

Projektrelevante Webkorpora & Analysesoftware

Dr. Elena Krotova

kroelebor@gmail.com

Überblick

- **Niveau 1**
erweiterte Google-Suche, online Wörterbücher, Ngram-Viewer
- **Niveau 2**
DeReKo, DWDS, Sketch Engine
- **Niveau 3**
Reguläre Ausdrücke in DeReKo & Sketch Engine
Python für Frequenzlisten

Niveau 1: Phrasemkonstruktionen

- es sei denn X [link](#)
- ob X oder nicht [link](#)
- und ob er es versteht - und ob X! [link](#)
- durch und durch X / X durch und durch - [link](#)
- erst recht / [und] erst recht [nicht] X / und] X erst recht [nicht] - [link](#)
- noch lange nicht X - noch lange nicht X / kein[e] N / noch längst nicht X / kein[e] N - [link](#)
- so was von - so was von Adj/Adv [+Partizip Perfekt]/N - [link](#)

Niveau 2: Phrasemkonstruktionen

- sei es X, sei es Y - sei es X oder Y - [link](#)
- du und dein Wagen - Pronpers nom/N'nom und dein(e)/sein(e)/ihr(e)/euer(eure)... N"nom / Pronpers nom/N'nom [schon wieder/immer] mit deinem/seinem/deiner/seiner/seinen/ihrem... N"nom / Pronpers nom/N'nom mit deinem/seinem/deiner/seiner/seinen/ihrem... N"nom [schon wieder?/!] - [link](#)
- und das nennt sich N - [Und] so etwas/so eine(r)/so was/das nennt sich Nnom[!/?] / [Und] so etwas/so eine(r)/so was/das schimpft sich Nnom[!/?] / das / so etwas nenne ich X[!] - [link](#)

Niveau 3: Phrasemkonstruktionen

- Kind ist Kind - X' ist/sind [immer] X' [link](#)
- Fragen über Fragen (Gründe über Gründe) - N' über N' - [link](#)
- Tod bleibt Tod - X'nom bleib(s)t/bleibe/bleiben X'nom - [link](#)
- X ist nicht gleich X - X'nom ist/sind nicht gleich X'nom - [link](#)
- Ring der Ringe - N' der N'gen pl / N' aller N'gen pl - [link](#)
- X hin oder her - X hin oder her / X' hin, X' her / X hin, Y her [link](#)

[PhK Liste](#) (nach Niveau geteilt)

Niveau 1

- erweiterte Google-Suche
- online Wörterbücher
- Ngram-Viewer

 Google Books Ngram Viewer

 Linguee

 Reverso

 Glosbe
Dictionary

USW.

Niveau 1: erweiterte Google-Suche

google.com →

Settings →

Advanced Search

[link](#)

this exact word or phrase:

"ob * oder nicht"

any of these words:

none of these words:

numbers ranging from:

Then narrow your results
by...

language:

German

Niveau 1: online Wörterbücher

Glosbe [results](#)

Linguee [results](#)

Reverso [results](#)



Linguee



USW.

Externe Quellen (nicht geprüft)

Von einer wirklichen Beteiligung der nichtstaatlichen Akteure im Allgemeinen und der Frauenorganisationen im Besonderen kann allerdings **noch lange nicht** gesprochen werden.
⇒ eur-lex.europa.eu

Jednakże wciąż daleko jeszcze do faktycznej partycypacji podmiotów niepaństwowych generalnie oraz organizacji kobiet w szczególności.

⇒ eur-lex.europa.eu

Niveau 1: Ngram-Viewer

Instruktion

Link: ob * oder nicht

Search in Google Books

ob gewollt oder nicht



1800 - 1919

1920 - 1992

1993 - 1997

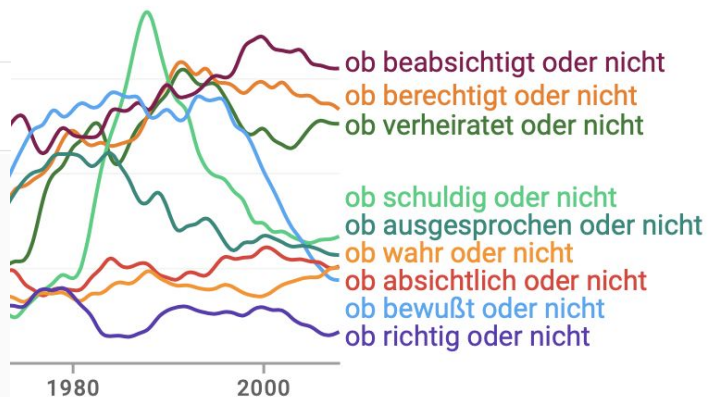
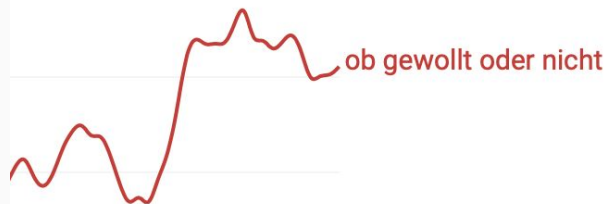
1998 - 2002

