

Úvod do korpusové lingvistiky

11



KORPUSY A KVANTITATIVNÍ DATA

Frekvenční seznam

– slovních tvarů – slov – lemmat – pos – tag



- Rozdíl mezi kvantitativní a kvalitativní analýzou korpusu spočívá v tom, že kvantitativní data přirozeně čerpatelná z korpusových textů nejsou součástí lingvistických rysů, které se datům přiřazují.
- Jsou pouze bází pro analýzu, která musí pokračovat dále. V kvalitativně zaměřené analýze jsou málo frekventované jevy zkoumány se stejnou pozorností jako jevy silně frekventované.
- Cílem analýzy korpusu není konstatování obvyklých a řídkých jevů v jazyce, nýbrž detailní popis jazyka jako celku.

příklad : adjektivum *rudý*



- kvantitativní analýza odhalí pouze počet výskytů
- kvalitativní analýza se zaměří na významy užití (nejen barva, ale i např. politická příslušnost atd.)

Reprezentativnost korpusu



- Ve velmi malých korpusech je možné, že se okrajové jevy vůbec nevyskytnou a frekventované jevy nebudou zastoupeny dostatečně.
- Velký korpus zaručuje možnost dobře zkoumat frekventované jevy.
- Nicméně platí, že výskyt hapaxových jevů je stabilní (zvětšujeme-li rozsah textů, neklesá podíl – kvantitativní i kvalitativní – tzv. hapax legomena, ale i dalších hapaxových jevů).

Vzorky



- Výběr vzorku – neexistuje obecně platná metoda, jak určit reprezentativnost vzorku.
- Podobné metody výběru vzorku zaručí, že data z korpusů zpracovaných co do výběru vzorků stejnými metodami budou srovnatelná navzájem.

Zpracování dat kvantitativními metodami



- V korpusové lingvistice jde kvantitativní analýza ruku v ruce s analýzou kvalitativní.
- Běžně užívané techniky matematické statistiky, které v rámci KL následují za prostým počítáním frekvenčních výskytů jazykových jevů obsažených v korpusu.
- Díky těmto metodám se lingvisté snaží získat z korpusů nejen prostá kvantitativní data, ale dojít k interpretaci jejich závažnosti, a to pomocí exaktních matematicky ověřených postupů.

metody matematické statistiky užívané v KL



- Jsou to např. metody, při jejichž užití je možné brát zřetel na takové okolnosti, jako je typ okolí jednotky (kolokace), vzorku (žánr) atd.
- Přehled je pouze omezený (nejsem matematický statistik a úvod do mat. stat. není cílem naší přednášky).
- (více: Statistics for Corpus Linguistics v řadě edinburských učebnic empirické lingvistiky, internet).

Frekvenční analýza



- matematické sečítání počtu jednotek (tokens)
- v případě klasifikovaných jednotek typů (lemmat, tagů, pos, ...)
- u anotovaných korpusů obecně můžeme počítat a) se snazší prací a b) s lepšími výsledky
- u anotovaných korpusů je třeba mít na zřeteli, že počítáme pouze výsledky anotací, nikoliv to, co skutečně v korpusu je

Proporcionalita



- Prosté počítání frekvencí je jen jako první krok další analýzy.
- Hlavní nevýhodou prostých frekvenčních výpočtů je, že výsledky, které jimi získáme, se mohou značně lišit v případě, kdy jeden a týž jev spočítáme v různých korpusech (např. psaném a mluveném).
- Jak získané výsledky porovnat?
- Výsledky ze dvou korpusů, které nejsou stejně velké: vypočítáme frekvenci jako procento z celkového počtu tokens v korpusu. Výsledek srovnání procentuálního zastoupení nám může říci něco spolehlivého.

Porovnání výskytu tokenů v korpusech různého rozsahu



- Např. zjistíme, že jev A se v korpusu psaného jazyka o 1 mil. slovních tvarů vyskytuje 500 krát a v korpusu mluveného jazyka o 100 000 slovních tvarů 50 krát.

Vypočítáme percentuální výskyt, a to takto:



- mluvený korpus $(50: 100\ 000) \times 100 = 0,05\%$
- psaný korpus $(500: 1000\ 000) \times 100 = 0,05\%$
- V obou případech nám vyjde stejný výsledek.
- Vypočítali jsme, že s ohledem na různost proporcí vzorků je frekvence stejná.
- Vždy se vychází z poměru mezi velikostí vzorku a počtem výskytů.
- $\text{ratio} = \text{počet výskytů typů} / \text{počet výskytů tokens}$ v celém vzorku

Testování významnosti výsledků frekvenčních analýz



- chi-square test
- MI-score
- T-score
- z- score

Kolokace



- Metody se používají pro vyhledávání statisticky významných kolokací.
- Kolokace (souvýskyt slov) jsou z lingvistického hlediska zajímavé.
- Gramatika
- Lexikon – idiomatika
- MWE

