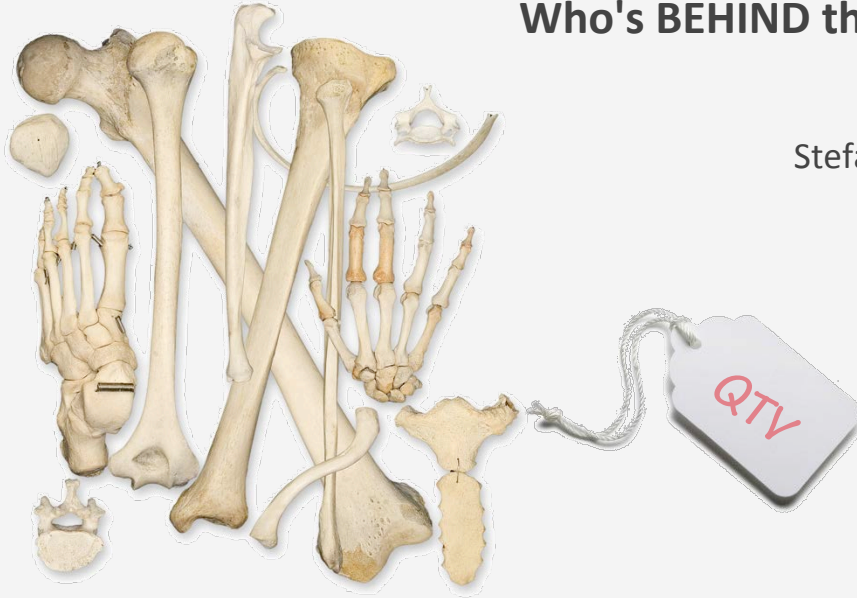
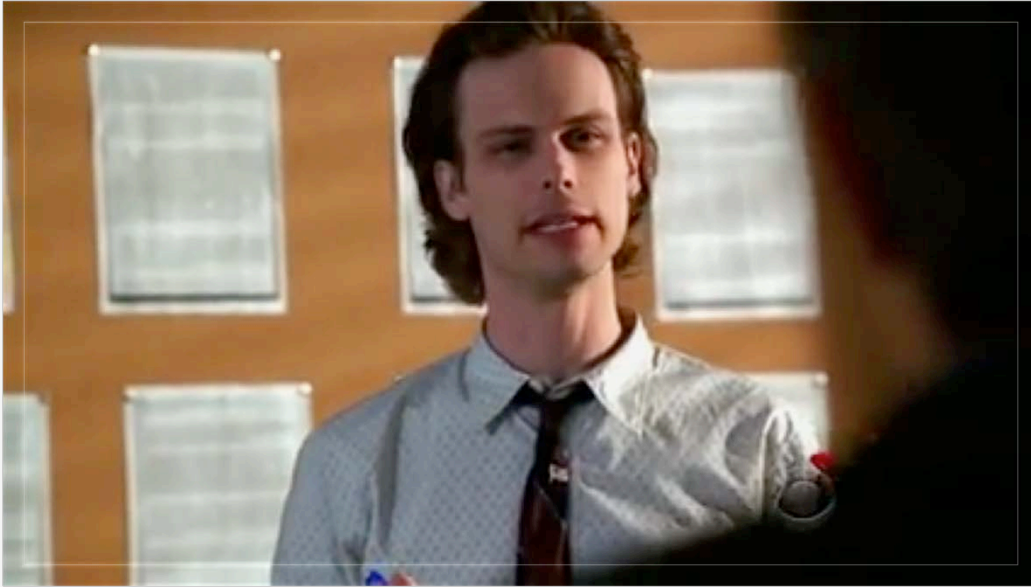


Linguistischer Workshop: Who's BEHIND the bad guys? – QTV auf dem Seziertisch der forensischen Linguistik

Stefan Diemer, Universität des Saarlandes



Intro



<https://www.youtube.com/watch?v=hE0-UyXSdXc>



- Forensische Linguistik
- Was ist anders beim QTV?
- Die Tools
- Beispiele
- Your turn
- Fazit





- Detail vs. Überblick
- Extrapolieren vs. Generalisieren
- Linguistik vs. Literatur...



-

$$f_{\mu_0} \sim \frac{\bar{x} - \mu_0}{s^2 + \dots}$$



-
- A close-up photograph showing a hand holding a magnifying glass over a printed document. The magnifying glass is positioned over a paragraph of text, making the words appear larger and clearer. The text visible through the lens includes phrases like "either mechanical or co", "typeface (such as Arial. P", "screen letters to avoid crowding", "seriously dark and should be an", "the figure. Point size should", "out the figure labels are 12-point", "uple, if axis labels are 12-point", "than 8 points, the minimum ac", "type also affects legibility. Ev", "and become less legible w", "letters generally are easier", "quires several distinct", "is acceptable". The background text, which is out of focus, discusses topics like "white paper", "output", and "legibility".



