



CZECH NATIONAL
CORPUS

Introduction to Text Corpora and Their Applications

Basic notions of corpus linguistics

Lucie Chlumská, Ph.D.

lucie.chlumska@korporus.cz





OUTLINE:

1. LECTURE

- representativeness
- concordance and context
- frequency and frequency distribution
- collocation and colligation

2. SEMINAR

- reading (Stubbs): Lexical semantics (meaning is use, frequency)
- how is meaning constituted? why is frequency important?





LECTURE





Key notions of corpus linguistics



Representativeness

- corpora aim to be **representative** > of what?
- general corpora: to fully capture the variability of language
 - written v. spoken language > proportion?
 - text types, genres to be included?
 - samples v. full texts?
- how to choose texts into a corpus? surveys?
 - production v. reception
- the larger the corpus is, the more likely it reflects prototypicality
 - centre v. periphery
- representativeness is **relative** > character of a scale...



Concordance

- **KWIC** = key word in context
 - not only a word or phrase, but also a string of characters (e.g. suffix)
- **concordances** = lines of context with the keyword centered
 - basic form of display in a corpus
 - can be sorted and analyzed
 - KWIC view v. sentence view
- **concordancer** = a computer program that can produce a concordance from a specified text or corpus
 - corpus alone solves few (if any) problems for a linguist > its potential is unlocked by tools such as concordancers



Concordance: KWIC view

Výběr řádků: základní ▼

1	<input type="checkbox"/>	Written books and periodicals	it ; or receive it . They like cats especially , if seems , because	cats	are apparently good strong dreamers . But …' The doctor rose slowly to his
2	<input type="checkbox"/>	Written books and periodicals	that . Like his compatriot Nelson Piquet , Emerson in the cockpit is a somnolent	cat	.
3	<input type="checkbox"/>	Written books and periodicals	This occupation has had little influence on its size . The North African will	cat	weight's only slightly more than the average mcggie .
4	<input type="checkbox"/>	Written books and periodicals	softly up the stairs and into our room with less noise than a pair of	cat	burglars . I open the curtains to let the night in . But there's
5	<input type="checkbox"/>	Written miscellaneous	was always black and white ? Remember how Your parents drew black lines Around the	cats	and mice ? Remember how As you grew to your teens Grey shades revealed themselves
6	<input type="checkbox"/>	Spoken demographic	n't long started . It did n't start till ten past ten . Who 's	cat	? So He said do you want a pint ? I said , no I
7	<input type="checkbox"/>	Spoken demographic	? Mm . Oh . Well I was n't rolling on the floor , the	cat	was . The cat , aye . Well see a black one still that we
8	<input type="checkbox"/>	Written books and periodicals	cut o'f my kitty 's privates . The People headline Player involved in allegor	cat	castration was a Leeds reserve .
9	<input type="checkbox"/>	Written books and periodicals	. 'With a smile , Doyle said , ' More ways of killing a	cat	…' But Cow ey , with a sharp comment , wiped the smile from Doyle
10	<input type="checkbox"/>	Written books and periodicals	to Florida last month for a fortnight 's training camp .	cat	's hearing ? The cat 's ears are much more sensitive than those of its
11	<input type="checkbox"/>	Written books and periodicals	are well-illustrated and helpful . FEARS OF A CLCWN OVERPOPULATION , NOISE , CARS AND	CATS	ALL Trouble SPIKE MILLIGAN . HE PUT HIS PERSONAL PROGRAMME FOR THE ENVIRONMENT TO TREVOR
12	<input type="checkbox"/>	Written books and periodicals	transcriptional level in Hep33 cells .	CAT	assays <p>The cDNA clone for C/EBP^δ in a eukaryotic expression vector , under the control
13	<input type="checkbox"/>	Written books and periodicals	: Julia Raymond , sales manager	Cats	in Art , see picture) , address books , notebooks , carrier bags ,
14	<input type="checkbox"/>	Written books and periodicals	are normally extremely efficient at regulating the amount of food they take in . These	cats	are far less common than these dogs (or these people) So if
15	<input type="checkbox"/>	Spoken demographic	, she 's , she 's typical of the Now now ! of the eight	cats	Oh I know she is Mm living on your own syndrome . So
16	<input type="checkbox"/>	Written books and periodicals	Commander . 'Bel began to laugh . ' Help ? You ? ' The	cat	jumped down and walked across the room . ' How can you help ? Why
17	<input type="checkbox"/>	Written books and periodicals	shall call you Peach , she thought . The trembling had stopped , though the	cat	did n't yet purr . It occurred to her that the way she was holding
18	<input type="checkbox"/>	Spoken demographic	, well I 've heard this before I know I 've noticed is that the	cat	meowing then ? tough luck if you want to go out , right so I
19	<input type="checkbox"/>	Written books and periodicals	the sandpit , ' said Emma . ' We could write a notice to say	cats	and dogs are not allowed in the sandpit . ' Sarah designed some swings :
20	<input type="checkbox"/>	Written books and periodicals	the woman , ' and should I try dangling a fish there to get the	cat	out , too ? ' Doctor ! Doctor ! I 've just swallowed a bluebottle
21	<input type="checkbox"/>	Written books and periodicals	over India to view his work . With his young assistant and a much-loved black	cat	we sat on the floor of his studio , sipped mint tea and listened to
22	<input type="checkbox"/>	Written books and periodicals	the thylacine ('marsupial ' wolf ') . Much later , the Europeans brought	cats	to Australia (which probably jumped ship in the seventeenth century) , and then
23	<input type="checkbox"/>	Written books and periodicals	within sentences ; in 3 we have seen how knowledge of the world (about	cats	and gates , for example) or of the culture (about social roles and relationships
24	<input type="checkbox"/>	Written miscellaneous	dog food .	cat	food products selling extremely well in the major multiples .
25	<input type="checkbox"/>	Written books and periodicals	va ves were already bent through test driving . ' Luckily , Jaguar had two other Big	Cats	on standby to show their paces .
26	<input type="checkbox"/>	Spoken demographic	she sat down Oh of course you did , mummy said no Does n't smell like	cat	wee It looks quite comfortable Have you got a carrier stuff in Yeah there 's
27	<input type="checkbox"/>	Written books and periodicals	physique . I churned out orgasms to order , squawking and spluttering like Sylvester the	Cat	on acid . Karen was too desperate to notice . By now the soggy British spring
28	<input type="checkbox"/>	Written books and periodicals	stay at my house and look after her . I know many other dog and	cat	owner's who feel the same .
29	<input type="checkbox"/>	Written books and periodicals	The bigger the flock , the smaller the chance that you end up as the	cat	's next victim . So , as the size of the flock increases individual birds



