

Úvod do korpusové lingvistiky

8/9



PARALELNÍ KORPUSY
WEBOVÉ KORPUSY

Terminologická poznámka



- Paralelní korpus
- Zarovnání
- Srovnatelný korpus

Projekt Intercorp



- <https://intercorp.korpus.cz/>
- teoretické studie
- lexikografii
- studentské práce
- výuka cizích jazyků
- počítačové aplikace
- překladatelé i veřejnost

Zdroje dat



Korpusy cizojazyčné

Psané paralelní korpusy (překlady a originály)

korpus	velikost (počet slov)	lemmatizace	morfologické značky	rok zveřejnění	charakteristika korpusu
InterCorp (verze 13ud, verze 15)	1,8 mld.	(✓)	(✓)	2008-2022	verzovaný paralelní korpus zahrnující kromě češtiny 41 dalších jazyků
Žalm 77	10 tis.	(✓)	(✓)	2023	paralelní korpus 11 verzí textu žalmu 77 v rumunštině, církevní slovanštině a řečtině

Jádrem korpusu InterCorp jsou ručně zarovnané, převážně beletristické texty. Kromě toho korpus obsahuje také nabídku automaticky zpracovaných textů, tzv. **kolekce**. Ve verzi 15 jsou k dispozici tyto kolekce:

- publicistické články a zpravodajství z webových stránek [Project Syndicate](#) a [VoxEurop](#) (dříve PressEurop)
- právní texty Evropské unie z korpusu [Acquis Communautaire](#)
- zápisy jednání Evropského parlamentu z let 2007–2011 z korpusu [Europarl](#)
- filmové titulky z databáze [OpenSubtitles](#)
- překlady Bible

Možnost vyhledávat přes rozhraní KonText



Část názvu nebo popisu

InterCorp v15 - Arabic	482k
InterCorp v15 - Belarusian	8.26M
InterCorp v15 - Bulgarian	36.1M
InterCorp v15 - Catalan	11.2M
InterCorp v15 - Czech	266M
InterCorp v15 - Danish	73.5M
InterCorp v15 - German	110M
InterCorp v15 - Greek	80.3M
InterCorp v15 - English	168M
InterCorp v15 - Spanish	144M
InterCorp v15 - Estonian	46.7M
InterCorp v15 - Finnish	61.5M
InterCorp v15 - French	

Ize užívat lemmatizaci i RV



Korpus: InterCorp v15 - French | Dotaz: .*er (7 792 378 výskytů) ~ [Podrobnosti](#)

Výskytů: **7 792 378** | i.p.m.: **65 064,55** (vztaženo k celému korpusu) | ARF: **5 206 831,09** | Výsledek je setříděn

1 / 194 810

Výběr řádků:

<input type="checkbox"/>	adams-stoparuv_pruvodc	famés du bout du bras occidental de la Galaxie ,	traîne	un petit soleil jaunâtre et minable . En orbite autour
<input type="checkbox"/>	adams-stoparuv_pruvodc	, traîne un petit soleil jaunâtre et minable . En	orbite	autour de celui -ci , à la distance approximative de
<input type="checkbox"/>	adams-stoparuv_pruvodc	distance approximative de cent cinquante millions de kilomètres , se	trouve	une petite planète bleu-vert et totalement négligeable dont les habitan
<input type="checkbox"/>	adams-stoparuv_pruvodc	la plupart du temps . Bien des solutions avaient été	suggérées	mais la plupart d' entre elles faisaient largement intervenir la
<input type="checkbox"/>	adams-stoparuv_pruvodc	largement intervenir la mise en circulation de petits bouts de	papier	vert , chose curieuse puisque en définitive ce n' étaient
<input type="checkbox"/>	adams-stoparuv_pruvodc	puisque en définitive ce n' étaient pas les bouts de	papier	vert qui étaient malheureux . Et donc le problème subsistait
<input type="checkbox"/>	adams-stoparuv_pruvodc	papier vert qui étaient malheureux . Et donc le problème	subsistait	; des tas de gens se sentaient minables et la
<input type="checkbox"/>	adams-stoparuv_pruvodc	à quartz numériques . Un nombre croissant d' entre eux	partageait	cette opinion selon laquelle leur plus grosse erreur aurait été
<input type="checkbox"/>	adams-stoparuv_pruvodc	le début de descendre des arbres . D' aucuns même	affirmaient	qu' avec les arbres déjà ... et qu' on aurait
<input type="checkbox"/>	adams-stoparuv_pruvodc	... et qu' on aurait mieux fait de ne jamais	quitter	les océans . Et puis , un beau jeudi ,