typisch

Sprache

Story

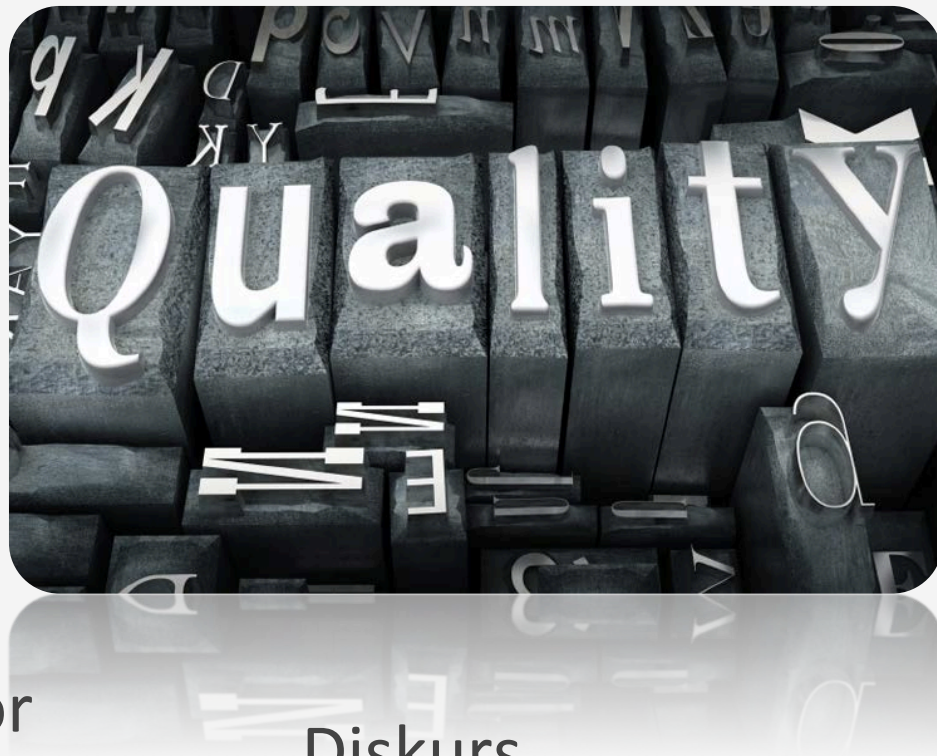
Adaption

künstlerisch

Thematik

Humor

Diskurs



Was ist unterschiedlich?



- Lexik
- Pragmatik
- Syntax
- Diskurs
- Semantik





- Eine Folge General Hospital hat 76,6 Wörter pro Minute, eine Folge Gilmore Girls hat 185 Wörter pro Minute (normale Sprechgeschwindigkeit ist 100-120 WpM)



Gilmore girls

Wortdichte



- In den ersten beiden Folgen der Sopranos kommt 87 Mal eine Variante von F*** vor - das ist einmal pro Minute (jedes 131. Wort)



THE
Sopranos

Fluchen



- In einem normalen Gespräch sind etwa 20% der Wörter seltene und akademische Begriffe, in The Bold and the Beautiful zum Beispiel sind es 19% (0% akademisch); in einer Folge The Big Bang Theory sind es 33%



Seltene Wörter



- In Staffel 3 von How I Met Your Mother kommen 119 Neologismen vor (Thewes 2014), das sind im Durchschnitt sechs Neologismen pro Folge, was bedeutet, dass alle vier Minuten ein neues Wort eingeführt wird

how i
met your
mother



Neue Wörter



- Jane Espenson hat als Drehbuchautorin zahlreiche QTV-Serien sprachlich geprägt. Mittels forensischer Analyse kann ihr Schreibstil von West Wing über Gilmore Girls bis hin zu Game of Thrones verfolgt werden.





- Eines ihrer Markenzeichen sind schnelle, überlappende Dialoge mit Einsatz von Diskursmarkern wie „well“

The screenshot displays the AntConc 3.4.3m (Macintosh OS X) 2014 interface. The main window shows a concordance search for the word "Well" in the file "GilmG.txt". The search results are displayed in a list format, with line numbers on the left and the corresponding text on the right. The word "Well" is highlighted in blue in the original image. The interface includes a menu bar at the top with options like "Corpus Files", "Concordance", "Concordance Plot", "File View", "Clusters/N-Gram", "Collocate", "Word List", and "Keyword List". A sidebar on the right shows a list of files, with "GilmG.txt" selected. At the bottom right, there is a "Window Size" button and a "Clone Results" button.

Line	Text
1	ly the same you know. MAX: Well, take it. LORELAI: Oh
2	ell me about school. RORY: Well, my French final went p
3	the thing is, if things go well, the way I hope they're
4	Well good. I'm glad. RORY: Well, better go. (gets up to
5	ou up to tonight. LORELAI: Well, I'm going to do a litt
6	AI: Ooh how sordid? RORY: Well, it's not the Rich Jame
7	o go towards fridge) RORY: Well, Madeline and Louise, F
8	is is so not you. LORELAI: Well, you don't know everyth
9	rules. RORY: Fine! Great! Well, maybe it's time for th
10	anymore, that's all. RORY: Well, it worked. PARIS: Yec
11	Lorelai Gilmore. LORELAI: Well, as long as we got that



- Korpusanalyse-Programme
- Konkordanzen
- Wortlisten
- Wortanalysetools
- Tagger
- Wortfeld-Datenbanken
- Referenzkorpora
- Webkorpus-Tools