Concordance: sentence view

Výběr řádků: základní ▼

- 1 *Writer books and periodicals* They like cats especially . it seems , because **cats** are apparently good strong dreamers .
- 2 *Writer books and periodicals* Like his compatriot Nelson Piquet , Emerson in the cockpit is a somnolent **cat** .</p>
- 3 *Writer books and periodicals* The North African wild **cat** weighs only slightly more than the average moggie .</p>
- 4 *Writer books and periodicals* We are softly up the stairs and into our room with less noise than a pair of **cat** burglars .
- 5 *Writer miscellaneous* Around the **cats** and mice ?
- 6 *Spoken demographic* Who's **cat** ?
- 7 *Spoken demographic* Well I was n't rolling on the floor , the **cat** was .
- 8 *Writer books and periodicals* Player involved in alleged **cat** castration was a Leeds reserve .</p>
- 9 *Writer books and periodicals* <p>With a smile , Doyle said , ' More ways of killing a **cat** …' </p>
- 10 *Writer books and periodicals* How sensitive is the **cat**'s hearing ?
- 11 *Writer books and periodicals* OVERPROLIFERATION , NOISE , CARS AND **CATS** All Trouble SPIKE MILLIGAN
- 12 *Writer books and periodicals* Recombinant Plasmids , transient transfections and **CAT** assays
- 13 *Writer books and periodicals* <p>Fine art illustrated record books , diaries (including **Cats** in Art , see picture) , address books , notebooks , carrier tags , giftwrap , photograph frames and albums , desk and home accessories NEC International Spring Fair : hall 6 , stand H1</p>
- 14 *Writer books and periodicals* Obese **cats** are far less common than obese dogs (or obese people) .
- 15 *Spoken demographic* of the eight **cats**
- 16 *Writer books and periodicals* The **cat** jumped down and walked across the room .
- 17 *Writer books and periodicals* The trembling had stopped , though the **cat** did not yet purr .
- 18 *Spoken demographic* is that the **cat** meowing then ? tough luck if you want to go out , right so I've got this , butter , sugar , flour eggs , salt , lemon juice , I have n't got any
- 19 *Writer books and periodicals* ' We could write a notice to say **cats** and dogs are not allowed in the sandpit . '
- 20 *Writer books and periodicals* <p>All right , ' said the woman , ' and should I try catching a fish there to get the **cat** out , too ? ' </p>
- 21 *Writer books and periodicals* With his young assistant and a much-loved black **cat** we sat on the floor of his studio , sipped mint tea and listened to his words relating to the ancient science .</p>
- 22 *Writer books and periodicals* Much later , the Europeans brought **cats** to Australia (which probably jumped ship in the seventeenth century) , and then foxes and rabbits (for sport) in the nineteenth century -- and these have out-competed and preyed upon a whole host of marsupials .
- 23 *Writer books and periodicals* In 2 we have looked at how the formal system operates between as well as within sentences ; in 3 we have seen how knowledge of the world (about **cats** and gates , for example) or of the culture (about social roles and relationships) enables people to make their language function as they intend and to understand how others do the same to them .
- 24 *Writer miscellaneous* <p>Our private label business in the supermarkets continues to grow with our **cat** food products selling extremely well in the major multiples .</p>
- 25 *Writer books and periodicals* <p>Luckily , Jaguar had two other Big **Cats** on standby to show their paces .</p>



Concordance: sorting (1L)



Dotaz Korpusy Uložit Konkordance Filtr Frekvence Kolokace Zobrazení Nápověda

Korpus: bnc | Dotaz: cat (5 398 výskyty) | Promíchat: ✓ | Třídění: ✓ (5 398 výskyty) | Třídění: ✓ (5 398 výskyty)

Výskytů: 5 398 | i.p.m. 0,48,52 (vztaženo k celému "bnc") | ARF 0,1 551,8 | Výsledek je seřazen