červený kolokace



Kolokace

Atribut: lemma ▾ V rozsahu od: 1 do: 1

Minimální frekvence v korpusu: 5

Minimální frekvence v daném rozsahu: 3

Zobrazit funkce:

- T-score
- MI
- MI3
- log likelihood
- min. citlivost
- logDice
- MI.log_f
- relativní frekvence

Seřadit dle:

- T-score
- MI
- MI3
- log likelihood
- min. citlivost
- logDice
- MI.log_f
- relativní frekvence

Vytvořit seznam

červený

frekvence a MI-score



	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n ,	1145	4.454	0.204
2. p/n a	777	13.931	0.999
3. p/n .	742	-5.934	-0.284
4. p/n víno	548	23.343	8.458
5. p/n kříž	527	22.932	9.855
6. p/n barva	384	19.463	7.206
7. p/n karta	282	16.704	7.566
8. p/n světlo	220	14.626	6.164
9. p/n krvinka	154	12.406	11.756
10. p/n koberec	149	12.171	8.416
11. p/n Kostelec	133	11.524	10.420
12. p/n cihla	132	11.464	8.855
13. p/n skvrna	125	11.144	8.282
14. p/n královna	125	11.120	7.532
15. p/n nebo	118	8.174	2.014
16. p/n trpaslík	111	10.517	9.165
17. p/n paprika	106	10.277	9.124
18. p/n oko	96	8.916	3.473
19. p/n být	91	-42.905	-2.459
20. p/n řepa	87	9.323	11.002
21. p/n květ	85	9.118	6.511
22. p/n šátek	76	8.695	8.554
23. p/n knihovna	74	8.470	6.021
24. p/n)	74	2.656	0.533
25. p/n nit	73	8.529	9.186

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n karkulka	32	5.655	12.015
2. p/n chitón	4	2.000	11.969
3. p/n krvinka	154	12.406	11.756
4. p/n Bedrník	3	1.732	11.746
5. p/n vejložek	3	1.732	11.746
6. p/n studebaker	3	1.732	11.746
7. p/n Řečice	29	5.383	11.637
8. p/n Pečky	37	6.081	11.534
9. p/n chitónu	9	2.999	11.384
10. p/n malvice	6	2.448	11.161
11. p/n řepa	87	9.323	11.002
12. p/n Karkulka	35	5.913	10.889
13. p/n TZ	54	7.344	10.752
14. p/n kapie	11	3.315	10.728
15. p/n kartounový	3	1.731	10.647
16. p/n bedrník	4	1.999	10.576
17. p/n lampas	5	2.235	10.536
18. p/n cínie	3	1.731	10.466
19. p/n nervatura	3	1.731	10.466
20. p/n sedma	39	6.241	10.440
21. p/n Kostelec	133	11.524	10.420
22. p/n kv	3	1.731	10.384
23. p/n Haut	3	1.731	10.306
24. p/n Hrádek	59	7.675	10.282
25. p/n fez	8	2.826	10.241

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n víno	548	23.343	8.458
2. p/n kříž	527	22.932	9.855
3. p/n barva	384	19.463	7.206
4. p/n karta	282	16.704	7.566
5. p/n světlo	220	14.626	6.164
6. p/n a	777	13.931	0.999
7. p/n krvinka	154	12.406	11.756
8. p/n koberec	149	12.171	8.416
9. p/n Kostelec	133	11.524	10.420
10. p/n cihla	132	11.464	8.855
11. p/n skvrna	125	11.144	8.282
12. p/n královna	125	11.120	7.532
13. p/n trpaslík	111	10.517	9.165
14. p/n paprika	106	10.277	9.124
15. p/n řepa	87	9.323	11.002
16. p/n květ	85	9.118	6.511
17. p/n oko	96	8.916	3.473
18. p/n šátek	76	8.695	8.554
19. p/n nit	73	8.529	9.186
20. p/n knihovna	74	8.470	6.021
21. p/n šaty	70	8.226	5.897
22. p/n nebo	118	8.174	2.014
23. p/n číslo	65	7.676	4.382
24. p/n Hrádek	59	7.675	10.282
25. p/n značka	60	7.553	5.329

rudý



		<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1.	p/n armáda	733	27.052	10.292
2.	p/n právo	318	17.728	7.419
3.	p/n ,	238	-10.426	-0.745
4.	p/n a	192	2.605	0.300
5.	p/n .	160	-16.008	-1.180
6.	p/n moře	151	12.227	7.650
7.	p/n náměstí	117	10.756	7.467
8.	p/n skvrna	110	10.473	9.416
9.	p/n růže	106	10.279	9.241
10.	p/n hvězda	97	9.773	7.014
11.	p/n vlas	77	8.665	6.323
12.	p/n barva	76	8.598	6.187
13.	p/n jako	69	5.888	1.780
14.	p/n záře	66	8.114	9.599
15.	p/n Khmer	65	8.062	14.100
16.	p/n planeta	64	7.955	7.468
17.	p/n vlajka	54	7.333	8.917
18.	p/n v	51	-9.305	-1.203
19.	p/n drak	51	7.119	8.294
20.	p/n oko	49	6.505	3.821
21.	p/n "	49	-11.066	-1.368
22.	p/n světlo	44	6.448	5.160
23.	p/n ret	44	6.567	6.638
24.	p/n prapor	40	6.314	9.162
25.	p/n garda	35	5.908	9.604