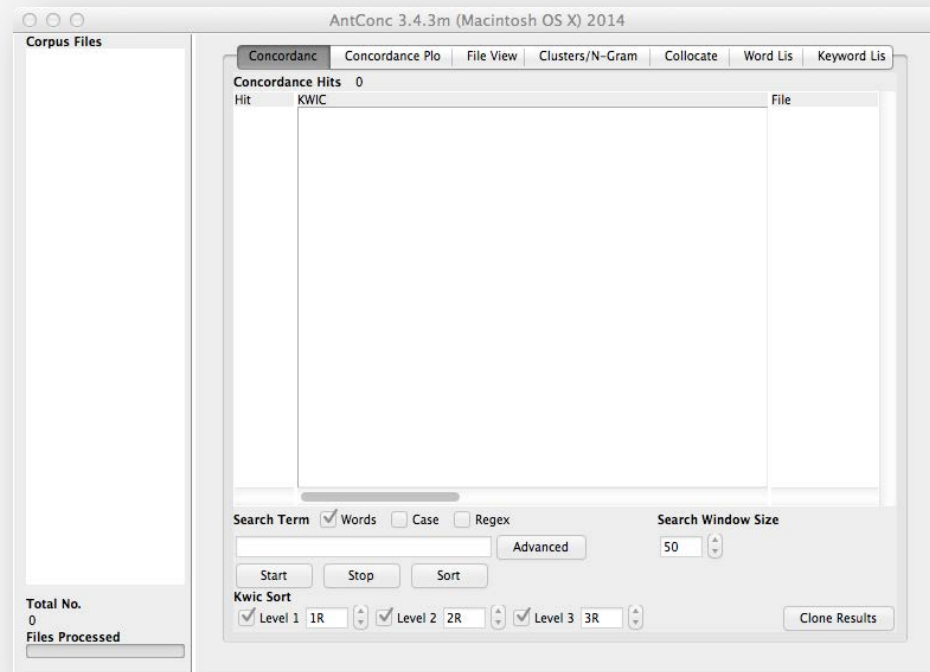




AntConc

- Erlaubt die einfache Umwandlung von Skripten in Korpora und deren Analyse
- Freeware für Windows, Mac OS X und Linux
- Alternative: WordSmith (kostenpflichtig)

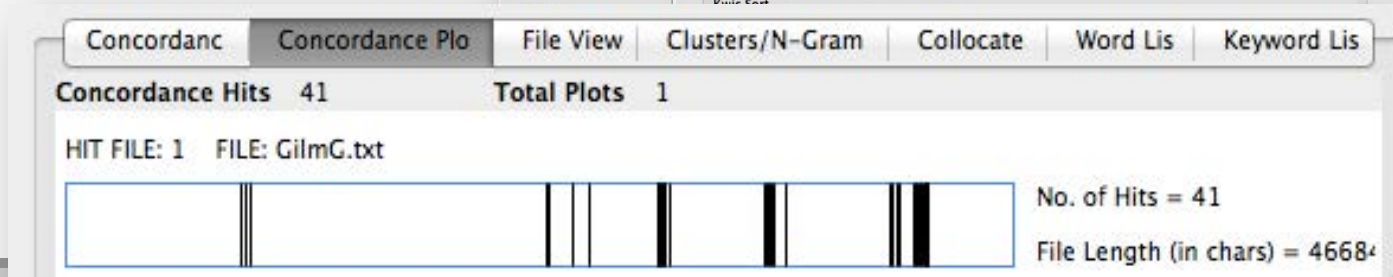
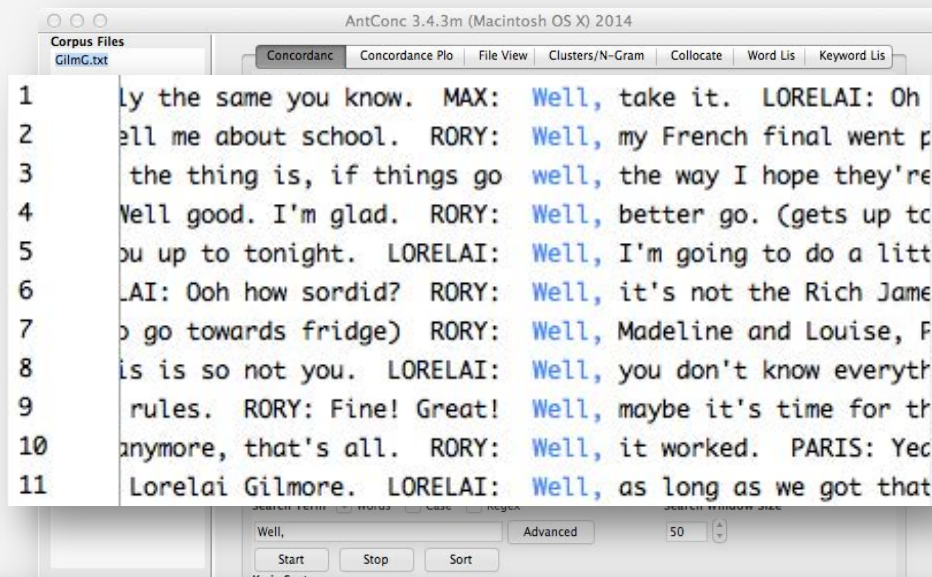
(http://www.laurenceanthony.net/antconc_index.html)



Konkordanzen



- Erlaubt Analyse von Wörtern und Phrasen im Kontext
- Möglichkeit, den Zeitpunkt und Kontext von Äußerungen zu analysieren
- Integriert in die meisten Korpusanalyse-Programme (wie WordSmith oder AntConc)