2003 - 2008

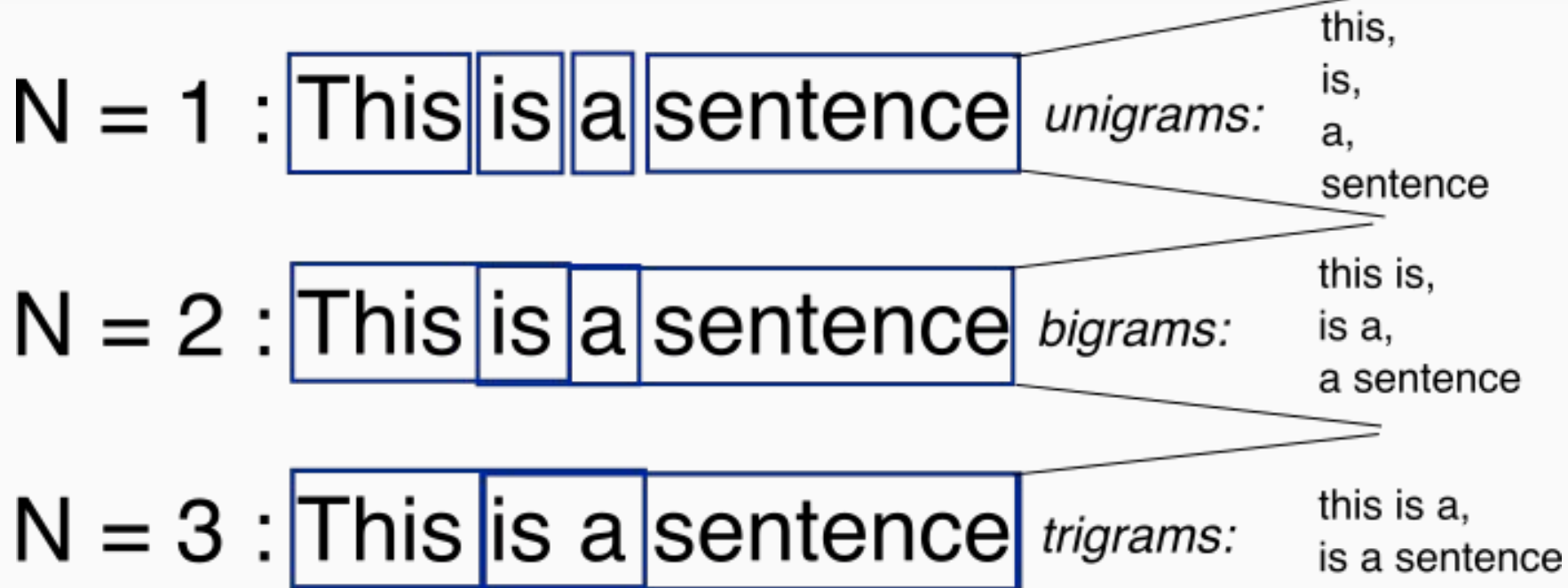
German

FOUND INSIDE – PAGE 143

... **Ob gewollt oder nicht**, er geht weder mit Nachfragen oder Ermahnungen auf ihre Aufregungen ein und entzieht sich damit als Übertragungsobjekt. Ob diese nüchterne Reaktion von Lauber 3.3 Ein Spezialist und eine Patientin 143.



n-grams



Niveau 1: Ngram-Viewer

- es sei denn X [link](#)
es sei denn * [results](#)
- ob X oder nicht [link](#)
ob * oder nicht, *_NOUN oder nicht [results](#)
- und ob er es versteht - und ob X! [link](#)
START und ob * [results](#)

NOUN	Noun
PROPN	Proper Noun
VERB	Verb
ADJ	Adjective
ADV	Adverb
PRON	Pronoun
DET	Determiner or article
ADP	An adposition: either a
NUM	Numeral
CONJ	Conjunction
PRT	Particle
ROOT	Root of the parse tree
START	Start of a sentence
END	End of a sentence

Niveau 2

Suchanfragen in:

- DeReKo
- DWDS
- Sketch Engine



COSMAS II



DeReKo

<https://cosmas2.ids-mannheim.de/cosmas2-web/>



&	Suche nach Lemmas: &haben
OR	Logischer Operator "oder": (du OR ihr)
/+w..	Distanz zwischen Wörtern: hin /+w2:3 her
%+w1	Ausschließen von Wörtern: von wegen %+w1 (, OR .)
“”	Suche nach <i>und, oder, nicht</i> : (hin "und" her) OR (hin "oder" her)
%s0	Ausschließen von bestimmten Strukturen im gleichen Satz: (hin /+w2:3 her) %s0 ((hin "und" her) OR (hin "oder" her))
MORPH()	Suche mit Tags: MORPH(N nn) /+w1:2 wie (er OR sie OR es) im (Buch OR Buche) &stehen
REG()	Operator für reguläre Ausdrücke: von wegen %+w1 (#REG([.](![:])))

Suchanfragen in DeReKo

<lemma> [ob] X [oder nicht]

<variants name 1> X oder nicht

<variants name 2> ob Y, ob X

<variants name 2> ob X oder Y

"oder" "nicht"

ob /+w2:3 ob

ob /+w2:3 "oder"

|< < Seite 1 von 50 > >|

n Geschäft mit dem Tod weiter mitverdienen oder nicht?

: einen verständigen Arbeitgeber.» Halbprofi oder nicht, eines ist für bei

Ob 40 Jahre ein Anlass zum Feiern seien oder nicht, darüber waren

er beschäftigt mich die Frage, ob es regnet oder nicht.

|< < Seite 1 von 35 > >|

Sie spielen ihre Lieder, ob Rock, ob Blues oder fast s

n sie ihr Ziel: «Jeder Gast, ob gross, ob klein, soll sich b

Jeder Bewohner, ob gesund, ob krank, gilt als

|< < Seite 1 von 50 > >|

ndlich völlig unwichtig, ob ein Mann oder eine Frau diese

Die Filme werden (ob Regen oder Sonne), um etwa

nd steckt das Theater, ob «frei» oder «etabliert», wirklich

Suchanfragen in DeReKo

du und dein Wagen

<lemma> Pronpers nom/N'nom und dein(e)/sein(e)/ihr(e)/euer(eure)... N"nom

<variants 1> Pronpers nom/N'nom [schon wieder/immer] mit deinem/seinem/deiner/seiner/seinen/ihrem... N"nom

<variants 2> Pronpers nom/N'nom mit deinem/seinem/deiner/seiner/seinen/ihrem... N"nom [schon wieder?!]

(MORPH(N) OR MORPH(PRON per irr)) "und" (&sein OR &ihr OR &dein OR &euer) MORPH(N nn)

(MORPH(N) OR MORPH(PRON per irr)) mit (&sein OR &ihr OR &dein OR &euer) MORPH(N nn)

(MORPH(N) OR MORPH(PRON per irr)) ((schon wieder) OR immer) mit (&sein OR &ihr OR &dein OR &euer)
MORPH(N nn)

wenn er immer mit seinen Kumpels rumhängt.

rd das Schloss immer mit seinen Augen sehen.“ Pe

s wollt ihr schon wieder mit euren Forderungen?

Du immer mit deinen Designerklamotten."

- und das nennt sich N - [Und] so etwas/so eine(r)/so was/das nennt sich Nnom[!/?] / [Und] so etwas/so eine(r)/so was/das schimpft sich Nnom[!/?] / das / so etwas nenne ich X[!] - [link](#)
- was heißt schon - Was heißt/hieß [schon/hier] X [?/!] - [link](#)
- Deutschland ist nicht USA - N ist nicht / kein N' - [link](#)
- da hast du N - Da hast du [dein/deinen/deine] / Da haben Sie [Ihr/Ihre] / Da habt ihr [eure/euer/eure] / Da haben wir / da habe ich NAkk [!] - [link](#)

Sketch Engine

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



Sketch Engine

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

Free Trial <https://auth.sketchengine.eu/#register/form?form=trial>

Sketch Engine: CQL

CONCORDANCE

Russian Web 2011 sample (ruTen

BASIC

ADVANCED

ABOUT

Query type ?

simple

lemma

phrase

word

character

CQL

CQL

[lemma="book"] [] {1,3} [tag="

Insert

[]

{ }

< >

" "

&

CQL BUILDER

⌵

A user-friendly tool for writing CQL queries.

lemma



X hin, X her.
(Irrelevanz, Einräumung)
Ex.: Lüge hin, Lüge her!