		<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1.	p/n Khmer	65	8.062	14.100
2.	p/n Khmér	4	2.000	13.965
3.	p/n zamžení	8	2.828	13.827
4.	p/n Sonja	8	2.827	11.242
5.	p/n karafiát	22	4.688	10.965
6.	p/n mžitka	7	2.644	10.702
7.	p/n posuv	25	4.997	10.659
8.	p/n plyš	7	2.644	10.618
9.	p/n gardista	13	3.603	10.301
10.	p/n armáda	733	27.052	10.292
11.	p/n Ural	14	3.739	10.236
12.	p/n stříkanec	3	1.730	10.039
13.	p/n hlinka	4	1.998	9.894
14.	p/n přísvit	5	2.234	9.840
15.	p/n rtěnka	27	5.190	9.735
16.	p/n baň	9	2.996	9.715
17.	p/n samet	19	4.353	9.616
18.	p/n garda	35	5.908	9.604
19.	p/n záře	66	8.114	9.599
20.	p/n mustang	4	1.997	9.545
21.	p/n DVDMAG	3	1.730	9.532
22.	p/n lemování	3	1.730	9.497
23.	p/n zář	10	3.158	9.459
24.	p/n skvrna	110	10.473	9.416
25.	p/n cákanec	3	1.729	9.317

		<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1.	p/n armáda	733	27.052	10.292
2.	p/n právo	318	17.728	7.419
3.	p/n moře	151	12.227	7.650
4.	p/n náměstí	117	10.756	7.467
5.	p/n skvrna	110	10.473	9.416
6.	p/n růže	106	10.279	9.241
7.	p/n hvězda	97	9.773	7.014
8.	p/n vlas	77	8.665	6.323
9.	p/n barva	76	8.598	6.187
10.	p/n záře	66	8.114	9.599
11.	p/n Khmer	65	8.062	14.100
12.	p/n planeta	64	7.955	7.468
13.	p/n vlajka	54	7.333	8.917
14.	p/n drak	51	7.119	8.294
15.	p/n ret	44	6.567	6.638
16.	p/n oko	49	6.505	3.821
17.	p/n světlo	44	6.448	5.160
18.	p/n prapor	40	6.314	9.162
19.	p/n garda	35	5.908	9.604
20.	p/n jako	69	5.888	1.780
21.	p/n dres	34	5.807	7.947
22.	p/n květ	33	5.679	6.464
23.	p/n ďábel	32	5.637	8.135
24.	p/n šátek	28	5.276	8.431
25.	p/n koberec	28	5.258	7.322

???



	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n ,	417	4.932	0.399
2. p/n a	269	8.863	1.121
3. p/n oko	257	15.860	6.546
4. p/n .	177	-8.303	-0.700
5. p/n oblek	121	10.989	9.910
6. p/n zóna	105	10.222	8.681
7. p/n vlas	102	10.024	7.063
8. p/n zákal	80	8.943	12.904
9. p/n barva	79	8.795	6.577
10. p/n eminence	73	8.543	14.004
11. p/n ekonomika	54	7.308	7.496
12. p/n nebo	48	5.587	2.369
13. p/n kámen	48	6.857	6.604
14. p/n kůra	46	6.769	8.964
15. p/n jako	40	3.805	1.328
16. p/n mlha	40	6.301	8.061
17. p/n mrak	38	6.142	8.118
18. p/n země	34	5.268	3.374
19. p/n obloha	32	5.618	7.175
20. p/n hmota	31	5.526	7.067
21. p/n plášt'	31	5.536	7.439
22. p/n kalhoty	30	5.443	7.310
23. p/n (30	2.640	0.949
24. p/n až	28	3.978	2.010
25. p/n šaty	28	5.221	6.227

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n eminence	73	8.543	14.004
2. p/n zákal	80	8.943	12.904
3. p/n gabardénový	3	1.732	12.036
4. p/n litina	15	3.872	11.517
5. p/n flanel	3	1.731	11.206
6. p/n flanelový	13	3.603	10.685
7. p/n baloňák	4	1.999	10.644
8. p/n metalíza	4	1.999	10.392
9. p/n plstěný	8	2.826	10.220
10. p/n žaket	3	1.731	10.140
11. p/n kalhotový	5	2.234	10.098
12. p/n oblek	121	10.989	9.910
13. p/n převlečník	3	1.730	9.902
14. p/n Austin	4	1.998	9.644
15. p/n sešitek	3	1.730	9.636
16. p/n břidlicový	5	2.233	9.577
17. p/n tvídový	5	2.233	9.517
18. p/n myška	17	4.117	9.461
19. p/n kápě	6	2.445	9.246
20. p/n bradka	7	2.640	8.969
21. p/n rolák	3	1.729	8.968
22. p/n kůra	46	6.769	8.964
23. p/n chevrolet	3	1.728	8.911
24. p/n tepláky	6	2.444	8.835
25. p/n sedan	7	2.640	8.770

