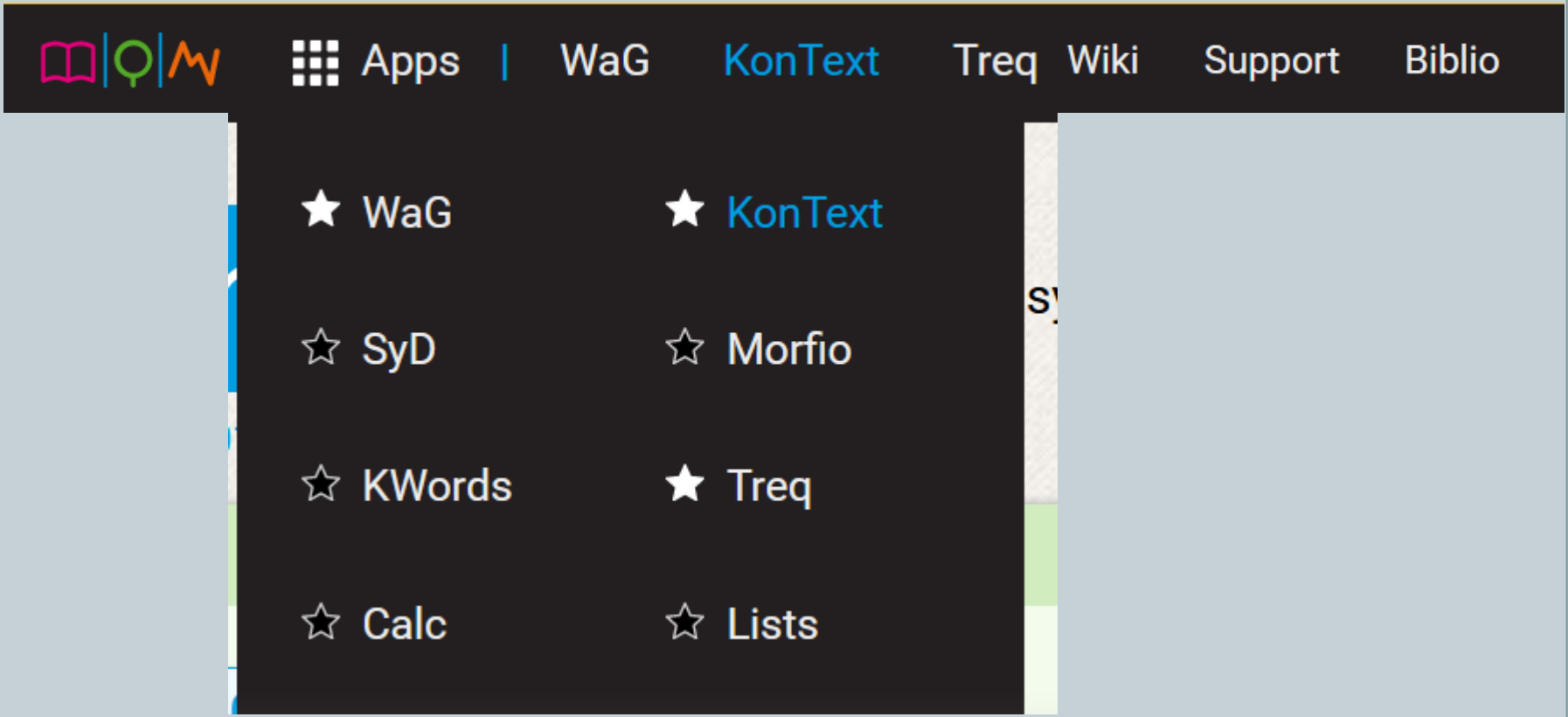
# Úvod do korpusové lingvistiky 5



KORPUSOVÉ NÁSTROJE II

# Další nástroje přístupné přes rozhraní KonText





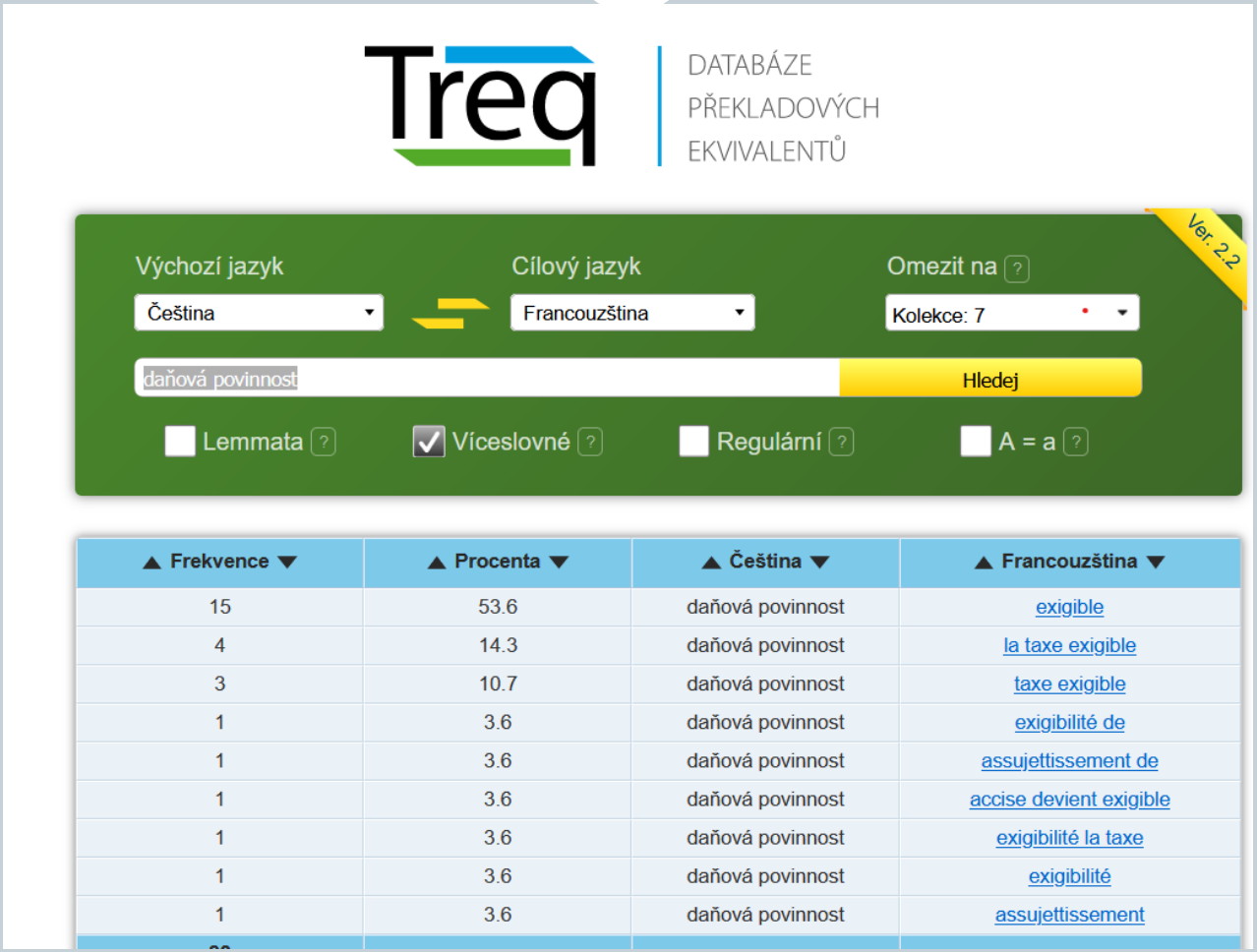
# WaG/Word at a Glance/Slovo v kostce



- Find a word *(sníh, sucho, válka)*
- Compare words (*snídaně, oběd, večeře/ Německo, Rusko/ chytrý, hloupý*)
- Search in two languages / translate *(host, opilec)*