Projekt Kapradí



- <https://www.phil.muni.cz/kapradi/>

GENERAČNÍ DĚLENÍ

1. generace	(nejstarší překlady)	Karel Hynek Thám, Josef Kajetán Tyl, Josef Jiří Kolár
2. generace	(muzejní edice, pol. 19. st.)	František Doucha, Jan Josef Čejka, Ladislav Čelakovský, Jakub Malý
3. generace	(přelom 19. a 20. st.)	Josef Václav Frič, Josef Václav Sládek, Jaroslav Vrchlický
4. generace	(počátek 20. st.)	Otokar Fischer, Antonín Fencel, Bohumil Štěpánek
5. generace	(polovina 20. st.)	Erik Adolf Saudek, Jiří Valja, František Nevrla, Otto František Babler
6. generace	(60.–70. léta 20. st.)	Aloys Skoumal, Zdeněk Urbánek, Josef Topol, Václav Renč, Břetislav Hodek
7. generace	(moderní překlady)	Alois Bejblík, Martin Hlinský, Jiří Josek, Milan Lukeš, Antonín Přidal

<https://corpora.fi.muni.cz/kapradi/run.cgi/locate>



Texty

Vyhledat

Projekt

Databáze se upravuje. Pracujete s nekompletními texty. Text: Topol_RJ.cz [Browse text](#)

SCE: 2.2

[Block/KWIC](#)

2.2

Julie. Jen podle jména jsi můj nepřítel.
Ty nejsi Montek, ty sám jsi jen ty.
Vždyť Montek není ruka ani tvář,
není to paže, noha, cokoli,
z čeho je člověk. Tak se přejmenuj!
Co to je jméno? To, **co růží** zvem,
nazváno jinak bude vonět stejně.
Kdyby ti neříkali Romeo,
nic na tvé vzácné dokonalosti
by nebylo. To jméno dej pryč.
Za to, co není ani částí tebe,
vem si mě celou.

Webové korpusy



- Řada Araneum
- Řada TenTen

Sketch Engine <https://www.sketchengine.eu/>



LOG IN | Sign up

FREE Trial

subscribe to news

CORPORA

Home

News & Events

Pricing

Guide

About us

Contact



LEARN HOW LANGUAGE WORKS

WHAT IS SKETCH ENGINE

LOG IN

SKETCH

THESAURUS

Volně pro MU



- Přístup k řadě korpusů
- Přístup k word sketch /slovní profily
- Porovnání slovních profilů
- Vizualizace vztahů slov /Thesaurus



DASHBOARD

Brown Corpus



BROWN CORPUS



Word Sketch

Collocations and word combinations



Thesaurus

Synonyms and similar words



Parallel Concordance

Translation search



N-grams

Multiword expressions (MWEs)



Trends

Diachronic analysis, neologisms

CORPUS INFO

MANAGE CORPUS



Word Sketch Difference

Compare collocations of two words



Concordance

Examples of use in context



Wordlist

Frequency list



Keywords

Terminology extraction



Text type analysis

Statistics of the whole corpus

Concordance

CONCORDANCE

Brown Corpus



Get more sp

BASIC

ADVANCED

ABOUT

Query type ?