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n oko	257	15.860	6.546
2. p/n oblek	121	10.989	9.910
3. p/n zóna	105	10.222	8.681
4. p/n vlas	102	10.024	7.063
5. p/n zákal	80	8.943	12.904
6. p/n a	269	8.863	1.121
7. p/n barva	79	8.795	6.577
8. p/n eminence	73	8.543	14.004
9. p/n ekonomika	54	7.308	7.496
10. p/n kámen	48	6.857	6.604
11. p/n kůra	46	6.769	8.964
12. p/n mlha	40	6.301	8.061
13. p/n mrak	38	6.142	8.118
14. p/n obloha	32	5.618	7.175
15. p/n nebo	48	5.587	2.369
16. p/n plášt'	31	5.536	7.439
17. p/n hmota	31	5.526	7.067
18. p/n kalhoty	30	5.443	7.310
19. p/n země	34	5.268	3.374
20. p/n šaty	28	5.221	6.227
21. p/n vlk	26	5.083	8.281
22. p/n ,	417	4.932	0.399
23. p/n odstín	23	4.772	7.628
24. p/n stín	23	4.717	5.919
25. p/n nebe	20	4.392	5.806

//intext.konze.cz/view?q=~nD6NlyId5...nanz.com

Odhadněte kolokáty adjektiva

růžový



růžový



	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n ,	404	2.695	0.208
2. p/n ·	282	-2.141	-0.173
3. p/n a	280	8.561	1.034
4. p/n zahrada	272	16.455	8.782
5. p/n barva	114	10.591	6.961
6. p/n keř	103	10.132	9.222
7. p/n květ	87	9.292	8.051
8. p/n sad	86	9.259	9.346
9. p/n nebo	81	7.858	2.979
10. p/n až	62	6.898	3.012
11. p/n panter	61	7.809	12.357
12. p/n tvář	56	7.282	5.218
13. p/n víno	56	7.410	6.674
14. p/n dům	50	6.654	4.084
15. p/n brýle	47	6.831	8.094
16. p/n lístek	40	6.279	7.117
17. p/n šaty	37	6.015	6.484
18. p/n rtěnka	33	5.740	10.214
19. p/n být	33	-24.898	-2.415
20. p/n poupě	32	5.654	11.062
21. p/n odstín	31	5.545	7.913
22. p/n jako	28	1.961	0.668
23. p/n dřevo	28	5.221	6.240
24. p/n trikot	27	5.194	11.271
25. p/n tričko	25	4.979	7.887

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n slamáček	4	2.000	12.668
2. p/n panter	61	7.809	12.357
3. p/n bahýnko	3	1.732	11.813
4. p/n palouček	8	2.828	11.668
5. p/n taft	5	2.235	11.440
6. p/n trikot	27	5.194	11.271
7. p/n slaměnka	6	2.448	11.203
8. p/n poupě	32	5.654	11.062
9. p/n pantoflíček	9	2.998	10.945
10. p/n mašlička	7	2.644	10.926
11. p/n plameňák	15	3.871	10.875
12. p/n poupátko	5	2.235	10.648
13. p/n kavalír	10	3.160	10.638
14. p/n grep	3	1.731	10.506
15. p/n svetřík	10	3.160	10.476
16. p/n rtík	3	1.731	10.417
17. p/n stínidlo	5	2.234	10.290
18. p/n šatičky	9	2.998	10.253
19. p/n mašle	12	3.461	10.215
20. p/n rtěnka	33	5.740	10.214
21. p/n vestička	4	1.998	10.049
22. p/n tvářička	20	4.468	10.046
23. p/n mikina	15	3.869	9.944
24. p/n pelargonie	4	1.998	9.921
25. p/n čumáček	5	2.234	9.915

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n zahrada	272	16.455	8.782
2. p/n barva	114	10.591	6.961
3. p/n keř	103	10.132	9.222
4. p/n květ	87	9.292	8.051
5. p/n sad	86	9.259	9.346
6. p/n a	280	8.561	1.034
7. p/n nebo	81	7.858	2.979
8. p/n panter	61	7.809	12.357
9. p/n víno	56	7.410	6.674
10. p/n tvář	56	7.282	5.218
11. p/n až	62	6.898	3.012
12. p/n brýle	47	6.831	8.094
13. p/n dům	50	6.654	4.084
14. p/n lístek	40	6.279	7.117
15. p/n šaty	37	6.015	6.484
16. p/n rtěnka	33	5.740	10.214
17. p/n poupě	32	5.654	11.062
18. p/n odstín	31	5.545	7.913
19. p/n dřevo	28	5.221	6.240
20. p/n trikot	27	5.194	11.271
21. p/n tričko	25	4.979	7.887
22. p/n ret	22	4.608	5.827
23. p/n tvářička	20	4.468	10.046
24. p/n světlo	22	4.460	4.349
25. p/n nádech	20	4.458	8.322

válka



válka czTenTen12 frekvence = 837,766 (165.25 v milionu)