- Erlaubt die Untersuchung des Vokabulars nach Häufigkeit
- Möglichkeit der thematischen Zuordnung von Texten
- Integriert in die meisten Korpusanalyse-Programme
- Beispiel: Gilmore Girls, The Sopranos

Word Types: 1383		Word
Rank	Freq	Word
1	347	lorelai
2	344	i
3	321	you
4	262	rory
5	221	to
6	218	the
7	172	s
8	154	it
9	151	a
10	148	max
11	140	and
12	117	that
13	110	t
14	78	not
15	75	m
16	75	me

Word Types: 1298		Word T
Rank	Freq	Word
185	5	long
186	5	never
187	5	ones
188	5	pay
189	5	phone
190	5	serge
191	5	somebody
192	5	sorry
193	5	stop
194	5	sure
195	5	talk
196	5	though
197	5	too
198	5	wanna
199	5	where
200	5	yo





- Erlaubt die Einordnung von Wörtern und den Vergleich mit anderen Korpora
- Online verfügbar, z.B. <http://www.wordandphrase.info/> (Beispiel: The Big Bang Theory)

WORD AND PHRASE . INFO

FREQUENCY LISTS - ANALYZE TEXTS - ALL GENRES - ACADEMIC

DAVIES | BYU | COCA

LOG OUT **HELP**

ENTER TEXT BELOW -SAMPLES- MY TEXTS

Leonard: That'll be all. That was fun.
Raj: How long can he keep this up?
Leonard: I heard about this professor at MIT who melted down, bought a van, and spent the rest of his life as a mobile dog groomer.
Raj: He never went back to the university?

☒ WORD ☐ PHRASE

Select individual words in the text to see "word sketches"

SEE LISTS

FREQ RANGE	1-500	501-3000	> 3000	ACAD
1184 WORDS	59 %	10 %	31 %	2 %

Scene: The apartment. Sheldon is stood in the middle of the room. His whiteboard is behind him. Every few moments he turns round suddenly.

Penny: Whatcha doing ?

Sheldon: I'm attempting to view my work as a fleeting peripheral image so as to engage the superior colliculus of my brain .

Penny: Interesting. I usually just have coffee. You've been up all night ?

Sheldon: Is it morning ?

Penny: Yes .

Sheldon: Then I've been up all night .

CLICK TO SEE ENTRIES FOR **PERIPHERAL (EXACT)** **ADJ (8508)** PHRASE (HELP)

SYNONYMS (click to see) [?]

minor	
3478	secondary
6176	marginal
8508	peripheral
8434	superficial
12572	incidental
13122	unimportant
27433	tangential
outlying	
6176	marginal
8508	peripheral
10050	exterior
15768	outlying
39291	bordering

PERIPHERAL (RANK 8508, FREQ 2367)

	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
CLICK BAR TO LIMIT					
STORED	15	18	30	17	122
MORE	122	231	468	170	1376

DEFINITIONS (WORDNET) (BAD ENTRY?)

1. on or near an edge or constituting an outer boundary **2.** related to the key issue but not of central importance

COLLOCATES (click to see with PERIPHERAL)

noun vision, nerve, disease, area, blood, region, neuropathy, cell, issue, tumor, sheath, role **misc** vascular, central, malignant, such, nervous, rather, than, treat, mononuclear, due, main, total

CONCORDANCE LINES

GENRE	TEXT	WORD	CONTEXT
1	ACAD	that favor task-specific repetitive training , targeting both peripheral and central nervous systems (2) . # Even though the	
2	NEWS	known universe . It makes all those other places elsewhere seem peripheral and distant . # Brown was no visitor : He was virtually	
3	MAG	she came in fourth , the 47-year-old self-proclaimed " fat , peripheral and provincial " politician continues to fight against	
4	SPOK	system receptors . The opiates are clearly very important in the peripheral and the central nervous system experience of pain FLATOW	





- Tagger markieren die grammatische Funktion der Wörter in einem Skript
- Erlauben die gezielte Suche nach syntaktischen und phraseologischen Mustern, z.B. Verb + Präposition, Interjektionen, Formen von „to be“, Fragen...
- Online verfügbar, z.B. CLAWS <http://ucrel.lancs.ac.uk/claws/> (Uni Lancaster)
- Beispiel: The Big Bang Theory

```
Penny_NP0 :_PUN Whatcha_NP0 doing_VDG ?_SENT -----_PUN
Sheldon_NP0 :_PUN Im_VVB attempting_VVG to_TO0 view_VVI my_DPS work_NN1 as_PRP
a_AT0 fleeting_AJ0 peripheral_AJ0 image_NN1 so_AV0 as_AV0 to_TO0 engage_VVI
the_AT0 superior_AJ0 colliculus_NN1 of_PRF my_DPS brain_NN1 ._SENT -----_PUN
Penny_NP0 :_PUN Interesting_AJ0 ._SENT -----_PUN
I_PNP usually_AV0 just_AV0 have_VHB coffee_NN1 ._SENT -----_PUN
Youve_NN1 been_VBN up_AVP all_DT0 night_NN1 ?_SENT -----_PUN
Sheldon_NP0 :_PUN Is_VBZ it_PNP morning_NN1 ?_SENT -----_PUN
Penny_NP0 :_PUN Yes_ITJ ._SENT -----_PUN
```





- Erlauben den Vergleich von im Skript benutzten Begriffen und Phrasen mit anderen Korpusdaten, Wörterbüchern und Thesauri
- Informationen zu Register, Häufigkeit, Synonymen,
- Online verfügbar, z.B. <http://www.wordandphrase.info/>
- Beispiel: *peripheral* (The Big Bang Theory)

SYNONYMS (click to see) [?]

minor	
3478	secondary
6176	marginal
8508	peripheral
8434	superficial
12572	incidental

PERIPHERAL *j* (RANK 8508, FREQ 2367)

	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
CLICK BAR TO LIMIT					
STORED	15	18	30	17	122
MORE	122	231	468	170	1376

DEFINITIONS (WORDNET)

(BAD ENTRY?)