# Treq

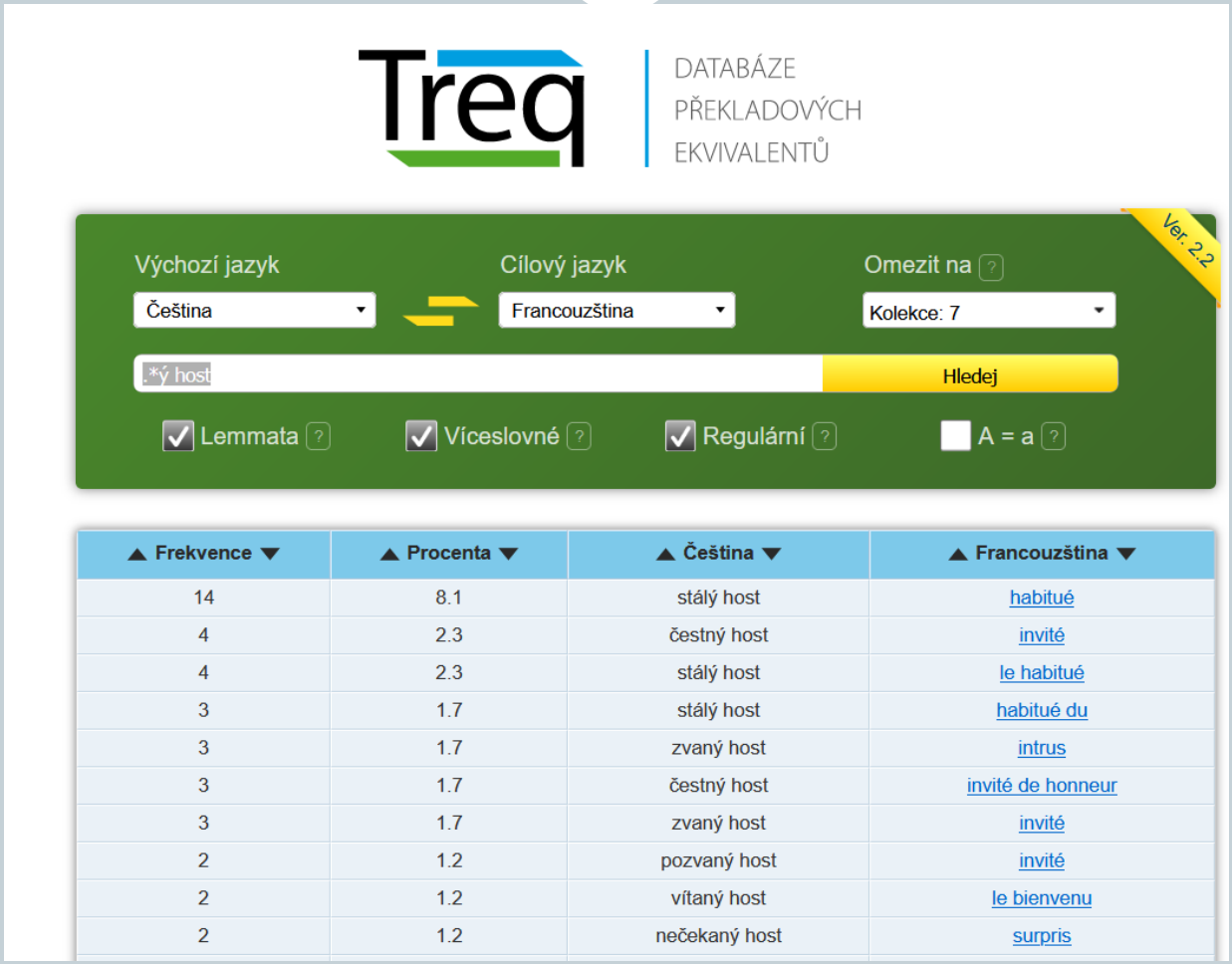





| ▲ Frekvence ▼ | ▲ Procenta ▼ | ▲ Čeština ▼ | ▲ Francouzština ▼ |
|---|---|---|---|
| 15 | 53.6 | daňová povinnost | exigible |
| 4 | 14.3 | daňová povinnost | la taxe exigible |
| 3 | 10.7 | daňová povinnost | taxe exigible |
| 1 | 3.6 | daňová povinnost | exigibilité de |
| 1 | 3.6 | daňová povinnost | assujettissement de |
| 1 | 3.6 | daňová povinnost | accise devient exigible |
| 1 | 3.6 | daňová povinnost | exigibilité la taxe |
| 1 | 3.6 | daňová povinnost | exigibilité |
| 1 | 3.6 | daňová povinnost | assujettissement |