22 / 180

Přejít na

Výběr řádků: základní

631	Written books and periodicals	against Christians .	Why do we say that someone is grinning like a Cheshire	Cat	?<p>In Lewis Carroll 's Alice in Wonderland we encounter a large cat lying
632	Written books and periodicals	customs , music , literature , and so on : What do you call a half Cheshire		Cat	? Yul Grinner </p><p>How do ghosts like to travel ? By British Scareways </p>
633	Written books and periodicals	wood , and after a while she came to a little house .	The Cheshire	Cat	<p>There was a boy outside the door , with a large letter in his hand
634	Written books and periodicals	as mad as he was in March .	Suddenly , there was the Cheshire	Cat	again , sitting in another tree . Alice jumped in surprise .</p><p> Do you
635	Written books and periodicals	if not a celebration of special effects from white roses being painted red to Cheshire		Cats	dismantling before your eyes ?</p><p>And apart from the giggles from the party frocked section of
636	Written books and periodicals	this but says : ' That 's very nice . ' And smiles her Cheshire		cat	grin .</p><p> I get a little far mail . Men have been known to
637	Written books and periodicals	Cod bless you .	School 's out for Dahl 's BFG Tracey Kandohla<p>THE Cheshire	Cat	has vanquished the Big Friendly Giant in the classroom battleground of English literature .</p><p>Education
638	Written books and periodicals	have created such expectations . But the foundations for this belief were like the Cheshire		Cat	in Alice of Wonderland ; the more one looked at them , the less substance
639	Written books and periodicals	.</p><p>She was thinking about pigs and children when she suddenly saw the Cheshire		Cat	in a tree . The cat grinned at her and she went nearer to it
640	Written books and periodicals	the Toraja 's attentions , and the splendid Werner Meyer , looking like the Cheshire		Cat	in safari costume . There was also the party of representatives from the Indonesian President
641	Written miscellaneous	side of things i.e. proverbs , as well as the nonsensical side of things . The Cheshire		cat	is an odd character and he causes confusion when he literally disappears into thin air
642	Written books and periodicals	It was distinctly confusing . The Tory vote began to seem like the Cheshire		Cat	that Alice came across in her travels . ' It vanished quite slowly , beginning
643	Written books and periodicals	gives them a get-out option but the CPI and Marjorie Roberts are grinning like Cheshire		cats	to have extracted such a promise . Of course , they may have to write next to
644	Written books and periodicals	of leftover paint . See the diagram over the page for full details .	Christmas	CATS	IAN LUCAS could 'nt resist expanding his catfish collection Catfish Tank – Ian Lucas IAN
645	Written books and periodicals	advice in my column will answer all your questions .	TROPICAL INFORMATION Up and Coming	CATS	MICHAEL RODSON can't resist the new catfish that appear on the market . Here
646	Written books and periodicals	more of them are taking their place on that very same stage .	In Cool	Cats	: 25 Years Of Rock 'N' Roll Style , published in the early '80s .
647	Written miscellaneous	also decorate by dipping half in chocolate .	Competitions<p>Let's face it ! The Cool	Cats	Club is a pretty childish name , eh ! So we've decided to have
648	Written miscellaneous	enjoyed her stay with you .	Find enclosed the member 's card for the Cool	cats	club and two newsletters ! hope you enjoy reading !</p><p>I have joined a drama
649	Written books and periodicals	was a case of genetic dwarfism </p><p>Among the pedigree breeds , the Maine Coon		cat	from North America is one of the biggest , with some exceptional individuals weighing as
650	Written books and periodicals	wind . Higher in the sky , seagulls hover opportunistically with patronising casualness . Cringing		cats	desperately hug buildings for cover , dealing with their inherited<CORR>agoraphobia</CORR>in the best way
651	Written-to-be-spoken	move on to a progressive prison ; to a Co cat , or even D		cat	</p><p>Male speaker<p>But it 's still the system . No matter where you go , there
652	Written-to-be-spoken	move on to a progressive prison ; to a Co cat , or even D		cat	</p><p>Male speaker<p>But it 's still the system . No matter where you go , there
653	Written books and periodicals	have drowned them : for four days searched barn and hedge and cornfield . Damn		cat	's getting sneaky –got kittens somewhere , hidden . Well , I 'm finding homes
654	Written books and periodicals	<p>Local councillor Michael Wigley said since the village was included in the blacklist of Darlington		Cats	Protection League , many residents had come forward to protest .</p><p>The league said it
655	Written miscellaneous	frosted glass back door : a good , clean , slightly sweet smell .	Definitely	cat	. A shape moved beyond the glass and then disappeared .</p><p>I waited but did
656	Written miscellaneous	and unpasteurised goat 's milk products . Unwashed , uncooked fruit and vegetables . Dirty		cat	litter . Soil contaminated with cat faeces .<p>If you love cats you can still
657	Written books and periodicals	and building contractor gangs who charge inflated prices .	Pet subject<p>THE Newtown and District	Cats	Protection League is to raise £120 as a matter of urgency to pay for a

Tokens and types

- **token**= any single, particular instance of an individual word in a text or a corpus
- **type** = a single particular wordform; one type may occur many times in a corpus (many tokens)

Count types and tokens in the following sentence:

There is a cat at the window; the cat is watching birds.

how many tokens (running words)?

how many types (different words)?