1. on or near an edge or constituting an outer boundary **2.** related to the key issue but not of central importance

COLLOCATES (click to see with PERIPHERAL)

noun vision, nerve, disease, area, blood, region, neuropathy, cell, issue, tumor, sheath, role **misc** vascular, central, malignant, such, nervous, rather, than, treat, mononuclear, due, main, total





- Ermöglichen den Vergleich von Quality-TV-Serien mit anderen Genres
- Referenzkorpora zu verschiedenen Varianten des Englischen, zu innovativer Sprache und New Englishes sowie zu TV-Serien
- Online verfügbar, z.B. <http://corpus.byu.edu/>

CORPUS OF CONTEMPORARY AMERICAN ENGLISH
450 MILLION WORDS, 1990-2012 [DOWNLOAD ALL 190,000 TEXTS]

DISPLAY: ☐ LIST ☒ CHART ☐ KWIC ☐ COMPARE

SEARCH STRING: []

WORD(S): []

COLLOCATES: []

POS LIST: []

SECTIONS:

1 IGNORE
SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

2 IGNORE
SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

CORPUS OF AMERICAN SOAP OPERAS
100 MILLION WORDS, 1990-2012

CORPUS OF GLOBAL WEB-BASED ENGLISH (GloWbE)
20 COUNTRIES / 1.9 BILLION WORDS [DOWNLOAD ALL 1,800,000 TEXTS]

DISPLAY: ☐ LIST ☒ CHART

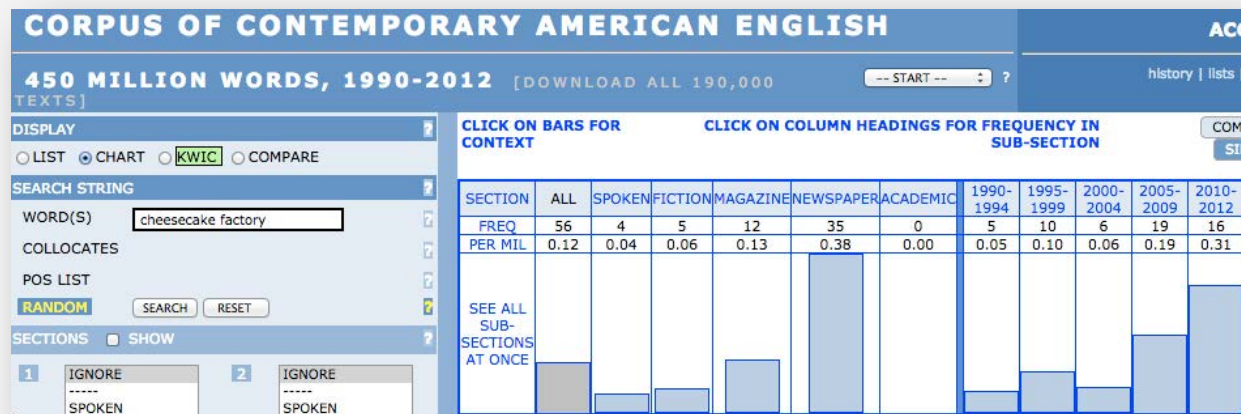
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK
1	HI	110193	16750	5786	24680	7747	12830	7011	7349	1345	1215	1420	5144	3213	4845	1705





- Varianten des Englischen: z.B. COCA und BNC
- Überprüfung von Wörtern und Phrasen auf Herkunft und Gebrauch im amerikanischen und britischen Englisch sowie regionale Varianten
- Online verfügbar: <http://corpus.byu.edu/coca/>
- Beispiel: *Cheesecake Factory* (The Big Bang Theory) im COCA





- Innovative Sprache und New Englishes: GloWbE
- Gebrauch von Wörtern in webbasierter Kommunikation und im internationalen Englisch, Neologismen
- Online verfügbar. <http://corpus.byu.edu/glowbe/>
- Beispiel: „wallah“, House of Cards

CORPUS OF GLOBAL WEB-BASED ENGLISH (GloWbE) ACCESS: 4/5

20 COUNTRIES / 1.9 BILLION WORDS [DOWNLOAD ALL 1,800,000 -- START -- : ?] history | lists | profile | logout

TEXTS

DISPLAY ☒ LIST ☐ CHART ☐ KWIC ☐ COMPARE

SEARCH STRING

WORD(S)