Lemma	Skóre	Frekvence
boj	0.404	741,906
konflikt	0.351	184,846
krize	0.294	454,749
bitva	0.274	199,198
útok	0.274	537,701
událost	0.251	574,877
spor	0.250	259,427
akce	0.247	1,807,080
situace	0.245	1,791,853
život	0.244	3,293,700
politika	0.239	695,600
revoluce	0.232	179,016
násilí	0.232	187,259
vztah	0.223	1,255,394
vláda	0.222	1,391,430
volba	0.218	1,069,431
jednání	0.217	803,701
smrt	0.216	667,221
vývoj	0.216	935,325
armáda	0.216	366,346
aktivita	0.210	729,500
svět	0.208	2,902,894
činnost	0.208	1,287,802
léto	0.208	5,618,796
doba	0.206	5,168,774
změna	0.205	1,810,145
úřad	0.204	2,475,500

Slovní profily – word-sketch (kolokace i s ohledem na fce)



válka

czTenTen12 frekvence = [837,766](#) (165.25 v milionu)

a_modifier	445,516	-0.80	adv_modifier	462,626	-0.50	prec_prep	172,508	-0.50	gen_2
světový	205,241	11.64	světový	204,655	11.62	během	8,581	7.99	vypuknutí
třicetiletý	16,543	10.04	třicetiletý	16,090	9.95	po	52,267	7.48	skončení
studený	19,220	9.72	studený	18,722	9.65	koncem	1,281	6.96	Studená
občanský	28,115	9.48	občanský	27,632	9.44	před	11,446	6.58	vyhlášení
husitský	7,731	8.98	husitský	7,409	8.87	proti	3,992	6.18	hrůza
hvězdný	6,654	8.54	hvězdný	6,643	8.49	za	22,408	5.99	rozpoutání
napoleonský	3,852	8.10	napoleonský	3,763	8.01	uprostřed	416	5.28	sklonek
vietnamský	2,721	7.46	vietnamský	2,688	7.39	kvůli	882	5.03	konec
žabomyší	1,753	6.99	žabomyší	1,710	6.90	začátkem	145	4.15	oběť
korejský	1,871	6.93	korejský	1,805	6.83	do	12,023	4.11	ukončení
dívčí	1,758	6.77	dívčí	1,735	6.71	následkem	99	4.09	nesmyslnost
krvavý	1,735	6.66	válka_roku	1,361	6.59	mezi	1,428	4.05	vypovězení
prohraný	1,480	6.64	svatý	3,179	6.58	ohledně	160	3.96	počátek
svatý	3,261	6.64	šestidenní	1,347	6.55	bez	1,606	3.91	Klonová
šestidenní	1,373	6.64	španělská_občanská_válka	1,329	6.55	počátkem	83	3.54	začátek
sedmiletý	1,386	6.58	sedmiletý	1,285	6.42	z	7,163	3.54	bůh
punský	1,136	6.37	prohraný	1,224	6.32	o	6,231	3.52	pokraj
prusko-rakouský	1,114	6.35	punský	1,114	6.29	vůči	148	3.50	krutost
zákopový	1,109	6.34	vlastenecký	1,148	6.25	díky	270	3.38	veterán
vlastenecký	1,168	6.32	zákopový	1,065	6.23	v	22,936	3.33	odpůrce
			partyzánský	1,061	6.18	k	5,227	3.32	důsledek



