Úvod do korpusové lingvistiky 13

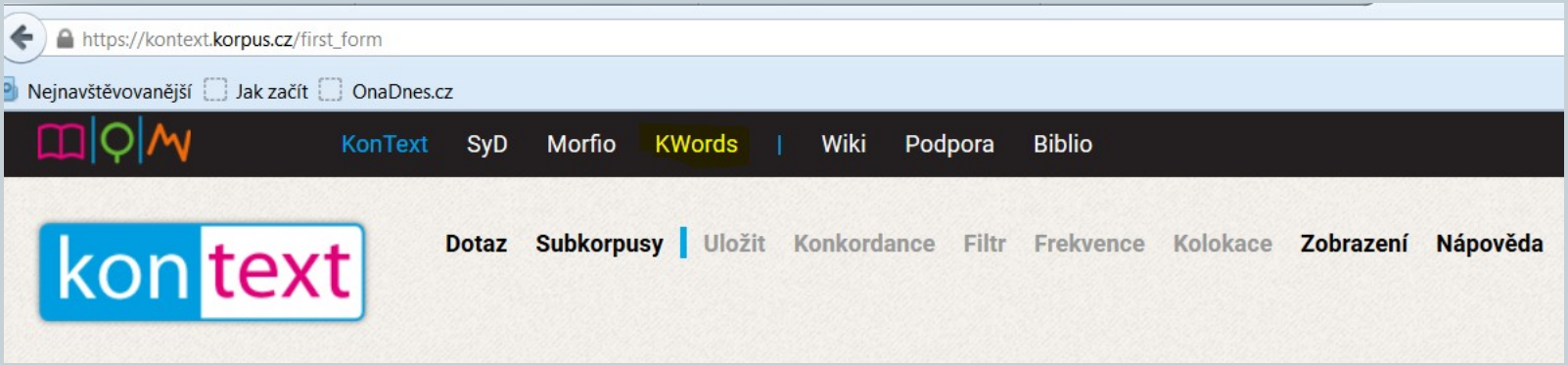
LINGVISTIKA LITERATURA KORPUSY

Využití jazykových korpusů v literárněvědném výzkumu

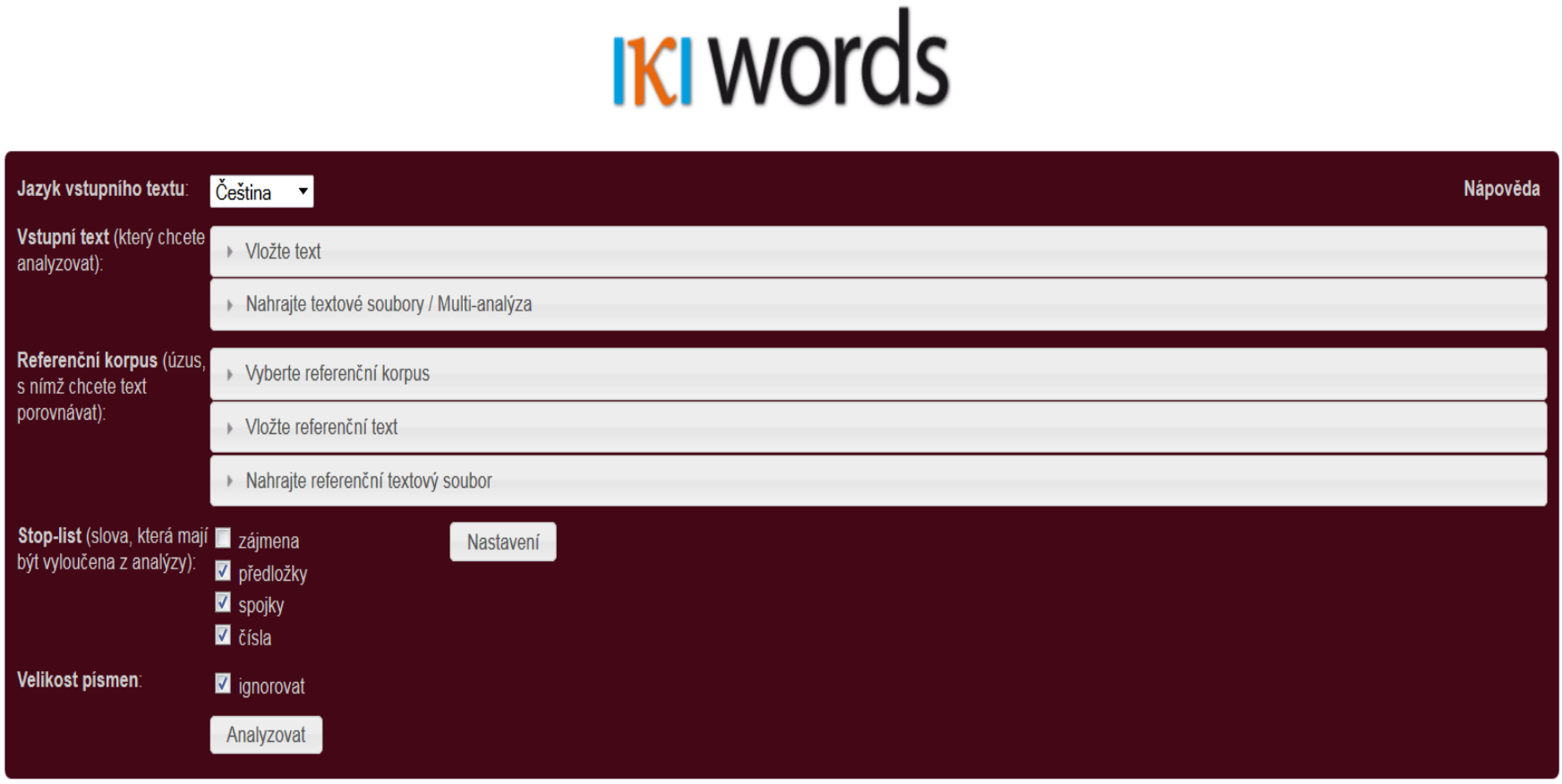
- J. R. R. Tolkien: **Valedictory address to the University of Oxford**
- In a Bestiary more nearly reflecting the truth *Lang* and *Lit* would appear **as Siamese Twins**, Jekyll-Hyde and Hyde-Jekyll, indissolubly joined from birth, **with two heads, but only one heart, the health of both being much better when they do not quarrel.** This allegory at least resembles more closely our older statute: *Every candidate will be expected to show a competent knowledge of both sides of the subject, and equal weight in the examination will be attacked to each.*