COLLOCATES

POS LIST

RANDOM SEARCH RESET

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

COMPARE

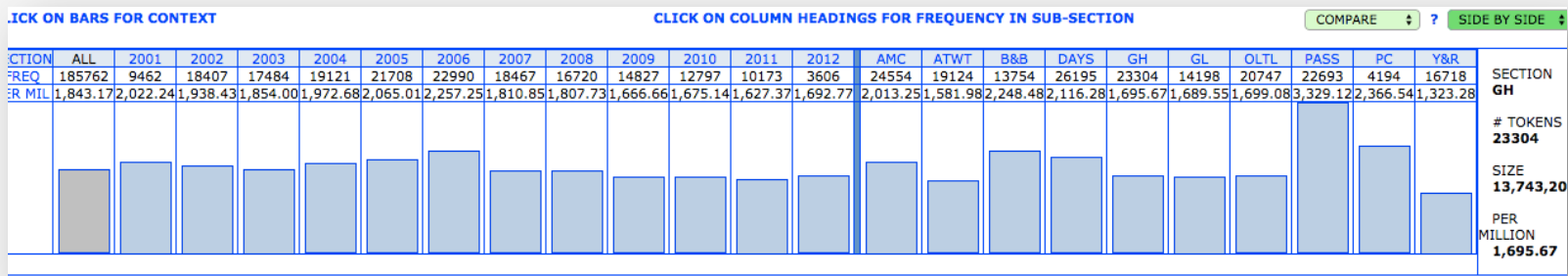
	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
1	WALLAH	172	10	39	3	15	7	26	2	37	11	1	7	2	1	2	1		2	3	3	

0.447 seconds



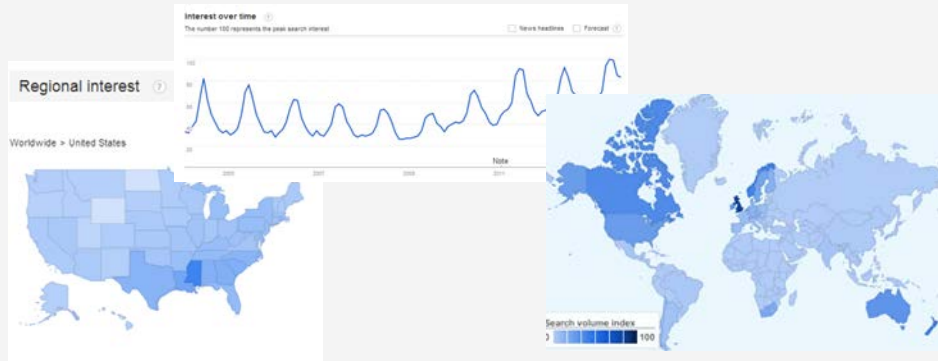


- TV-Serien: Corpus of American Soap Operas
- Direkter Vergleich zwischen Soaps (z.B. Days of our Lives, Bold and Beautiful, Passions, General Hospital) und QTV: Wortwahl oder -anzahl, Satzlänge, Komplexität...
- Online verfügbar: <http://corpus.byu.edu/soap/>
- Beispiel: *love*





- Google Trends: <http://www.google.de/trends/explore>
- Quantitative & graphische Darstellung von Google-Suchbegriffen
- Verknüpfung mit geographischen & diachronischen Daten

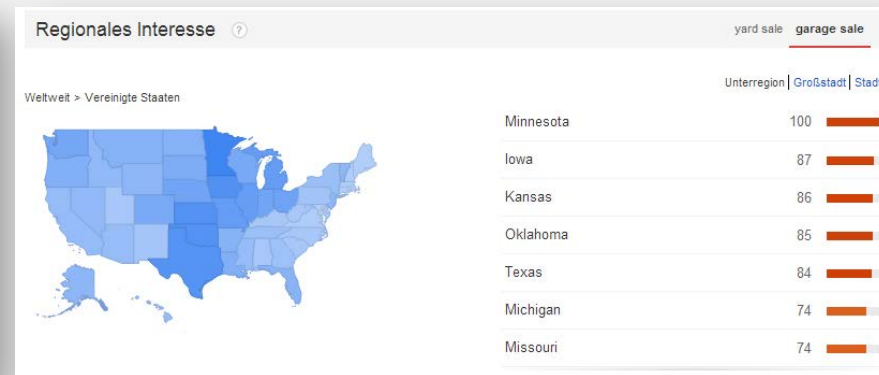
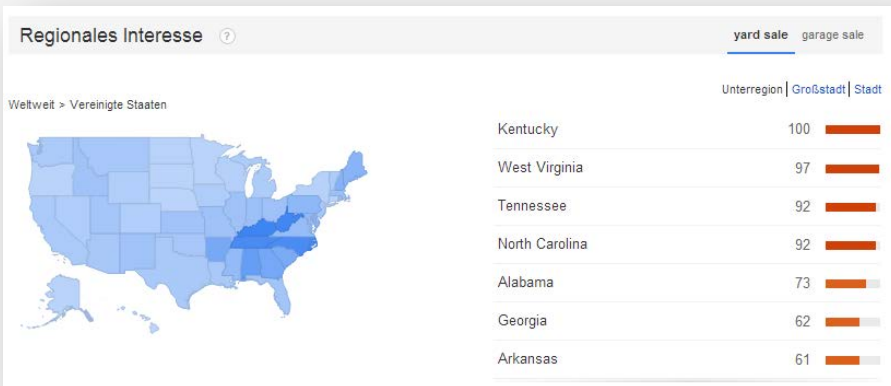


Google
Trends





Beispiel: regionale Varianten: Dialekte *yard sale* - *garage sale*





Beispiel: Beliebtheit und Rezeption *Quality TV*

Themen

Abonnieren



House o...