<u>post_v</u>	<u>46,715</u>	<u>-0.80</u>	<u>is_obj4_of</u>	<u>33,598</u>	<u>-0.30</u>	<u>prec_v</u>	<u>23,593</u>	<u>-0.40</u>	<u>prec_verb</u>	<u>21,606</u>	<u>-0.30</u>
Irák	<u>7,804</u>	10.86	rozpoutat	<u>1,919</u>	9.91	přerůst	<u>448</u>	8.24	zuřit	<u>1,203</u>	9.72
Vietnam	<u>3,582</u>	10.53	vyhlásit	<u>3,475</u>	8.83	vyústit	<u>289</u>	7.26	vypuknout	<u>626</u>	8.20
záliv	<u>3,185</u>	9.72	prohrát	<u>956</u>	7.53	zahynout	<u>159</u>	6.17	rozpoutat	<u>346</u>	7.73
Afgánistán	<u>2,078</u>	9.31	vyhlašovat	<u>614</u>	7.45	zmítat	<u>86</u>	6.13	vyhlašovat	<u>497</u>	7.31
Jugoslávie	<u>875</u>	8.58	přežít	<u>1,179</u>	7.37	bojovat	<u>877</u>	5.77	svatat	<u>250</u>	7.22
Bosna	<u>677</u>	8.26	přečkat	<u>229</u>	7.13	spojenec	<u>153</u>	5.60	rozpoutávat	<u>98</u>	7.02
Čečensko	<u>481</u>	8.20	vyprovokovat	<u>209</u>	6.85	hrdinství	<u>49</u>	5.47	vypukat	<u>73</u>	6.70
Libye	<u>444</u>	7.60	vypovědět	<u>242</u>	6.83	vítězství	<u>779</u>	5.46	propukat	<u>67</u>	6.31
Libanon	<u>352</u>	7.56	vyhrát	<u>2,160</u>	6.79	zvrhnout	<u>46</u>	5.46	prohrávat	<u>114</u>	6.06
Korea	<u>574</u>	7.46	ukončit	<u>1,094</u>	6.68	porážka	<u>260</u>	5.42	vyhlásit	<u>427</u>	5.88
Pacifik	<u>297</u>	7.35	vyvolat	<u>947</u>	6.56	přerůstat	<u>42</u>	5.35	zmítat	<u>65</u>	5.80
Gruzie	<u>314</u>	7.14	vést	<u>5,466</u>	6.48	vojín	<u>44</u>	5.34	konvenčnět	<u>68</u>	5.77
Španělsko	<u>591</u>	7.04	zažít	<u>602</u>	6.21	padnout	<u>377</u>	5.33	vést	<u>3,059</u>	5.65
Tichomoří	<u>167</u>	6.80	prožít	<u>440</u>	6.18	zvrát	<u>71</u>	4.98	prusko-rakouskat	<u>33</u>	5.64
Gaze	<u>178</u>	6.80	prohrávat	<u>134</u>	6.01	obrat	<u>204</u>	4.95	propuknout	<u>62</u>	5.54
Afgánistán	<u>125</u>	6.39	rozpoutávat	<u>63</u>	5.81	angažovat	<u>65</u>	4.94	rozhořet	<u>35</u>	5.27
Afganistán	<u>116</u>	6.28	zahájit	<u>806</u>	5.78	umírat	<u>129</u>	4.81	sužovat	<u>38</u>	4.81
Kongo	<u>137</u>	6.13	odvrátit	<u>107</u>	5.45	ocitnout	<u>115</u>	4.79	hrozit	<u>385</u>	4.80
Afghanistán	<u>103</u>	6.11	ospravedlnit	<u>54</u>	5.28	zvítězit	<u>185</u>	4.76	vyhrávat	<u>77</u>	4.55
Indočína	<u>101</u>	6.10	zavinit	<u>61</u>	5.21	voják	<u>287</u>	4.48	odsuzovat	<u>54</u>	4.49
Afrika	<u>326</u>	6.02	riskovat	<u>117</u>	5.00	zemřít	<u>202</u>	4.48			