- Type/token ratio (TTR) as a basic indicator of repetition



Tokens and types

- **token**= any single, particular instance of an individual word in a text or a corpus
- **type** = a single particular wordform; one type may occur many times in a corpus (many tokens)

Count types and tokens in the following sentence:

There is a cat at the window; the cat is watching birds.

how many tokens (running words)? 14

how many types (different words)? 11 (*the, cat, is*)

- Type/token ratio (TTR) as a very basic indicator of repetition



Frequency

- **frequency of occurrence** = the most basic statistics we can provide for a word or pattern

raw frequency (to be interpreted within one text/corpus)

normalized frequency – **ipm/ppm**, pmw/ptw (to compare texts/corpora)

- **frequency lists** = very useful indicators of core vocabulary

crucial for L2 teaching and learning

top frequency words in any language: function words

the most frequent words in English?

different for written and spoken language



Top frequency words (written v. spoken)

	<u>Filtr</u>	<u>word</u>	<u>Frekvence</u>
1.	p/n	the	4 525 933
2.	p/n	of	2 456 681
3.	p/n	and	2 045 595
4.	p/n	to	2 016 743
5.	p/n	a	1 654 614
6.	p/n	in	1 412 212
7.	p/n	is	767 638
8.	p/n	was	750 744
9.	p/n	that	749 455
10.	p/n	for	643 264
11.	p/n	it	549 477
12.	p/n	on	534 481
13.	p/n	The	532 354
14.	p/n	's	529 068
15.	p/n	with	523 293
16.	p/n	be	518 809
17.	p/n	I	516 733
18.	p/n	by	418 536
19.	p/n	as	412 256
20.	p/n	he	375 273
21.	p/n	had	373 892
22.	p/n	are	356 062
23.	p/n	at	353 059
24.	p/n	not	351 078
25.	p/n	his	349 635
26.	p/n	have	335 294
27.	p/n	from	328 561
28.	p/n	which	306 126
29.	p/n	you	303 020
30.	p/n	or	288 198

	<u>Filtr</u>	<u>word</u>	<u>Frekvence</u>
1.	p/n	I	168 001
2.	p/n	you	117 520
3.	p/n	it	112 297
4.	p/n	the	110 186
5.	p/n	's	107 730
6.	p/n	n't	77 481
7.	p/n	and	75 121
8.	p/n	that	71 614
9.	p/n	to	70 776
10.	p/n	a	70 232
11.	p/n	do	47 846
12.	p/n	Yeah	43 469
13.	p/n	in	43 059
14.	p/n	he	40 079
15.	p/n	they	37 588
16.	p/n	of	34 989
17.	p/n	was	33 874
18.	p/n	on	32 466
19.	p/n	Oh	31 923
20.	p/n	know	31 614
21.	p/n	have	31 245
22.	p/n	is	30 786
23.	p/n	got	30 216
24.	p/n	we	28 573
25.	p/n	she	27 597
26.	p/n	there	27 087
27.	p/n	one	24 074
28.	p/n	what	24 061
29.	p/n	No	23 719
30.	p/n	've	23 267



Distribution (dispersion)

- **frequency distribution** = shows how a word or phrase is distributed in the corpus

high frequency due to overuse in one text v. even distribution

(*křižník* ‘cruiser’ v. *nedorozumění* ‘misunderstanding’)

indicated by no. of texts or ARF (average reduced frequency)

- examples of uneven distribution

Table 2.10 Distribution of a selection of adjectives over subcorpora in COCA: per million words

	<i>Spoken</i>	<i>Fiction</i>	<i>Magazines</i>	<i>Newspapers</i>	<i>Academic</i>	<i>Total</i>
appalling	3.5	3.1	3.3	3.5	2.3	3.2
bizarre	17.3	12.2	13.4	11.0	6.5	12.3
conceptual	0.7	0.9	5.0	2.7	50.4	11.9
purple	7.2	42.1	27.2	15.0	5.8	19.7



Collocation

- **collocation** = a co-occurrence relationship between two words; words are said to collocate with one another in one is more likely to occur in the presence of the other than elsewhere
 - association measures
 - t-score, MI-score, LogDice etc.
 - no measure is perfect
- **colligation** = a co-occurrence between a word and a grammatical category or context

What collocation is on a lexical level of analysis, colligation is on a syntactic level. The term does not refer to the repeated combination of concrete word forms but to the way in which word classes co-occur or keep habitual company in an utterance.

Ute Römer



Association measures (-1 cat)

logDice

MI-score (mutual information)

t-score (absolute frequency)

	Filtr	Frekvence	T-score	MI	logDice
1.	p/n Cheshire	40	6.319	10.079	7.944
2.	p/n pet	31	5.561	9.764	7.586
3.	p/n wild	49	6.976	8.206	7.545
4.	p/n domestic	53	7.249	7.882	7.428
5.	p/n Siamese	19	4.358	12.594	7.125
6.	p/n tabby	17	4.123	12.912	6.972
7.	p/n stray	16	3.996	9.956	6.774
8.	p/n Stray	12	3.464	13.263	6.477
9.	p/n black	58	7.528	6.436	6.474
10.	p/n tom	11	3.316	11.921	6.339
11.	p/n Manx	10	3.161	11.399	6.194
12.	p/n pedigree	10	3.159	10.053	6.148
13.	p/n tortoiseshell	9	2.999	11.824	6.052
14.	p/n big	40	6.228	6.036	6.044
15.	p/n feral	9	2.999	11.060	6.038
16.	p/n fat	15	3.839	6.828	5.990
17.	p/n Big	12	3.442	7.302	5.961
18.	p/n female	17	4.068	6.213	5.771
19.	p/n purring	7	2.644	10.698	5.675
20.	p/n de-clawed	6	2.449	14.409	5.485
21.	p/n pussy	6	2.449	11.310	5.468
22.	p/n alley	6	2.444	8.693	5.371
23.	p/n Black	10	3.112	5.977	5.284
24.	p/n Tom	10	3.110	5.927	5.260
25.	p/n Peke-faced	5	2.236	14.369	5.223
26.	p/n starveling	5	2.236	13.253	5.220
27.	p/n spotted	6	2.429	6.877	5.114
28.	p/n Practical	5	2.229	8.343	5.101
29.	p/n Wild	5	2.225	7.641	5.029
30.	p/n white	20	4.330	4.975	4.997