Klíčová slova v korpusu a klíčová slova autorská

- Autorské korpusy a autorské slovníky
- nástroj KWords



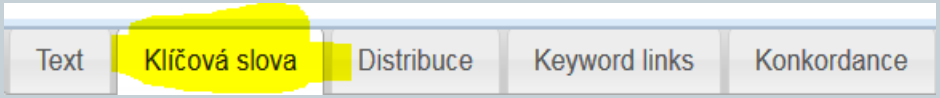
Jak vypadá KWords?



Co dělá KWords?

- **KWords** slouží k identifikaci klíčových slov (tzv. keywords): to jsou slova (resp. slovní tvary), která jsou úzce spojena s hlavními tématy textu a s jeho žánrem. Klíčová slova tedy nejsou uživatelem předvybrána (jak by mohlo naznačovat jiné užití tohoto termínu např. ve vyhledávacích aplikacích), naopak, jsou to slova, která nástroj v textu odhaluje jako nejvíc prominentní. Tato aplikace umožňuje provádět corpus-driven výzkum slov, která jsou klíčová pro vyznění textu a následnou interpretaci, při minimální interferenci ze strany badatelovy intuice.

O čem asi pojednávají vložený text?



O čem asi pojednávají vložený text?

Text | Klíčová slova | Distribuce | Keyword links | Konkordance

nejfrekventovanější slova na předcházející pozici	klíčové slovo	nejfrekventovanější slova na následující pozici
aniž (2) • mohl (2) • že (2) • který (2)	by (25)	to (4) • ji (3) • jí (2) • s (2)
, (3) • to (2) • samozřejmě (1) • konců (1)	byla (16)	to (4) • obyčejná (1) • opravdu (1) • taky (1)
druhou (1) • novou (1) • ještě (1) • odhodila (1)	cigaretu (4)	, (2) • ? (1) • od (1)
sladké (2) • irmgardiny (1)	dětičky (3)	, (3)
a (2) • to (1) • " (1) • číšnice (1)	řekla (6)	: (4) • - (1) • jí (1)
nebo (1) • tu (1) • , (1)	gaby (3)	? (1) • , (1) • ((1)
pro (3)	hosty (3)	, (3)
o (2) • " (1) • nevzala (1) • přece (1)	chlapce (5)	, (3) • si (1) • " (1)
nad (1) • sympatickým (1) • s (1)	chlapcem (3)	, (2) • přestěhovali (1)
se (8) • že (3) • , (2) • která (2)	jí (29)	to (2) • políbit (1) • líbil (1) • zaplatit (1)
ještě (5) • jí (1) • , (1)	jednou (7)	, (3) • se (1) • pokynul (1) • večer (1)
, (2) • přece (2) • tam (1) • a (1)	ještě (15)	jednou (5) • oříškový (1) • cigaretu (1) • kdesi (1)
v (3)	kavárně (3)	mohla (1) • se (1) • v (1)
do (2) • z (1)	kavárny (3)	. (1) • , (1) • venku (1)
uhlazovat (1) • ukázat (1) • , (1)	košile (3)	, (2) • a (1)
ulici (1) • nerušeně (1) • zvyk (1)	kouřit (3)	, (2) • na (1)
druhou (1) • nesla (1) • , (1)	kávu (3)	pro (1) • a (1) • , (1)
manželské (1) • do (1) • z (1)	ložnice (3)	do (1) • vstoupí (1) • s (1)
to (1) • tu (1) • , (1)	lotte (3)	nebo (2) • ? (1)
a (1) • kde (1) • ten (1)	milí (3)	zástupce (1) • soudcové (1) • právní (1)
ten (3)	milý (3)	chlapec (1) • pan (1) • právní (1)
. (4) • a (2) • , (2) • : (1)	možná (10)	, (6) • spolu (1) • slučovací (1) • dvacet (1)
vedle (3) • na (2) • po (1) • s (1)	ní (9)	, (2) • přes (1) • zmeškal (1) • zůstat (1)
. (5) • , (4) • " (3) • tady (1)	ne (23)	, (8) • . (6) • " (3) • : (1)
tak (3)	neberte (3)	. (2) • " (1)
, (2) • on (1) • a (1)	nechal (4)	si (1) • ji (1) • zatím (1) • zpopelnit (1)
všem (1) • cos (1) • nic (1)	nechtěla (3)	ukázat (1) • ani (1) • , (1)
, (2) • ; (1) • by (1)	nemohla (4)	pracovat (1) • se (1) • mu (1) • by (1)
? (1) • že (1) • , (1)	nepochopil (3)	, (3)
smrt (4)	nerozdělí (4)	" (3) • první (1)
dobu (1) • jí (1) • a (1) • mladá (1)	číšnice (5)	, (3) • řekla (1) • právě (1)
s (4) • před (1) • za (1)	ním (6)	ulehla (1) • hrát (1) • . (1) • skutečně (1)

O čem asi pojednávají vložený text?