<u>prec_po</u>	<u>19,596</u>	<u>-3.10</u>	<u>coord</u>	<u>18,315</u>	<u>-0.10</u>	<u>gen_1</u>	<u>16,746</u>	<u>-0.10</u>	<u>post_s</u>	<u>15,935</u>	<u>-0.60</u>
těsna	<u>194</u>	6.55	hladomor	<u>161</u>	7.51	Roseová	<u>152</u>	8.17	mlok	<u>394</u>	9.13
Němec	<u>277</u>	5.47	okupace	<u>509</u>	7.40	gang	<u>500</u>	8.08	Prusko	<u>278</u>	8.50
léto	<u>1,845</u>	4.73	mír	<u>725</u>	7.04	ninj	<u>57</u>	6.36	terorismus	<u>657</u>	8.35
nedlouho	<u>16</u>	4.58	násilí	<u>774</u>	6.94	sever	<u>627</u>	6.33	Dominion	<u>114</u>	7.81
desetiletí	<u>64</u>	4.19	holocaust	<u>151</u>	6.85	Rakousko-Uhersk	<u>33</u>	5.85	Irák	<u>663</u>	7.67
Beneš	<u>29</u>	4.13	nepokoj	<u>215</u>	6.78	Ares	<u>39</u>	5.76	Írán	<u>344</u>	7.51
Československo	<u>57</u>	3.80	zbrojení	<u>90</u>	6.78	nerv	<u>233</u>	5.76	turek	<u>267</u>	7.38
Sudety	<u>12</u>	3.68	genocida	<u>126</u>	6.54	Áres	<u>28</u>	5.75	teror	<u>185</u>	7.00
toužit	<u>49</u>	3.61	fašismus	<u>117</u>	6.52	nacista	<u>83</u>	5.66	napoleon	<u>72</u>	6.69
Německo	<u>150</u>	3.61	militarismus	<u>52</u>	6.41	policaajt	<u>84</u>	5.64	Gruzie	<u>128</u>	6.64
monarchie	<u>20</u>	3.54	terorismus	<u>167</u>	6.33	Jom	<u>24</u>	5.44	Iráno	<u>45</u>	6.46
Burian	<u>11</u>	3.53	krveprolití	<u>54</u>	6.21	pohlaví	<u>158</u>	5.43	Hizballáh	<u>54</u>	6.36
nacista	<u>20</u>	3.53	revoluce	<u>532</u>	6.21	hněv	<u>72</u>	5.42	Irán	<u>54</u>	6.30
zaniknout	<u>25</u>	3.52	teror	<u>97</u>	6.00	diadoch	<u>22</u>	5.41	Rusko	<u>450</u>	6.27
pohraničí	<u>16</u>	3.52	bída	<u>126</u>	5.93	růže	<u>97</u>	5.29	Japonsko	<u>224</u>	6.02
emigrovat	<u>12</u>	3.45	protektorát	<u>68</u>	5.91	Švédová	<u>21</u>	5.26	Peršan	<u>39</u>	5.99
příměří	<u>12</u>	3.41	konflikt	<u>480</u>	5.87	džihád	<u>29</u>	5.23	Prus	<u>38</u>	5.99
komunista	<u>57</u>	3.37	zabíjení	<u>94</u>	5.83	Rakousku-Uhersek	<u>19</u>	5.17	Voldemort	<u>56</u>	5.91
Stalin	<u>13</u>	3.33	mor	<u>81</u>	5.81	neválek	<u>17</u>	5.05	Izrael	<u>251</u>	5.79
Libanon	<u>10</u>	3.27	holocaust	<u>42</u>	5.77	Rusová	<u>19</u>	5.01	Persie	<u>40</u>	5.79
civilista	<u>15</u>	3.25	rozbroj	<u>36</u>	5.70	klan	<u>66</u>	4.93	Uher	<u>26</u>	5.54
odsun	<u>14</u>	3.22	epidemie	<u>109</u>	5.66	terorismus	<u>53</u>	4.71	Francie	<u>248</u>	5.54
baťa	<u>11</u>	3.14	hlad	<u>136</u>	5.60	kopí	<u>20</u>	4.65	Taliban	<u>37</u>	5.48
Irák	<u>29</u>	3.11	agrese	<u>74</u>	5.58	klon	<u>31</u>	4.54	Turecko	<u>109</u>	5.38
zbyt	<u>38</u>	3.11	nacismus	<u>60</u>	5.52				Kartágo	<u>25</u>	5.37

<u>post_proti</u>	<u>13,312</u>	<u>-8.20</u>	<u>prec_do</u>	<u>10,979</u>	<u>-0.60</u>	<u>prec_z</u>	<u>9,668</u>	<u>-0.50</u>	<u>post_na</u>	<u>8,057</u>
terorismus	<u>2,964</u>	10.59	narukovat	<u>137</u>	8.23	veterán	<u>445</u>	7.32	Balkán	<u>543</u>
teror	<u>701</u>	9.01	zaplést	<u>148</u>	7.66	vzpamatovávat	<u>72</u>	7.08	Kavkaz	<u>105</u>
Irák	<u>1,145</u>	8.50	zatáhnout	<u>235</u>	7.14	Vader	<u>24</u>	6.13	gehenna	<u>23</u>
terorizmus	<u>68</u>	7.18	zavléct	<u>40</u>	6.36	trauma	<u>84</u>	6.01	bojiště	<u>74</u>
Írán	<u>166</u>	6.52	vtáhnout	<u>98</u>	6.28	vysloužilec	<u>23</u>	5.98	východ	<u>557</u>
turek	<u>132</u>	6.44	vstoupit	<u>850</u>	6.21	pomník	<u>128</u>	5.60	spadnutí	<u>20</u>
droga	<u>516</u>	6.43	zabřednout	<u>29</u>	6.06	navrátilce	<u>25</u>	5.50	fronta	<u>392</u>
Libye	<u>100</u>	6.30	rukovat	<u>21</u>	5.87	výjev	<u>56</u>	5.43	balkán	<u>10</u>
Saddám	<u>65</u>	6.28	povolat	<u>89</u>	5.83	spolubojovník	<u>21</u>	5.38	poloostrov	<u>73</u>
Jugoslávie	<u>80</u>	5.98	uvrhnout	<u>36</u>	5.69	zajatec	<u>49</u>	5.27	maloměsto	<u>14</u>
Napoleon	<u>39</u>	5.93	vehnat	<u>24</u>	5.63	Vadero	<u>11</u>	5.15	kontinent	<u>92</u>
islám	<u>151</u>	5.78	narukování	<u>14</u>	5.36	munice	<u>40</u>	4.91	Sicílie	<u>15</u>
Srbsko	<u>70</u>	5.63	odvelet	<u>16</u>	5.35	patrona	<u>19</u>	4.87	Srí	<u>20</u>
Hizballáh	<u>25</u>	5.43	zapojit	<u>288</u>	5.14	Jed	<u>11</u>	4.87	Ukrajina	<u>35</u>
okupant	<u>35</u>	5.38	příplést	<u>20</u>	5.10	vzpamatovat	<u>33</u>	4.75	hřbitov	<u>69</u>
Taliban	<u>29</u>	5.27	naverbovat	<u>15</u>	5.07	strach	<u>237</u>	4.69	území	<u>481</u>
Hitler	<u>67</u>	5.15	nahnat	<u>24</u>	4.95	bunkr	<u>23</u>	4.57	západ	<u>115</u>
nacismus	<u>40</u>	5.11	verbovat	<u>12</u>	4.83	bojiště	<u>28</u>	4.56	náves	<u>14</u>
fašismus	<u>38</u>	5.08	vmanipulovat	<u>11</u>	4.82	zotavovat	<u>11</u>	4.48	přelom	<u>53</u>
Izrael	<u>148</u>	5.05	přerůst	<u>30</u>	4.77	návrat	<u>311</u>	4.41	sever	<u>79</u>
Parth	<u>14</u>	5.05	vstupovat	<u>159</u>	4.74	zápiska	<u>13</u>	4.38	pevnina	<u>10</u>
nemuslimus	<u>14</u>	5.04	zaplétat	<u>12</u>	4.72	obava	<u>150</u>	4.30	Slovensko	<u>62</u>
Japonsko	<u>110</u>	5.02	zatažení	<u>11</u>	4.56	invalida	<u>10</u>	4.26	jih	<u>63</u>