	Filtr	Frekvence	T-score	MI	logDice
1.	p/n de-clawed	6	2.449	14.409	5.485
2.	p/n Peke-faced	5	2.236	14.369	5.223
3.	p/n Stray	12	3.464	13.263	6.477
4.	p/n starveling	5	2.236	13.253	5.220
5.	p/n tabby	17	4.123	12.912	6.972
6.	p/n Siamese	19	4.358	12.594	7.125
7.	p/n tailless	4	2.000	12.384	4.896
8.	p/n brindled	3	1.732	12.310	4.483
9.	p/n Giraffe	3	1.732	12.310	4.483
10.	p/n tom	11	3.316	11.921	6.339
11.	p/n tortoiseshell	9	2.999	11.824	6.052
12.	p/n Manx	10	3.161	11.399	6.194
13.	p/n pussy	6	2.449	11.310	5.468
14.	p/n feral	9	2.999	11.060	6.038
15.	p/n Abyssinian	3	1.731	10.859	4.474
16.	p/n purring	7	2.644	10.698	5.675
17.	p/n Cheshire	40	6.319	10.079	7.944
18.	p/n pedigree	10	3.159	10.053	6.148
19.	p/n long-haired	3	1.730	10.007	4.463
20.	p/n stray	16	3.996	9.956	6.774
21.	p/n Fold	4	1.998	9.774	4.865
22.	p/n pet	31	5.561	9.764	7.586
23.	p/n alley	6	2.444	8.693	5.371
24.	p/n Practical	5	2.229	8.343	5.101
25.	p/n wild	49	6.976	8.206	7.545
26.	p/n ginger	4	1.993	8.164	4.791
27.	p/n contented	3	1.725	7.950	4.390
28.	p/n domestic	53	7.249	7.882	7.428
29.	p/n Wild	5	2.225	7.641	5.029
30.	p/n raining	3	1.723	7.595	4.363

	Filtr	Frekvence	T-score	MI	logDice
1.	p/n the	904	24.137	2.342	2.709
2.	p/n a	535	20.312	3.037	3.402
3.	p/n The	167	11.300	2.993	3.350
4.	p/n black	58	7.528	6.436	6.474
5.	p/n domestic	53	7.249	7.882	7.428
6.	p/n wild	49	6.976	8.206	7.545
7.	p/n Cheshire	40	6.319	10.079	7.944
8.	p/n big	40	6.228	6.036	6.044
9.	p/n A	40	5.676	3.285	3.594
10.	p/n pet	31	5.561	9.764	7.586
11.	p/n 's	66	5.559	1.663	2.019
12.	p/n like	29	4.676	2.925	3.229
13.	p/n Siamese	19	4.358	12.594	7.125
14.	p/n white	20	4.330	4.975	4.997
15.	p/n tabby	17	4.123	12.912	6.972
16.	p/n two	24	4.072	2.566	2.874
17.	p/n female	17	4.068	6.213	5.771
18.	p/n your	22	3.999	2.762	3.055
19.	p/n stray	16	3.996	9.956	6.774
20.	p/n fat	15	3.839	6.828	5.990
21.	p/n mother	15	3.679	4.320	4.390
22.	p/n little	16	3.616	3.381	3.595
23.	p/n my	19	3.486	2.320	2.624
24.	p/n Stray	12	3.464	13.263	6.477
25.	p/n Big	12	3.442	7.302	5.961
26.	p/n (34	3.419	1.273	1.624
27.	p/n old	14	3.364	3.309	3.511
28.	p/n tom	11	3.316	11.921	6.339
29.	p/n our	15	3.284	2.718	2.981
30.	p/n '	53	3.273	0.861	1.221



Thank you for your attention!

Questions?

Let's talk about language!





SEMINAR



Reading

common reading:

Stubbs, M. (2002). 2. Words, Phrases and Meanings in Use: Basic Concepts. In M. Stubbs, *Words and Phrases*, pp 24-53. Oxford: Blackwell.



Discussion

- What are the main concepts (not only) in phraseology that corpus linguistics might shed a new light on?
- How can we distinguish different senses of a word?
- Why is it important to distinguish between a word-form and a lemma in terms of collocation profiles?
- Why is a corpus relevant for semantic research?
- Why is it important to know about a word's frequency?
- What is the difference between content words and function words?
- What is a lexical density?
- How to decide what belongs to a core vocabulary?