# Jaký *host* ?






| ▲ Frekvence ▼ | ▲ Procenta ▼ | ▲ Čeština ▼ | ▲ Francouzština ▼ |
|---|---|---|---|
| 14 | 8.1 | stálý host | habitué |
| 4 | 2.3 | čestný host | invité |
| 4 | 2.3 | stálý host | le habitué |
| 3 | 1.7 | stálý host | habitué du |
| 3 | 1.7 | zvaný host | intrus |
| 3 | 1.7 | čestný host | invité de honneur |
| 3 | 1.7 | zvaný host | invité |
| 2 | 1.2 | pozvaný host | invité |
| 2 | 1.2 | vítaný host | le bienvenu |
| 2 | 1.2 | nečekaný host | surpris |

# Morfio



- Hledání n-tic se společnou a odlišnou částí
- Možnost sledovat slovotvornou produktivitu
- Možnost vyhledávat varianty

# ý/ota




Morfio
Jazyk: čeština
<+ společný
odlišný
+> Morf. specifikace:
vzor 1: .+ ý přídavná jména A.*
vzor 2: .+ ota podstatná jména N.*
Přidat vzor
Korpus: SYN2005
Frekvence vyšší než: 0
Hledat: lemmata
Vyhodnotit: lemmata
A = a
Alternace
Hledat
Nové zadání
Odkaz na toto zadání: http://morfio.korpus.cz/9pLHBjCF
Nápověda

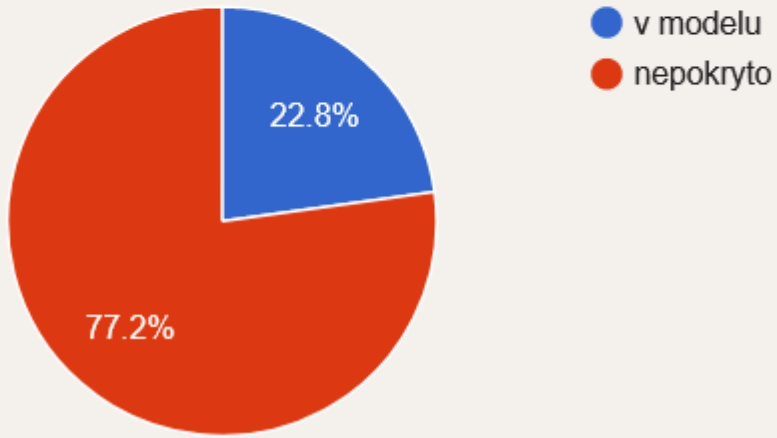
# Přegenerované výsledky



Souhrn | Výpis | Produktivita | vzor 1 | vzor 2

| Souhrn | celkem tokenů | typů | v modelu tokenů | typů |
|---|---|---|---|---|
| vzor 1 | 8471877 | 74252 | 655690 | 89 |
| vzor 2 | 115198 | 390 | 84722 | 89 |

**Odhad úplnosti daného modelu**




Souhrn | Výpis | Produktivita | vzor 1

| | vzor 1 (fq) | vzor 2 (fq) |
|---|---|---|
| 1 | hodný (5887) | hodnota (30888) |
| 2 | teplý (8320) | teplota (10804) |
| 3 | bý (3) | bota (7958) |
| 4 | jistý (30680) | jistota (7758) |
| 5 | samý (39) | samota (3726) |
| 6 | jedný (3) | jednota (3277) |
| 7 | nejistý (2) | nejistota (2822) |
| 8 | čistý (14480) | čistota (2195) |
| 9 | temný (6584) | temnota (1900) |
| 10 | ochý (1) | ochota (1687) |
| 11 | ný (32) | nota (1512) |
| 12 | hustý (5090) | hustota (1504) |
| 13 | prázdný (10074) | prázdnota (1476) |
| 14 | dobrý (103080) | dobrota (1170) |
| 15 | rý (1) | rota (1160) |
| 16 | nahý (4383) | nahota (532) |
| 17 | lhý (2) | lhota (527) |
| 18 | prostý (6758) | prostota (438) |
| 19 | nový (138367) | novota (390) |
| 20 | slepý (3573) | slepota (385) |