válka



	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n .	4859	44.857	1.488
2. p/n ,	4705	40.808	1.304
3. p/n a	1668	22.599	1.163
4. p/n v	1660	26.968	1.565
5. p/n být	1578	15.583	0.719
6. p/n se	915	6.264	0.335
7. p/n s	722	18.522	1.687
8. p/n proti	469	20.984	5.010
9. p/n na	362	-4.039	-0.278
10. p/n "	322	-15.732	-0.908
11. p/n -	310	9.791	1.172
12. p/n mezi	284	15.204	3.354
13. p/n ?	265	9.331	1.228
14. p/n o	209	4.392	0.522
15. p/n skončit	196	13.533	4.907
16. p/n (195	7.258	1.058
17. p/n mít	148	-2.230	-0.243
18. p/n jako	137	3.503	0.513
19. p/n)	136	3.254	0.472
20. p/n :	135	4.088	0.626
21. p/n z	120	-7.985	-0.790
22. p/n za	119	2.138	0.315
23. p/n i	113	-1.567	-0.198
24. p/n nebo	107	4.931	0.934
25. p/n	102	5.343	1.086

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n neválka	5	2.236	12.030
2. p/n rekvírovat	3	1.730	10.030
3. p/n Roseová	6	2.447	9.860
4. p/n prusko-rakouský	5	2.232	9.067
5. p/n zmítaný	15	3.862	8.406
6. p/n vybombardovat	3	1.724	7.808
7. p/n 1866	11	3.300	7.644
8. p/n sužovaný	6	2.437	7.582
9. p/n dezertovat	4	1.988	7.358
10. p/n area	4	1.987	7.302
11. p/n rozvrácený	5	2.221	7.223
12. p/n prohibice	3	1.720	7.172
13. p/n zpustošený	5	2.219	7.012
14. p/n věznit	12	3.428	6.593
15. p/n galský	3	1.714	6.571
16. p/n zničený	31	5.498	6.328
17. p/n vypuknout	31	5.491	6.177
18. p/n poznamenaný	9	2.958	6.158
19. p/n zajmout	12	3.416	6.156
20. p/n okupovat	7	2.609	6.154
21. p/n deportovat	3	1.704	5.971
22. p/n čaroděj	12	3.401	5.779
23. p/n gang	14	3.667	5.649
24. p/n vyznamenat	5	2.188	5.541
25. p/n vyhasnout	3	1.695	5.538

	<u>Frekvence</u>	<u>T-score</u>	<u>MI</u>
1. p/n .	4859	44.857	1.488
2. p/n ,	4705	40.808	1.304
3. p/n v	1660	26.968	1.565
4. p/n a	1668	22.599	1.163
5. p/n proti	469	20.984	5.010
6. p/n s	722	18.522	1.687
7. p/n být	1578	15.583	0.719
8. p/n mezi	284	15.204	3.354
9. p/n skončit	196	13.533	4.907
10. p/n -	310	9.791	1.172
11. p/n ?	265	9.331	1.228
12. p/n vyhrát	63	7.444	4.009
13. p/n (195	7.258	1.058
14. p/n začít	90	6.954	1.905
15. p/n bojovat	51	6.772	4.273
16. p/n sloužit	51	6.646	3.849
17. p/n se	915	6.264	0.335
18. p/n ;	77	6.257	1.801
19. p/n pokračovat	47	5.884	2.819
20. p/n zničený	31	5.498	6.328
21. p/n vypuknout	31	5.491	6.177
22. p/n přežít	33	5.425	4.167
23. p/n !	102	5.343	1.086
24. p/n sever	29	5.116	4.324
25. p/n pracovat	40	5.001	2.257