Discussion

<http://textalyser.net/>

Trump:

1. <http://www.politico.com/story/2016/06/transcript-donald-trump-national-security-speech-224273> (1811)
2. <http://www.latimes.com/politics/la-na-pol-donald-trump-immigration-speech-transcript-20160831-snap-htmllstory.html> (3829)

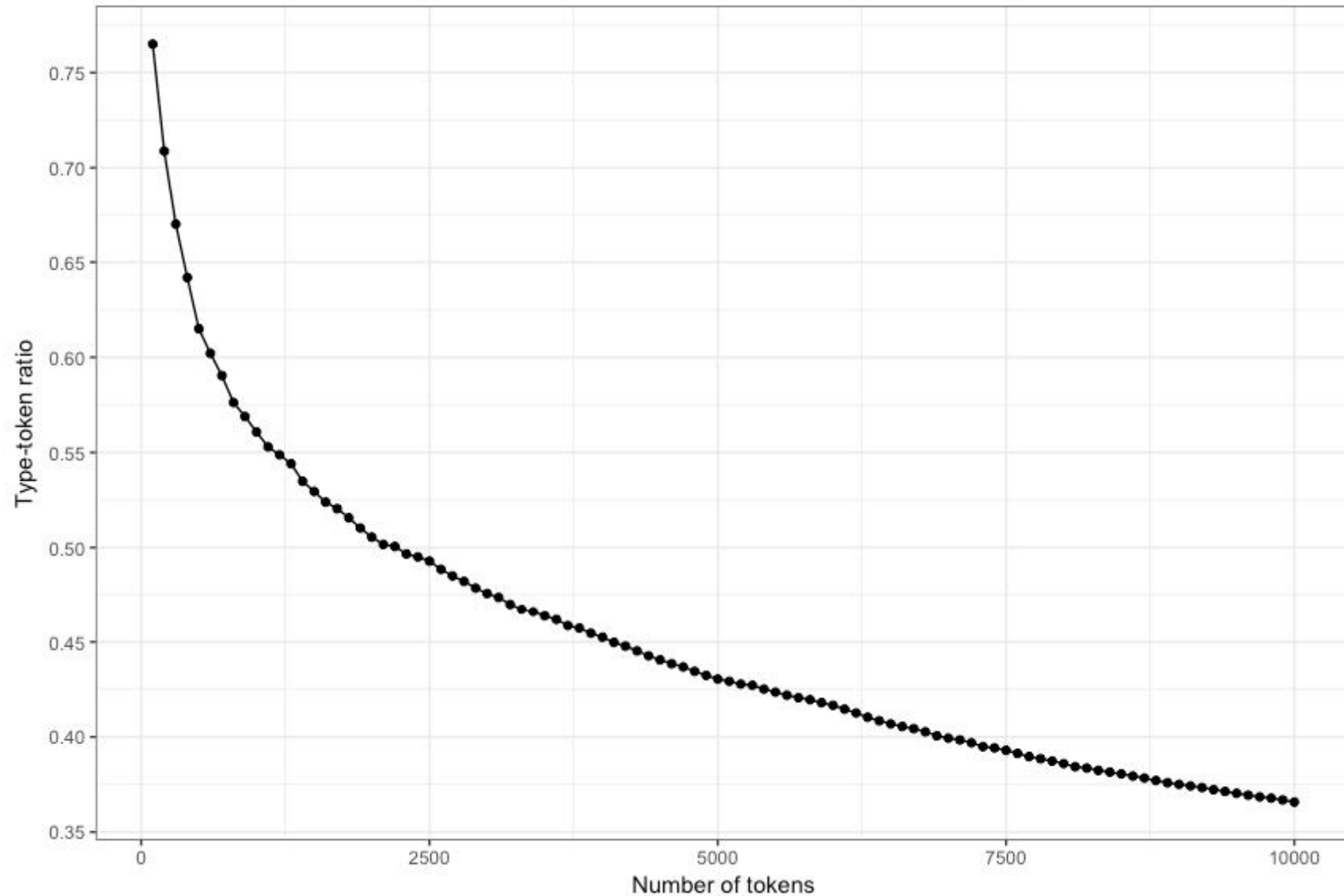
Clinton:

1. <http://www.politico.com/story/2016/08/transcript-hillary-clinton-alt-right-reno-227419> (1710)
2. <http://www.politico.com/story/2015/06/hillary-clinton-campaign-rally-speech-transcript-118973> (2986)