[lemma="hin"] [tag="SYM.*.Comma.*"]
[lemma!="und|oder"] [lemma="her"]

Zivilklausel	hin, Zivilklausel her	. Umge
rein, Deko	hin, Deko her	– muss
en – Laptop	hin, Laptop her	, die m
p> Gestank	hin, Milliardenkosten her	– auf d
> Eigentum	hin, Rechtsstaatlichkeit her	: Es sta
Mainstream	hin, Frühlingsgefühle her	! Der M
en – Prinzip	hin, Prinzip her	. Schlie
. Gutscheine	hin, Gutscheine her	(der ka
ohnäpchen	hin, Bierchen her	Dieses
ch nun bald	hin, bald her	im Sch



Da hast du [dein/deinen/deine] N Akk

(Vorweisen, Skepsis)

Ex.: Da habt ihr eure Krise!

[lemma="da"] [lemma="haben" &
tag="VFIN.*.2.*"] [tag="PRO.Pers.*.2.*"]
[lemma_lc="deine"]

normal token

[lemma="da"]

+

normal token

[lemma="haben" & tag="VFIN.*.2.*"]

+

normal token

[tag="PRO.Pers.*.2.*"]

+

normal token

[lemma_lc="deine"]

lle Bilder!	Da hast du deinen	Akkur
s Gesetz,	da hättest du deiner	Zunge
d töteten.	Da hast Du Deine	Schul
gerlicher -	da hast du deinen	Käs g
agte: "Da,	da hast du deinen	Weink
ug. Paff! "	da hast du deine	lächer
deutet ..."	Da hast Du Deinen	eigen
gap/> <p>	Da hast du dein	Baby
gap/> <p>	Da hast du dein	Bike s

Pron pers nom/N'nom und sein(e)/ihr(e)/euer(eure) N"nom

(Verärgerung, mangelnder Respekt)

Ex.: Charly und ihre Sturheit. Du und Deine Sorgen.

[tag="SYM.*"] [tag="N.Name.Nom.*" | tag="N.Reg.Nom.*"] [lemma="und"] [lemma_lc="seine|ihre|eure|deine"]
[lemma="N.*"] [tag="SYM.*"]



- und das nennt sich N - [Und] so etwas/so eine(r)/so was/das nennt sich Nnom[!/?] / [Und] so etwas/so eine(r)/so was/das schimpft sich Nnom[!/?] / das / so etwas nenne ich X[!] - [link](#)
- was heißt schon - Was heißt/hieß [schon/hier] X [?/!] - [link](#)
- Deutschland ist nicht USA - N ist nicht / kein N' - [link](#)
- da hast du N - Da hast du [dein/deinen/deine] / Da haben Sie [Ihr/Ihre] / Da habt ihr [eure/euer/eure] / Da haben wir / da habe ich NAkk [!] - [link](#)

Parallelkorpora in Sketch Engine

CREATE CORPUS



[CREATE CORPUS](#) > [ALIGNMENT](#) > [UPLOAD DATA](#) > [COMPILE](#)



The quality of automatic alignment depends heavily on the input data, mainly on the layout:

- documents containing only text in one column produce best result.
- documents with complex design such as advertisements, promotional leaflets or posters may produce poor result.

Unusual layout features may interfere with the alignment algorithm so much that no satisfactory result can be achieved, and aligning the documents manually is the only option.

Source language

German



Target language

Polish



Source corpus name

De-Pl_Test_Corpus, German

Target corpus name

De-Pl_Test_Corpus, Polish

CONFIRM

RESET



Choose a file or drag it here.

You can upload: .doc, .docx, .htm, .html, .pdf, .txt



Choose a file or drag it here.

You can upload: .doc, .docx, .htm, .html, .pdf, .txt

DWDS

<https://www.dwds.de/r>



Suchoperatoren

@	Wortform: @Buche
{ }	Alternative Wahl: @{hast,habt}
\$p=	Wortart: \$p= NE
#<1	Abstand weniger als: #<1

Da hast du [dein/deinen/deine] N Akk

(Vorweisen, Skepsis)

Ex.: Da habt ihr eure Krise!

"da @{hast,habt} @{du,ihr}

@{dein,deine,deinen,euer,eure,euren}"

- 2: Kant, Hermann: Die Aula, Berlin: Rütten & Loening 1965, S. 285
" Ja, und **da hast du deine** Riesenklappe gehal
- 3: Walser, Martin: Halbzeit, Frankfurt a. M.: Suhrkamp 1997 [1960]
Da hast du deine Beweiskette, sagte Edmund,
- 4: Walser, Martin: Halbzeit, Frankfurt a. M.: Suhrkamp 1997 [1960]
Erst nach zwei Stunden hatte er ein Einsehen u
die Firma Mayer und Co. zu machen; der Chef s:
Co. gleich groß, **da hast du deine** fünf und.
- 5: Koeppen, Wolfgang: Das Treibhaus. In: ders., Drei Romane, Fra
»**Da hast du deine** Politik und deine Händel, ur

"da @{hast,habt} @{du,ihr} @{dein,deine,deinen,euer,eure,euren}"

X hin, X her.