Charakter textu

- Žánr: povídka a věcný text
- Téma: ?
- Další charakteristiky

Poezie a korpusy

- Slovník rýmů

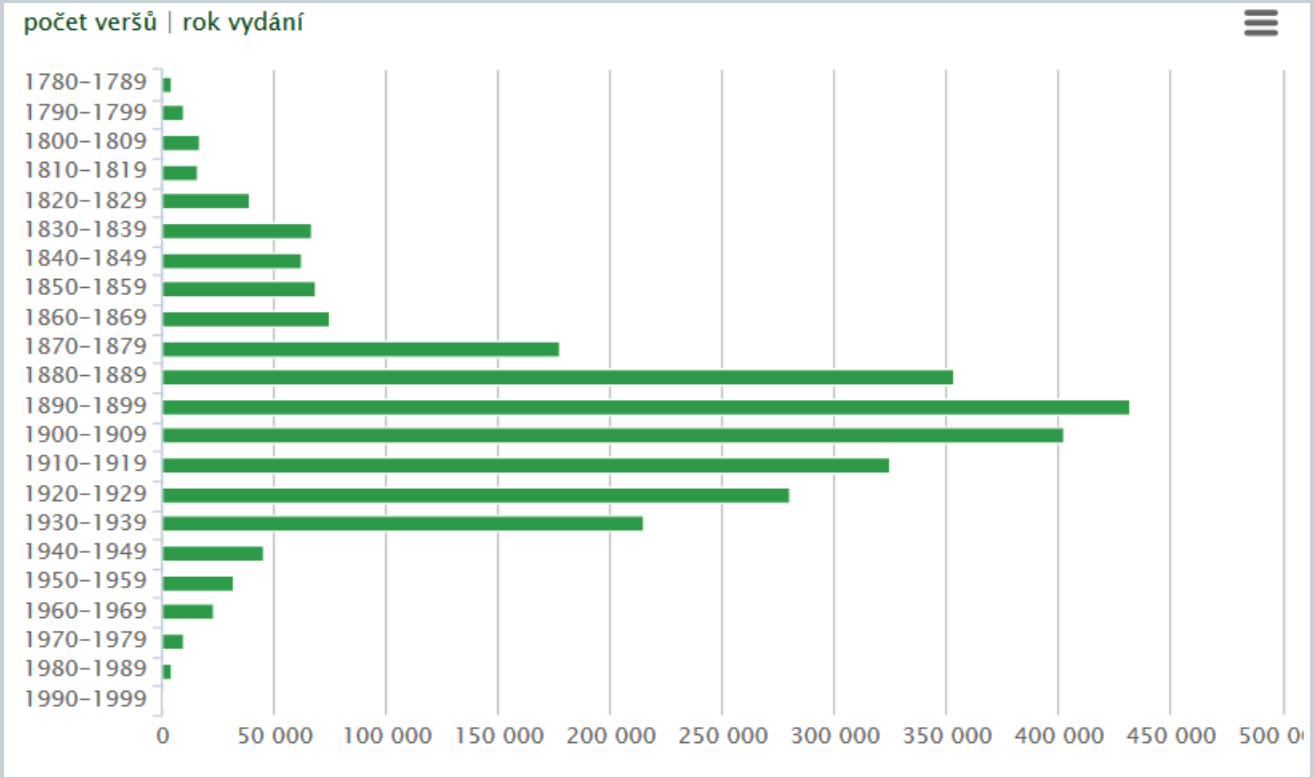
Korpus českého verše

- http://www.versologie.cz/kcv.html
- Korpus českého verše (KČV) je lemmatizovaný, foneticky, morfologicky, metricky a stroficky anotovaný korpus české poezie 19. a počátku 20. století.*
  Ke každé slovní jednotce v korpusu je připojena informace o jejím základním slovním tvaru (lemma), fonetickém přepisu a gramatických kategoriích, u každého verše je určeno metrum (jamb, trochej…), rozsah (*n*-stopý), typ klauzule (mužská, ženská…) a metrický vzorec. (V současnosti jsou z hlediska metriky anotovány pouze verše sylabotónické.) Na vyšších rovinách jsou pak anotovány rýmové dvojice, resp. *n*-tice a pevné formy (sonet, rondel…). V metrickém a strofickém popisu je možné vyhledávat prostřednictvím Databáze českých meter, rovina lemmatizace je částečně zpřístupněna prostřednictvím Frekvenčních slovníků, rýmové páry lze vyhledávat v aplikaci Gunstick

Základní charakteristika korpusu

- 1 689 básnických sbírek
- 78 391 básní
- 2 664 989 veršů
- 14 592 037 slov

Složení korpusu českého verše



Ukázka

strofa #1	text	Už je konec štěstí;					
verš #1	token	už	je	konec		štěstí	
	lemma	už	být	konec		štěstí	
rým #0	morfologická značka [?]	Db----------	VB-S---3P-AA---I	NNIS1-----A-----		NNNS2-----A-----	
	interpunkce					;	
	slabika (SAMPA) [?]	uS	je	kon	et_s	Sce	sci:
	přízvuk	*		*		*	
	metrický vzorec	S	W	S	W	S	W
	rozměr	T 3 z					

strofa #1	text	pohněvaná víla					
verš #2	token	pohněvaná				víla	
	lemma	pohněvaný				víla	
rým #1	morfologická značka [?]	AAFS1----1A-----				NNFS1-----A-----	
	interpunkce						
	slabika (SAMPA) [?]	po	h\Je	va	na:	vi:	la
	přízvuk	*				*	
	metrický vzorec	S	W	S	W	S	W
	rozměr	T 3 z					

Přečtěte si

- http://www.lexically.net/wordsmith/corpus_linguistics_links/R%F6mer%202006%20in%20Gerbig%20and%20M%FCller-Wood.pdf
- J. R. R. Tolkien: **Valedictory address to the University of Oxford (**http://www.e-reading.club/bookreader.php/130388/Tolkien_-_Valedictory_address_to_the_University_of_Oxford.html**)**
- (česky in *Netvoři a kritikové*, **Řeč na rozloučenou s Oxfordskou univerzitou** (*Valedictory Address to the University of Oxford*))
- **IBRAHIM, R. – PLECHÁČ, P. Báseň a počítač,** Praha: Academia 2014.