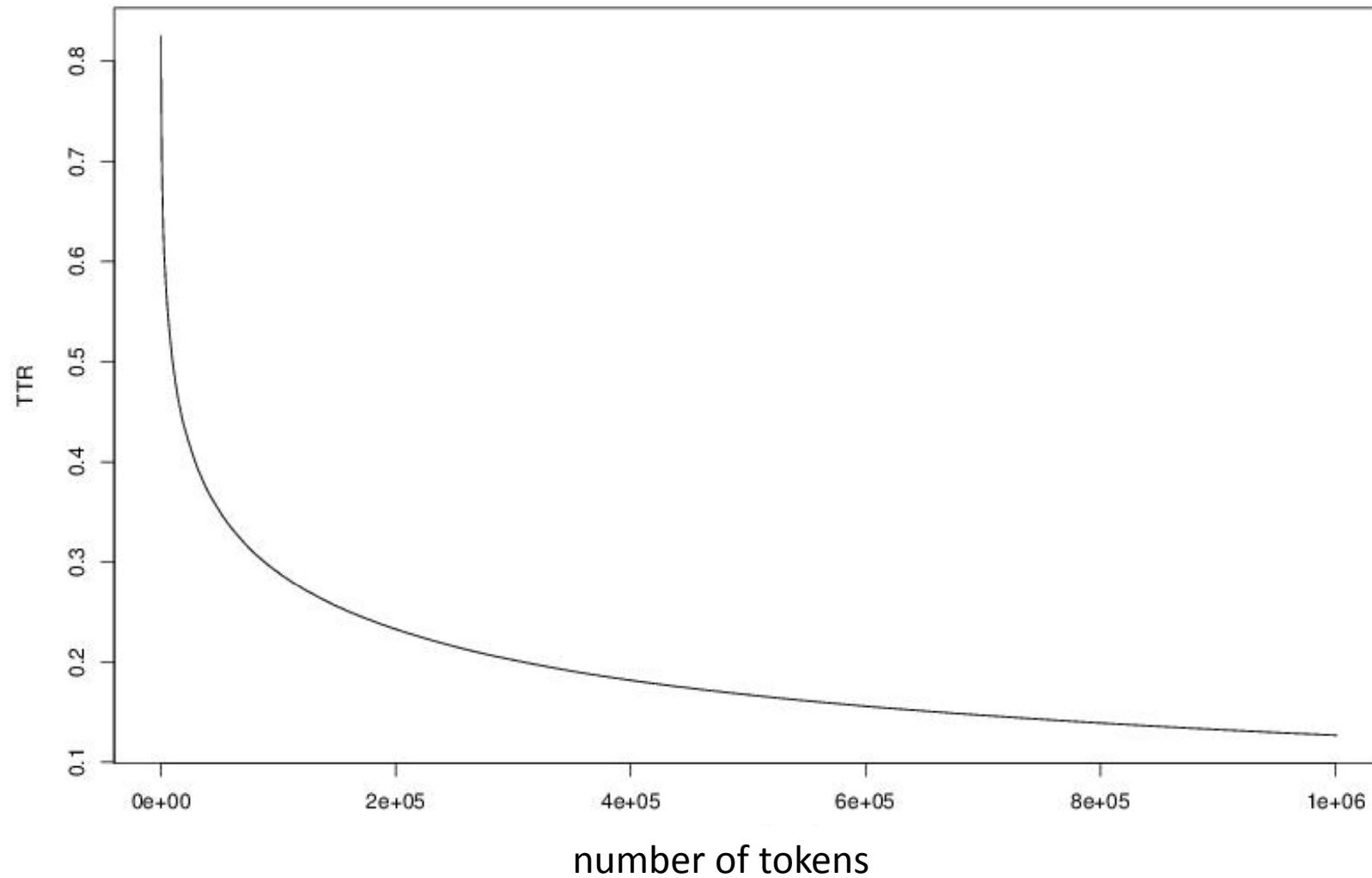
Discussion

TTR: text size matters!



Discussion

TTR: text size matters!