(Irrelevanz, Einräumung)

Ex.: Lüge hin, Lüge her!

"hin \, #<1 her"

"hin \, #<1 her"

- 1: Berliner Zeitung, 05.11.2005
Steppenpieper **hin**, Steppenpieper **her**: l
- 2: Berliner Zeitung, 28.09.2005
Müdigkeit **hin**, Müdigkeit **her** - es gibt Vo
- 3: Berliner Zeitung, 27.08.2005
2 MoMA **hin**, Goya **her**:
- 4: Berliner Zeitung, 17.08.2005
"Staub **hin**, Staub **her**, na und?"
- 5: Berliner Zeitung, 11.07.2005
Ball **hin**, Ball **her**

Pron pers nom/N'nom und sein(e)/ihr(e)/euer(eure) N''nom

(Verärgerung, mangelnder Respekt)

Ex.: Charly und ihre Sturheit. Du und Deine Sorgen.

"\$p=\$. \$p= NE und \$p= PPOSAT \$p= NN"

Korpusbelege Berliner Zeitung (1994–2005)

"\$p=\$. \$p= NE und \$p= PPOSAT \$p= NN"

: Berliner Zeitung, 28.12.2005

Foto: **Michelle und ihr Lunchmobil** am Strand von Honolulu.

: Berliner Zeitung, 21.12.2005

Das heißt: **Fritsche und seine Kollegen** sind in einem Hotel unfähig, wichtige Termine verlassen.

: Berliner Zeitung, 09.11.2005

Trotzdem: **Coenen und seine Mitarbeiter** waren auch damals

Niveau 3

Suchanfragen in:

- Reguläre Ausdrücke in Sketch Engine
- Python für Frequenzlisten



Python

High-level programming language

Google colab

Niveau 3: Phrasemkonstruktionen

- Kind ist Kind - X' ist/sind [immer] X' [link](#)
- Fragen über Fragen (Gründe über Gründe) - N' über N' - [link](#)
- Tod bleibt Tod - X'nom bleib(s)t/bleibe/bleiben X'nom - [link](#)
- X ist nicht gleich X - X'nom ist/sind nicht gleich X'nom - [link](#)
- Ring der Ringe - N' der N'gen pl / N' aller N'gen pl - [link](#)
- X hin oder her - X hin oder her / X' hin, X' her / X hin, Y her [link](#)

[PhK Liste](#) (nach Niveau geteilt)

Sketch Engine: Reguläre Ausdrücke für Wiederholungen

<https://www.sketchengine.eu/guide/regular-expressions/#toggle-id-2>

Ngram-Tool supports backreferences

PhK: Fragen über Fragen

([A-Z][^]*) über \1

	N-gram	Frequency ?
1	Feld über Feld	10 ...
2	Daten über Daten	8 ...

Left context

KWIC

Right context

It. </s><s> In dem **Feld über Feld** 3 jedes Exemplars

N-GRAMS

EUR-Lex 2/2016 parallel – German



BASIC

ADVANCED

ABOUT

N-gram length ?

2

3

4

5

6

Attribute ?

word

Frequency min ?

5

Frequency max ?

0

☐ Nest n-grams ?



☐ A = a ?



Subcorpus ?

none (the whole corpus)

☐ Key n-grams ?

Additional criteria ?

starting with letters

ending with letters

containing letters

starting with word

containing word

ending with word

matching regular expression

([A-Z][^]*) über \1

Insert

[]

{ }

|

^

Textdaten aus Dereko

Aktuelles Archiv: TAGGED-T - Archiv morphosyntakt. annotierter Korpora (TreeTagger)

Aktuelle Suchanfrage:

Treffer:

Archive Korpus **Suchanfragen** Wortform. Ergebnisse

Eingabe:

MORPH(N) über MORPH(N)

Bearbeiten mit Python

1	:	Fragen	-	10
2	:	Chancen	-	2
3	:	Zahlen	-	2
4	:	Hundertern	-	1
5	:	Umwege	-	1

Vielen Dank für Ihre Aufmerksamkeit!

kroelebor@gmail.com

elena_krotova@inbox.ru

[PhK Liste](#) (nach Niveau geteilt)