# Rychle je patrný rozdíl mezi morfem a pouhým řetězcem (homonymie)




| abc/cba | fq |
|---|---|
| hodnota | 30888 |
| sobota | 12048 |
| teplota | 10804 |
| bota | 7958 |
| jistota | 7758 |
| hmota | 6702 |
| samota | 3726 |
| jednota | 3277 |
| nejistota | 2822 |
| čistota | 2195 |
| porota | 2116 |
| temnota | 1900 |
| ochota | 1687 |
| nota | 1512 |
| hustota | 1504 |
| prázdnota | 1476 |
| dobrota | 1170 |
| rota | 1160 |
| pěchota | 1033 |
| toyota | 956 |
| nečistota | 821 |
| kapota | 658 |
| anekdota | 608 |
| neochota | 588 |

# Varianty



Morfio

# Podle frekvence vzor1/vzor2

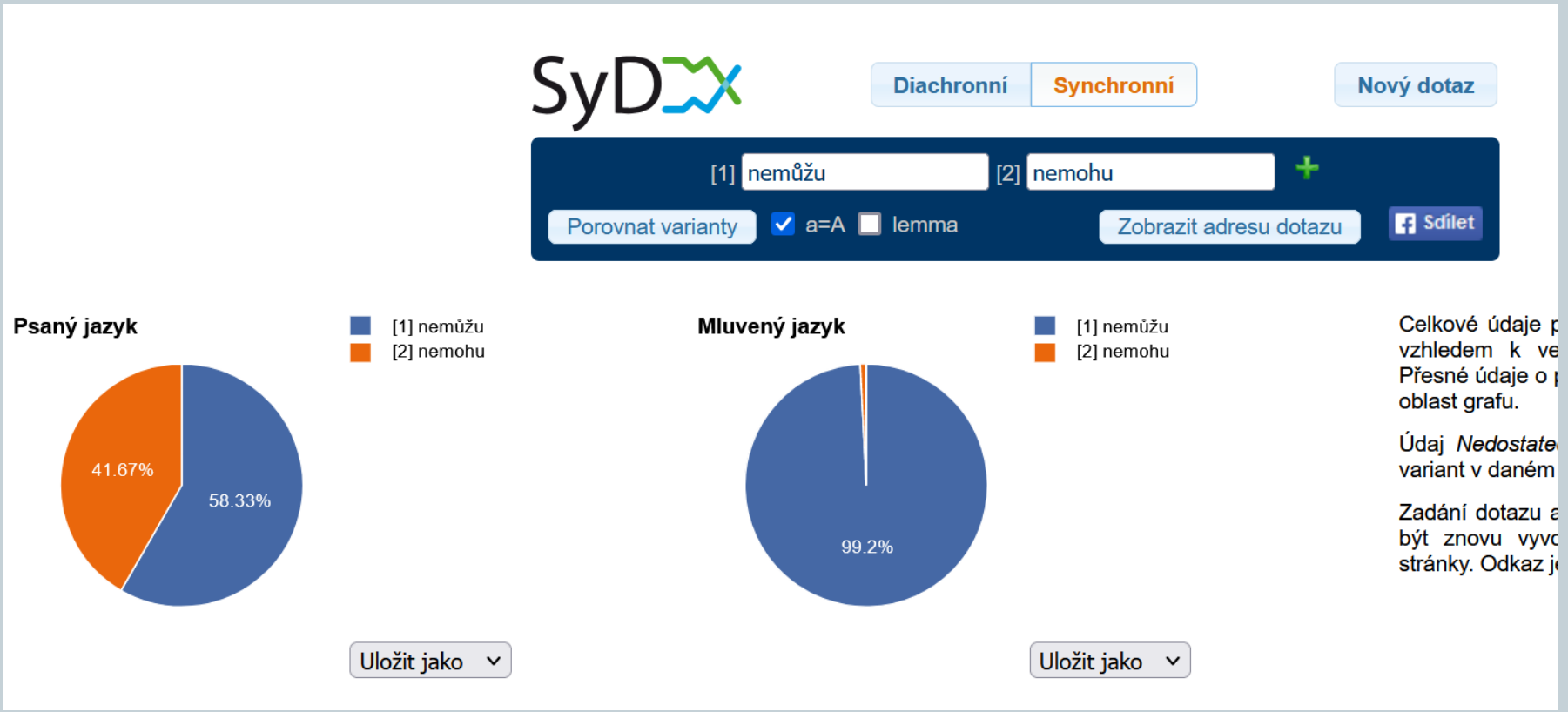


| | vzor 1 (fq) | vzor 2 (fq) |
|---|---|---|
| 1 | muži (11980) | mužové (471) |
| 2 | zástupci (3608) | zástupcové (8) |
| 3 | hasiči (2335) | hasičové (1) |
| 4 | psi (2272) | psové (15) |
| 5 | páni (1778) | pánové (2286) |
| 6 | francouzi (1327) | francouzové (18) |
| 7 | čtenáři (1250) | čtenářové (1) |
| 8 | hoši (1226) | hošové (1) |
| 9 | tvůrci (1154) | tvůrcové (14) |
| 10 | zloději (1067) | zlodějové (4) |
| 11 | soudci (954) | soudcové (104) |