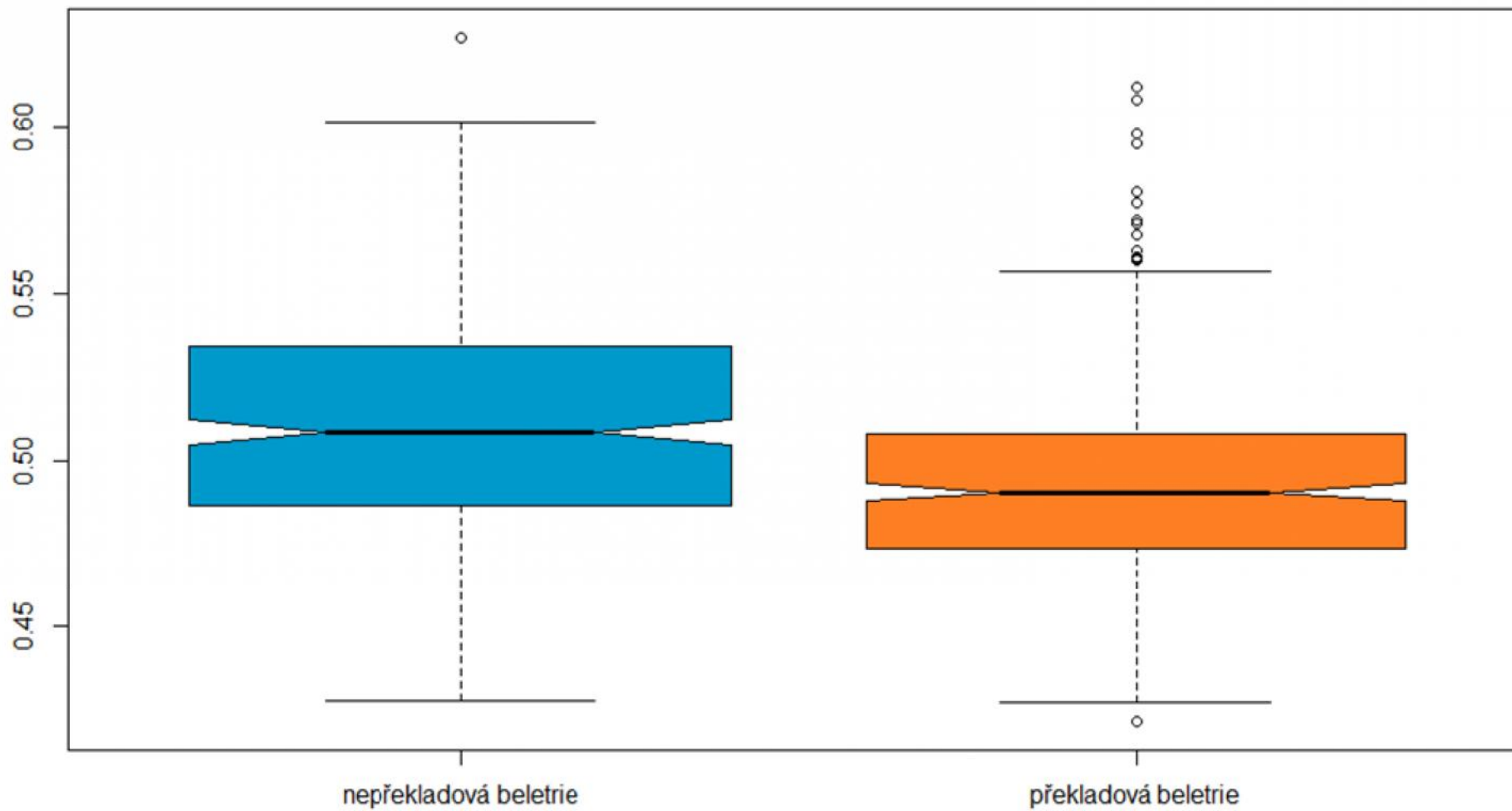
EXAMPLES OF LEXICAL DENSITY

ON NON-TRANSLATED AND TRANSLATED CZECH



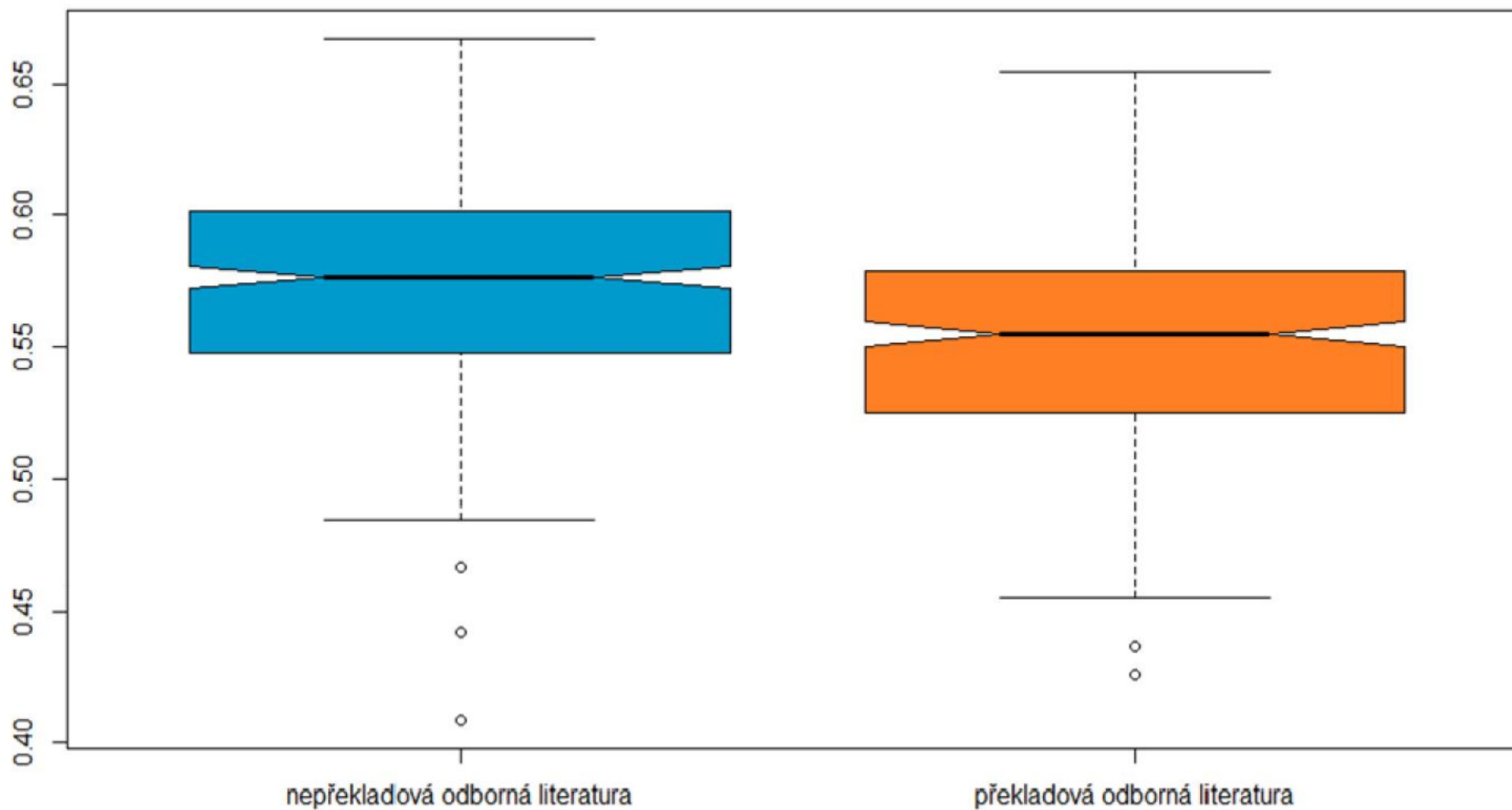
LD (fiction)

LEXICAL DENSITY



LD (non-fiction)

LEXICAL DENSITY



LD (total)