Simple

abc

simple

lemma

phrase

word

character

CQL

Subcorpus ?

none (the whole corpus)



Macro ?

none



Details

Left context

KWIC

Right context



various

ted in full text or fairly accurate excerpts behind the

Iron **Curtain**
Iron/NP Curtain/NP

./s><s>Then the Communes reply came back from



A Foreign Polic... ations to plant our influence on the other side of the

Iron **Curtain**
Iron/NP Curtain/NP

./s><s>There was the Hungarian Revolution which

Co jsou word sketch



WORD SKETCH

German Web 2020 (deTenTen20)



rot as adjective 2,444,351×



↔	☰	🔍	✕
Constructions			
"%w" in positive degree	...		
"%w" in nominative	...		
"%w" in dative	...		
"%w" in accusative	...		
"%w" in genitive	...		

↔	☰	🔍	✕
modifiers of "rot"			
glatt	...		
glatt Rot			
• especially: news			
• especially: sports			
ehemals	...		
ehemals roten			
wahlweise	...		
wahlweise rot			
zwischen	...		
zwischen rot			
gegen	...		
gegen Roter Stern			
• especially: news			
kräftiger	...		
kräftiger rot			

↔	☰	🔍	✕
nouns modified by "rot"			
Kreuz	...		
Roten Kreuzes			
• especially: news			
Faden	...		
ein roter Faden			
• especially: arts			
Armee	...		
der Roten Armee			
• especially: reference			
Karte	...		
Rote Karte			
• especially: sports			
• especially: news			
Teppich	...		
dem roten Teppich			
• especially: news			

Word sketch differ



WORD SKETCH DIFFERENCE German Web 2020

rot 2,444,351× **gelb** 796,834×

modifiers of "rot/gelb"				nouns modified by "rot/gelb"			
glatt	758	0	...	Armee	51,619	0	...
gegen	53	0	...	Faden	69,198	173	...
ehemals	221	39	...	Kreuz	127,861	567	...
wahlweise	108	25	...	Teppich	37,089	232	...
innen	600	272	...	Liste	29,525	337	...
zwischen	137	84	...	Farbe	20,786	9,670	...
kräftiger	47	26	...	Karte	54,026	45,320	...
zweimal	220	263	...	Blüte	3,253	6,475	...
zurecht	36	107	...	Trikot	2,756	9,964	...
5.	38	132	...	Tonne	404	7,029	...
16.	0	27	...	Ei	534	11,474	...
17.	0	25	...	Sack	147	13,582	...

Thesaurus



WORD SKETCH

German Web 2020 (deTenTen20)



Get more space



rot as adjective 2,444,351×



Number of collocates (10)



Fewer

More

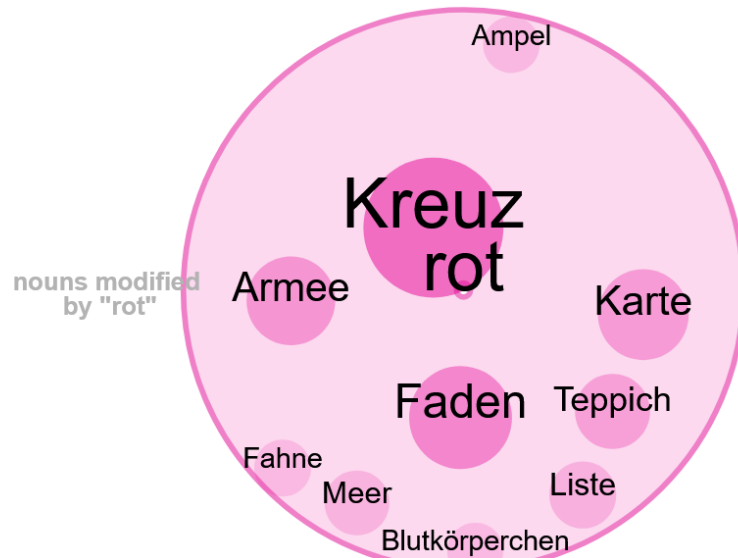
Constructions

modifiers of "rot"

nouns modified by "rot"

DOWNLOAD IMAGE

[About visualization](#)





Další volby



WORD SKETCH

German Web 2020 (deTenTen20)



Get more space +



rot as adjective 2,444,351×



Number of collocates (10)