| 12 | jedinci (942) | jedincové (1) |
| 13 | odpůrci (846) | odpůrcové (2) |
| 14 | milenci (786) | milencové (1) |
| 15 | samci (537) | samcové (1) |
| 16 | strážci (521) | strážcové (51) |
| 17 | domorodci (513) | domorodcové (1) |
| 18 | zastánci (506) | zastáncové (1) |
| 19 | vůdci (505) | vůdcové (167) |
| 20 | šamani (463) | šamanové (9) |

| | vzor 1 (fq) | vzor 2 (fq) |
|---|---|---|
| 1 | členi (4) | členové (6519) |
| 2 | páni (1778) | pánové (2286) |
| 3 | rusi (10) | rusové (1912) |
| 4 | briti (2) | britové (1603) |
| 5 | syni (4) | synové (1152) |
| 6 | otci (289) | otcové (675) |
| 7 | romi (1) | romové (662) |
| 8 | králi (78) | králové (518) |
| 9 | vítězi (86) | vítězové (513) |
| 10 | muži (11980) | mužové (471) |
| 11 | biskupi (3) | biskupové (456) |
| 12 | arabi (8) | arabové (383) |
| 13 | švédi (17) | švédové (355) |
| 14 | dáni (3) | dánové (295) |
| 15 | elfi (1) | elfové (258) |
| 16 | indi (5) | indové (255) |
| 17 | vůdci (505) | vůdcové (167) |
| 18 | posli (11) | poslové (147) |
| 19 | ďábli (11) | ďáblové (127) |
| 20 | lumpi (5) | lumpové (117) |