American Telev...

Breakin...

American Telev...

Game of...

American Telev...

How I M...

American TV Pr...

The Big ...

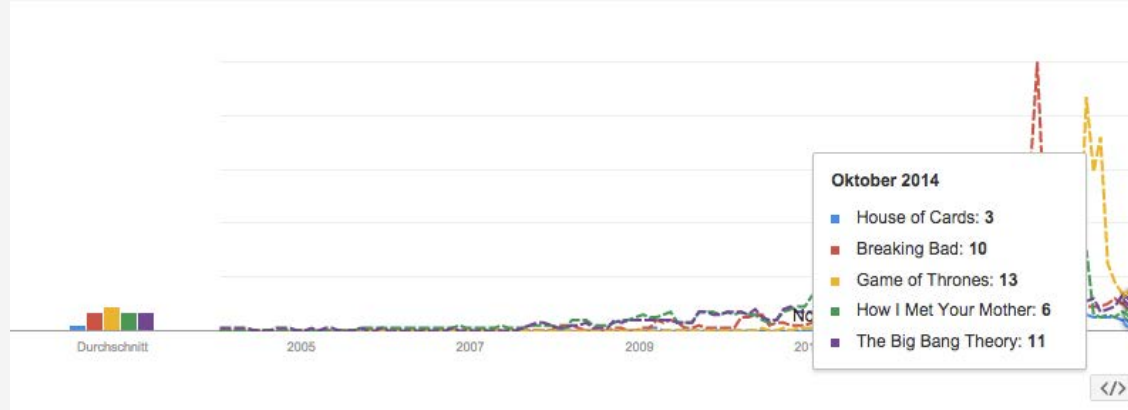
American TV Pr...

Beta: Mit dieser Betafunktion können Sie das Interesse an einem *Suchthema* abrufen und Sie erhalten schnell genaue Angaben zum gesamten Suchinteresse. Um das Interesse an einer bestimmten *Suchanfrage* abzurufen, wählen Sie die Option "Suchbegriff" aus. ?

Interesse im zeitlichen Verlauf ?

☐ Nachrichtenschlagzeilen ?

☐ Prognose ?



<http://www.google.de/trends/explore#q=/m/0h3rv9x.+m/03d34x8.+m/0524b41.+m/07zhji.+m/02r5qtm&cmpt=q&geo=US>





- Beispiel: Charakterbasierte Sprachmuster: Suche nach:
 - Lexikalisch (Vokabular, Wortlänge, Satzlänge)
 - Charakterbasiert (Humor, Stereotyp, Idiosynkrasien)
 - Syntaktisch (Satzteile, n-Grams, Satzstruktur)
 - Semantisch (Synonyme und Metaphern)





- Untersuchungsgegenstand? Lexik, Charaktere etc.
- Skript suchen oder kompilieren – die meisten QTV-Serien sind online verfügbar, oft jedoch in sehr unterschiedlicher Qualität und in variierendem Format.
- Tagging – ja oder nein?
- Auswahl der Korpusanalyse-Tools: Standard: Wortlisten und Konkordanzen
- Anwendung auf die Daten, ggf. Vergleich
- Visualisierung In Tabellen- oder Diagrammformat
- Interpretation der Variablen

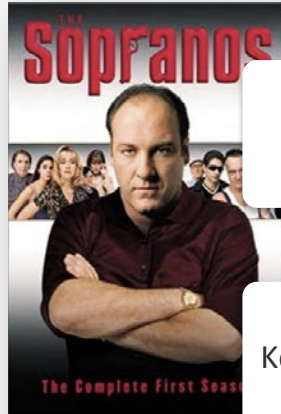




Beispiel: Lexikalische Analyse eines Skripts

The Sopranos, Season 1, Folge 2 (46 Long)

http://www.springfieldspringfield.co.uk/episode_scripts.php?tv-show=the-sopranos



1 Skript finden (oder kompilieren)



2 Qualität und Format evaluieren
und ggf. anpassen



3 Ins Korpusprogramm laden
(Rohversion)



4 Analysemöglichkeiten:
Konkordanz, Plot, Cluster, N-Gramm,
Kollokationen, Schlüsselwort...



Schwierigkeit des Skripts



Möglichkeit 1:

- Lexikalische Variation: STTR - Standard Type-Token Ratio:
 - Typen: Zahl der verschiedenen Wörter durch
 - Token: Gesamtwortzahl
- Vorteil: Berechnung erfolgt automatisch durch Korpusanalyseprogramm (z.B. WordSmith) oder leicht manuell.