Fewer

More

Constructions

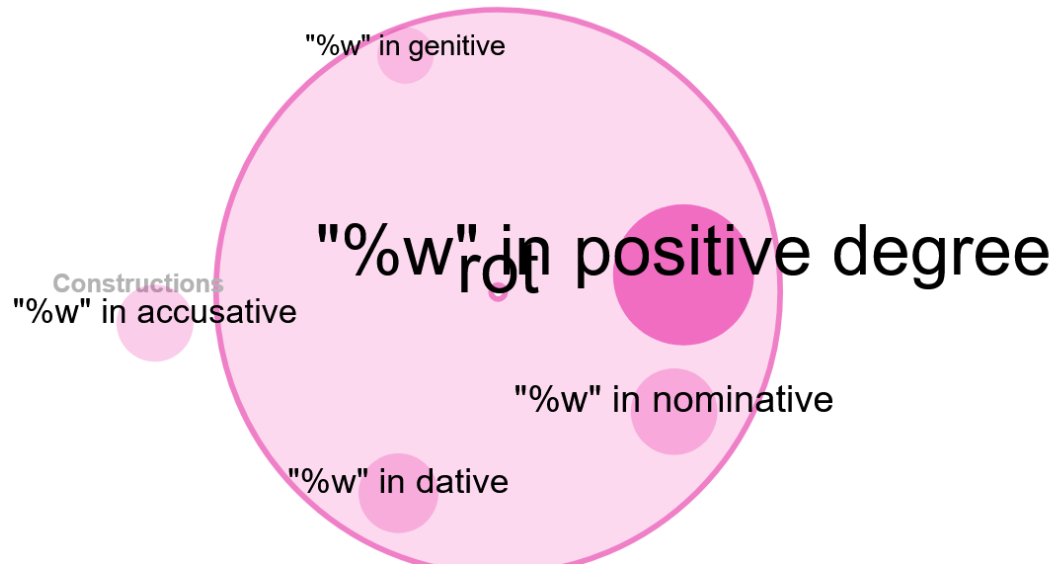
modifiers of "rot"

nouns modified by "rot"



DOWNLOAD IMAGE

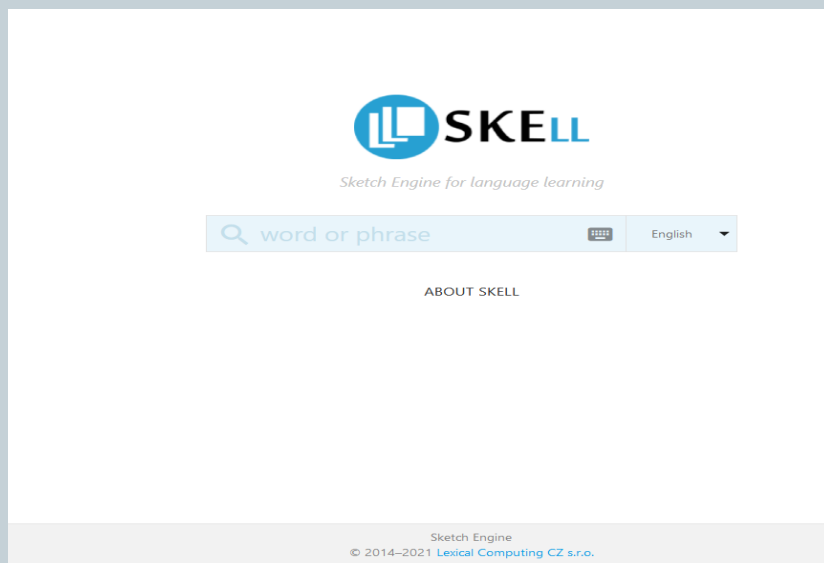
[About visualization](#)



Využití ve výuce




- Nástroj Skell (<https://skell.sketchengine.eu/>)



Nabídka



Sketch Engine for language learning

 **entwerfen**



Deutsch ▼

entwerfen in [Examples](#)

entwerfen in [Word sketch](#)

entwerfen in [Similar words](#)

