
| 8 | p / n | horou | Sněžkou | 1 | 0,01 |
| 9 | p / n | horu | Sněžku | 1 | 0,01 |
| 10 | p / n | hory | Sněžky | 2 | 0,02 |

K dispozici je tak málo dokladů s jinými než základními tvary, že z nich nelze vyvodit žádný závěr, spojení *hora Klínovec* v jiném než základním tvaru není doloženo vůbec.

Podobnou situaci jsme již měli i u některých tvaroslovných.

Potřebujeme **větší korpus**, který by byl **žánrově vyvážený**.

Vytvoříme si žánrově vyvážený SUBKORPUS v rámci velkého korpusu SYN v12. Žánrově vyvážený subkorpus vytvoříme sloučením korpusů SYN2000, SYN2005, SYN2010, SYN2015, SYN2020.

Postup:

1. Místo SYN2020 si nastavíme SYN v12

2. Kurzorem přejdeme k nástroji Korpusy a zvolíme Vytvořit nový subkorpus:



3. Nový subkorpus bude mít určité vlastnosti. Můžeme vytvořit:

– subkorpus textů jednoho typu: prózy



Nebo můžeme vytvořit korpus tvořený z dílčích korpusů žánrově vyvážených, tj SYN2000–SYN2020, nazveme ho Vyvážený korpus, zadáme pokyn Vytvořit subkorpus.





Důkaz, že subkorpus se vytvořil:



A teď v něm znovu vyhledáme **hory**:



A vytvoříme seznam pomocí Frekvence vlastní.

Výsledek:

|  | **Filtr** | **word** | **word** | **Freq** | **i.p.m.** |
| --- | --- | --- | --- | --- | --- |
| **1** | **p / n** | **hory** | **Milešovka** | **1** | **0** |
| **2** | **p / n** | **horu** | **Milešovku** | **2** | **0** |
| **3** | **p / n** | **hory** | **Praděd** | **6** | **0,01** |
| **4** | **p / n** | **hoře** | **Praděd** | **2** | **0** |
| ~~5~~ | ~~p / n~~ | ~~horu~~ | ~~Praděd~~ | ~~5~~ | ~~0,01~~ |
| **6** | **p / n** | **horou** | **Praděd** | **1** | **0** |
| **7** | **p / n** | **horou** | **Pradědem** | **3** | **0,01** |
| **8** | **p / n** | **hory** | **Pradědu** | **3** | **0,01** |
| ~~9~~ | ~~p / n~~ | ~~horu~~ | ~~Říp~~ | ~~62~~ | ~~0,11~~ |
| **10** | **p / n** | **horou** | **Říp** | **15** | **0,03** |
| **11** | **p / n** | **hoře** | **Říp** | **19** | **0,03** |
| **12** | **p / n** | **hory** | **Říp** | **51** | **0,09** |
| **13** | **p / n** | **horou** | **ŘÍP** | **1** | **0** |
| **14** | **p / n** | **hoře** | **Řípu** | **4** | **0,01** |
| **15** | **p / n** | **hory** | **Řípu** | **3** | **0,01** |
| **16** | **p / n** | **hoře** | **Sněžce** | **6** | **0,01** |
| **17** | **p / n** | **horou** | **Sněžkou** | **1** | **0** |
| **18** | **p / n** | **horu** | **Sněžku** | **11** | **0,02** |
| **19** | **p / n** | **hory** | **Sněžky** | **4** | **0,01** |

Srovnání poměrů výskytu variant a dublet v SYN2020 a ve vyváženém subkorpusu

**1. Srovnejte poměr výskytů variant *impuls* – *impulz*, *transfuze* – *transfúze* v SYN2020 a v nově vytvořeném vyváženém subkorpusu:**

|  |  |
| --- | --- |
| **SYN2020**  |  **vyvážený subkorpus** |
| impuls 902 – impulz 1208 **3:4** transfuze 163 – transfúze 64 **5:2** | impuls 6408 – impulz 4525 **3:2** transfuze 599 – transfúze 455 **4:3**  |

U variant *impuls – impulz* v novějších textech roste podíl varianty se Z, a naopak klesá podíl varianty se S.

U variant *transfuze – transfúze* v novějších textech klesá podíl varianty s dlouhým Ú.

**2. Srovnejte výskyt infinitivnich tvarů *péct* a *péci*, *upéct* a *upéci* v SYN2020 a v nově vytvořeném vyváženém subkorpusu:**

|  |  |
| --- | --- |
| **SYN2020**  |  **vyvážený subkorpus** |
| péct 425 – péci 36  **12:1**upéct 234 – upéci 13 **18:1**  | péct 1271 – péci 347 **4:1** (3,7:1) (píct 21)upéct 692 – upéci 129 **5:1** (upíct 17) |

V novějších textech výrazně ubylo dokladů infinitivu zakončeného na *-i*.