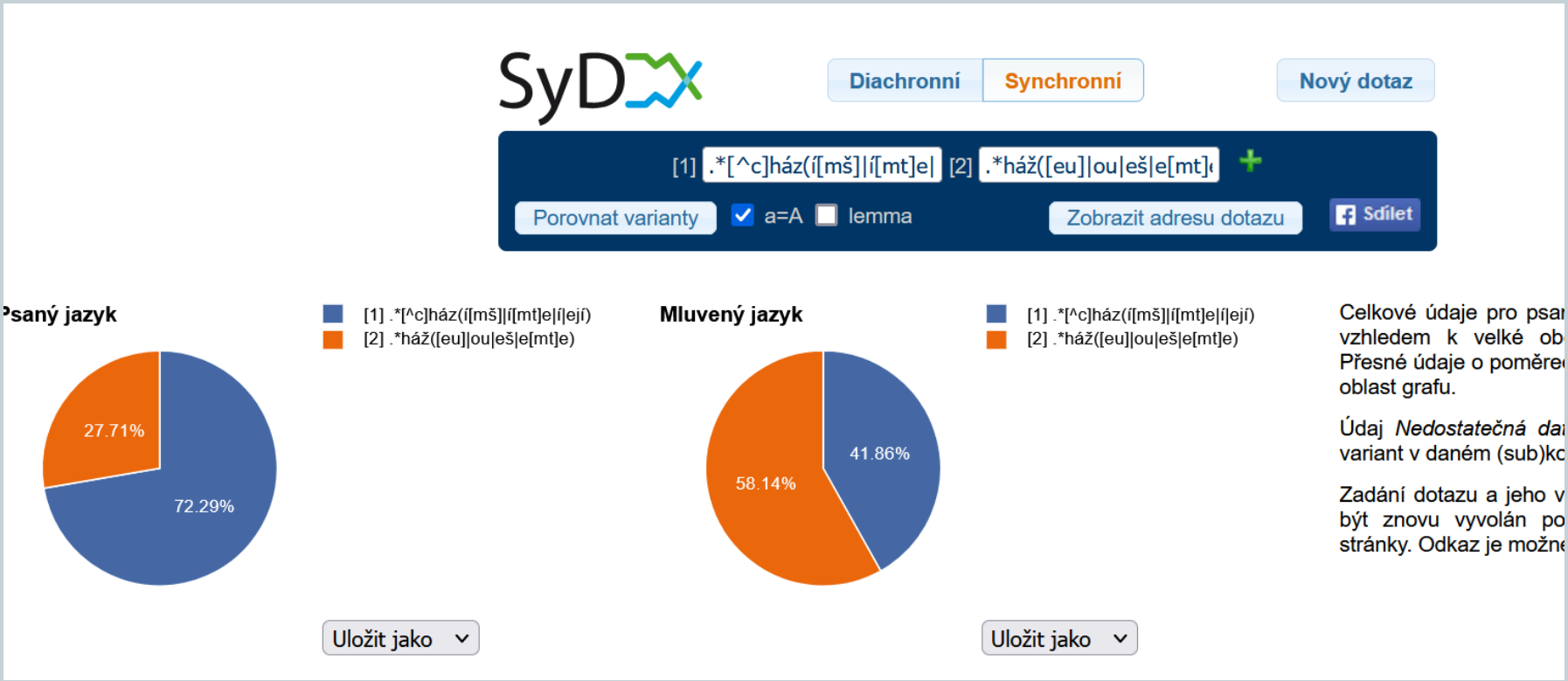
# SyD




SyD
Diachronní
Synchronní
Nový dotaz
[1] nemůžu
[2] nemohu
Porovnat varianty
a=A
lemma
Zobrazit adresu dotazu
Sdílet
Psaný jazyk
[1] nemůžu
[2] nemohu
41.67%
58.33%
Uložit jako
Mluvený jazyk
[1] nemůžu
[2] nemohu
99.2%
Uložit jako
Celkové údaje p
vzhledem k ve
Přesné údaje o p
oblast grafu.
Údaj Nedostate
variant v daném
Zadání dotazu a
být znovu vyvo
stránky. Odkaz je

# házíš/házeš




SyDaX
Diachronní
Synchronní
Nový dotaz
[1] .*[^c]ház(í[mš]|í[mt]e|
[2] .*háž([eu]|ou|eš|e[mt]
Porovnat varianty
a=A
lemma
Zobrazit adresu dotazu
Sdílet
saný jazyk
[1] .*[^c]ház(í[mš]|í[mt]e|í|ejí)
[2] .*háž([eu]|ou|eš|e[mt]e)
27.71%
72.29%
Uložit jako
Mluvený jazyk
[1] .*[^c]ház(í[mš]|í[mt]e|í|ejí)
[2] .*háž([eu]|ou|eš|e[mt]e)
41.86%
58.14%
Uložit jako
Celkové údaje pro psa
vzhledem k velké ob
Přesné údaje o poměre
oblast grafu.
Údaj Nedostatečná da
variant v daném (sub)ko
Zadání dotazu a jeho v
být znovu vyvolán po
stránky. Odkaz je možn

# Ukázka otázek v testu



- Lze ve všech nástrojích, které nabízí rozhraní Kontext, pracovat se zadáním dotazu pomocí regulárních výrazů?
- Vyber si jeden z nástrojů nabízených rozhraním KonText, který jsi vyzkoušel, a popiš úkol, který tě zaujal.
- Které z nástrojů nabízených rozhraním KonText nabízejí data čerpaná z paralelního korpusu?
- Jaká jsou omezení funkcí, které porovnávají data z psaných a mluvených korpusů?