UT SKELL

Příklady



entwerfen 19.36 hits per million

1. Lieber gut nachgebaut als schlecht selbst **entworfen** .
2. Alles wurde liebevoll von mir selbst **entworfen** und zusammengestellt.
3. Aber etwas kann man auch selbst **entwerfen** .
4. Die ist von uns selbst **entworfen** und angefertigt worden.
5. Hatte ich selbst bereits vor 2004 **entworfen** .
6. Er hat diese Route auch "erfunden" und **entworfen** .
7. Und genau diese haben wir als USB-Stick **entworfen** .
8. Es wurde eigens für unseren Park **entworfen** und gebaut.
9. Die hübsche Dekoration haben sie selbst **entworfen** und angefertigt.
10. Malen war gestern- heute wird **entworfen** !
11. Dabei sind sehr schöne Plakate **entworfen** worden.
12. Darum habe ich mich dran gemacht und jeweils ein Liste **entworfen** .
13. Crisis 2 ist für die Konsole **entworfen** worden.
14. Dann mussten wir unseren eigenen Stundenplan **entwerfen** .
15. Nur wie geht das eine Homepage zu **entwerfen** ?
16. Die Plakate dafür wurden jetzt schon **entworfen** und gedruckt.

Podobná slova



entwerfen

verb ▾



kreieren anfertigen fertigen konzipieren aussuchen realisieren basteln zusammenstellen erarbeiten gestalten erfinden malen erstellen
produzieren drucken überlegen bearbeiten ausprobieren auswählen anschauen erschaffen verfassen definieren umsetzen fotografieren
ansehen herstellen zeichnen aufbauen verwirklichen



Word sketch



entwerfen



Deutsch

Examples

Word sketch

Similar words

entwerfen

verb



Show context

modifiers of "entwerfen"

1. [eigens](#)
2. [extra](#)
3. [kurzerhand](#)
4. [selbst](#) selbst entworfen
5. [selber](#) selber entworfen
6. [zusammen](#)
7. [einst](#)
8. [seither](#)
9. [nebenbei](#)
10. [übrigens](#)
11. [quasi](#)
12. [hin](#)
13. [etc.](#)
14. [nämlich](#)
15. [ausschließlich](#)

nouns with "entwerfen" as genitive object

1. [Architekt](#)
2. [Identity](#)
3. [Kollektion](#)
4. [Outfit](#)
5. [Zukunft](#) der Zukunft zu entwerfen
6. [Jahrhundert](#)
7. [Wettbewerb](#)
8. [Hochschule](#)
9. [Gesellschaft](#)
10. [Generation](#)
11. [Partner](#)
12. [Schule](#) Schule entworfen
13. [Firma](#)
14. [Projekt](#)
15. [Frau](#)

dative objects of "entwerfen"

1. [Architekt](#) Architekten entworfen
2. [Designer](#) Designern entworfen
3. [Designerin](#)
4. [Reißbrett](#)
5. [T-Shirt](#)
6. [Grafiker](#)
7. [Atelier](#)
8. [Künstler](#)
9. [Werbeagentur](#)
10. [Künstlerin](#)
11. [Label](#)
12. [Sorgfalt](#)
13. [Handarbeit](#)
14. [Gesichtspunkt](#)
15. [Vorstellung](#) Vorstellungen zu entwerfen

accusative objects of "entwerfen"

1. [Logo](#)
2. [Kollektion](#)
3. [Plakat](#)
4. [Flyer](#) Flyer entworfen
5. [Kostüm](#)
6. [Szenario](#)
7. [Bühnenbild](#)
8. [Design](#)
9. [Mode](#)
10. [Kleid](#)
11. [Konzept](#)
12. [Möbel](#)
13. [Layout](#)
14. [Szenarium](#)
15. [Plan](#)

subjects of "entwerfen"

1. [Designerin](#) Designerin entwirft
2. [Architekt](#)
3. [Designer](#)
4. [Grafiker](#)
5. [Logo](#)
6. [Architektin](#)
7. [Lagerfeld](#)
8. [Visitenkarte](#)
9. [Mode](#)
10. [Shirt](#)
11. [Label](#)
12. [Ingenieur](#)
13. [Flyer](#)
14. [Künstlerin](#)
15. [Design](#)

Příklad otázek v testu