Schwierigkeit des Skripts



Möglichkeit 2:

- Lexikalische Dichte: Zahl der Inhaltswörter verglichen mit Zahl der Gesamtwörter (inklusive Funktionswörter):
 - Nachteil: schwieriger zu ermitteln, da die Inhaltswortliste erst mittels einer „Stoppliste“ erstellt werden muss
 - Vorteil: genauer als STTR
- Beides Standardverfahren in der forensischen Linguistik, z.B. zur Plagiatserkennung
- Anwendungen: Textvergleiche, Didaktik, Segmentierung, Klassifizierung





Schwierigkeit des Skripts



Beispiele für lexikalische Variation

- STTR ist niedriger in gesprochener als in geschriebener Sprache (vgl. z.B. Chafe & Danielewicz 1987, Granger & Wynne 2000)
- Hohe STTR deutet auf informativen Stil hin, niedrige auf involvierten Stil (Biber 1991)
- Problematik: Abhängig von Textlänge und Art des Textes, aber nützlich für lange Skripte

TTR in Telefongesprächen

- 0,36-0,40 (Biber 1991, Granger 2013)

TTR in freien Reden

- 0,44 (Biber 1991)

TTR in Tageszeitungen

- 0,56-0.60 (Westin 2002)

Lexikalische Variation



Beispiel 1



Schwierigkeit des Skripts



The Sopranos, Season 1, Folge 2 (46 Long): Lexikalische Variation (und Vergleiche)

The Sopranos

• 0,42

Bold and
Beautiful

• 0,31

Gilmore Girls

• 0,41

Concordanc	Concordance Plo	File View
Word Types: 430	Word Tokens: 1008	



Schwierigkeit des Skripts



Beispiele für lexikalische Dichte

- Lexikalische Dichte variiert stark je nach Texttyp und Register; typisch nach Stubbs (1996) ist z.B.

Belletristik

- 40-54%

Fachtexte

- 40-65%

Gesprochene Sprache

- unter 40%

Alltägliche Dialoge

- 23-34%

Politische Reden

- 48%

Kommentare

- 58%

Lexikalische Dichte



Beispiel 1



Schwierigkeit des Skripts



Beispiele für lexikalische Dichte: berechnet mit Wortliste

Word Types: 1298		Word
Rank	Freq	Word
185	5	long
186	5	never
187	5	ones
188	5	pay
189	5	phone
190	5	serge
191	5	somebody
192	5	erry

Alltägliche Dialoge

- 23-34%

Gespr. Sprache

- unter 40%

The Sopranos 1.02:

- 54%

Fachtexte

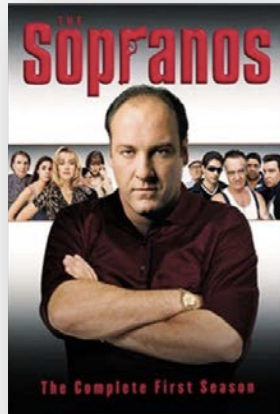
- 40-65%

200 5



Inhaltliche Schwerpunkte eines Skripts

- Nur ca. 200 Wörter, die zehnmal oder mehr vorkommen: einfache Analyse von inhaltlichen Schwerpunkten durch Eliminierung der Funktionswörter („Stopliste“)



know	oh	out
fuckin	like	tell
all	good	said
gonna	call	shit
but	how	these
right	mother	time
ma	some	ton
yeah	now	from
come	take	gotta
do	did	uncle
get	have	want
go	one	car
just	when	give
fuck	hello	home
so	his	make
no	let	okay



Schwierigkeitsgrad

- Wordandphrase.info: Vokabular nach Genre einteilen
- Sheldon: The Big Bang Theory 3.14



ACADEMIC

4: atoms, electrons
 3: pattern, structure
 2: engage, figure, table, theory
 1: acceptable, basal, classes, cognitive, consequences, contribute, database, degradation, determine, devaluation, fermions, fracture, function, ganglia, impaired, interference, intolerance, labouring, m, molecular, monuments, necessary, neutrons, particle, pathways, peripheral, phenomenon, plural, polymer, protons, ratio, regarding, relativity, sensitivity, similarly, simultaneously, sufficiently, task, terms, theoretical

FREQ RANGE	1-500	501-3000	> 3000		ACAD
971 WORDS	64 %	10 %	26 %		5 %

Sheldon (appearing in an **apron** and **carrying a tray**): Behind you .

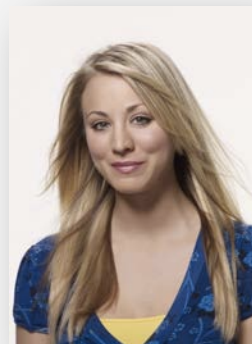
Sheldon (**drops tray**. A **nearby table claps**): Is that really **necessary**? Good Lord. The **interference pattern** in the **fracture**. The **motion** of the **wave** through the **molecular structure**. I've been looking at it all **wrong**. I can't consider the **electrons** as **particles**. They move through the **graphene** as a **wave**. It's a **wave**! The moment to **applaud** would be now. **Troglodytes** .

Sheldon (**popping his head up**): Bazinga. (Disappears, **pops up** in another place) Bazinga .(And again) Bazinga. (And again) Bazinga. (And again) Bazinga .



Schwierigkeitsgrad

- Wordandphrase.info: Vokabular nach Genre einteilen
- Penny: The Big Bang Theory 3.14



ACADEMIC

1: physics

FREQ RANGE

1-500

501-3000

> 3000

ACAD

245 WORDS

68 %

10 %

22 %

0 %

And you're **stuck** ?

Do I look **okay** ?

For me, the worst part was when people saw us leave with them .

Hey, guys, **sorry** you had to wait, but we are **swamped**. What's this ?

Hey, you know what, **Leonard**, I know I said I could **handle** your **roommate**, but I was **wrong**. We're going to have to break up .



Wiederkehrende Phrasen

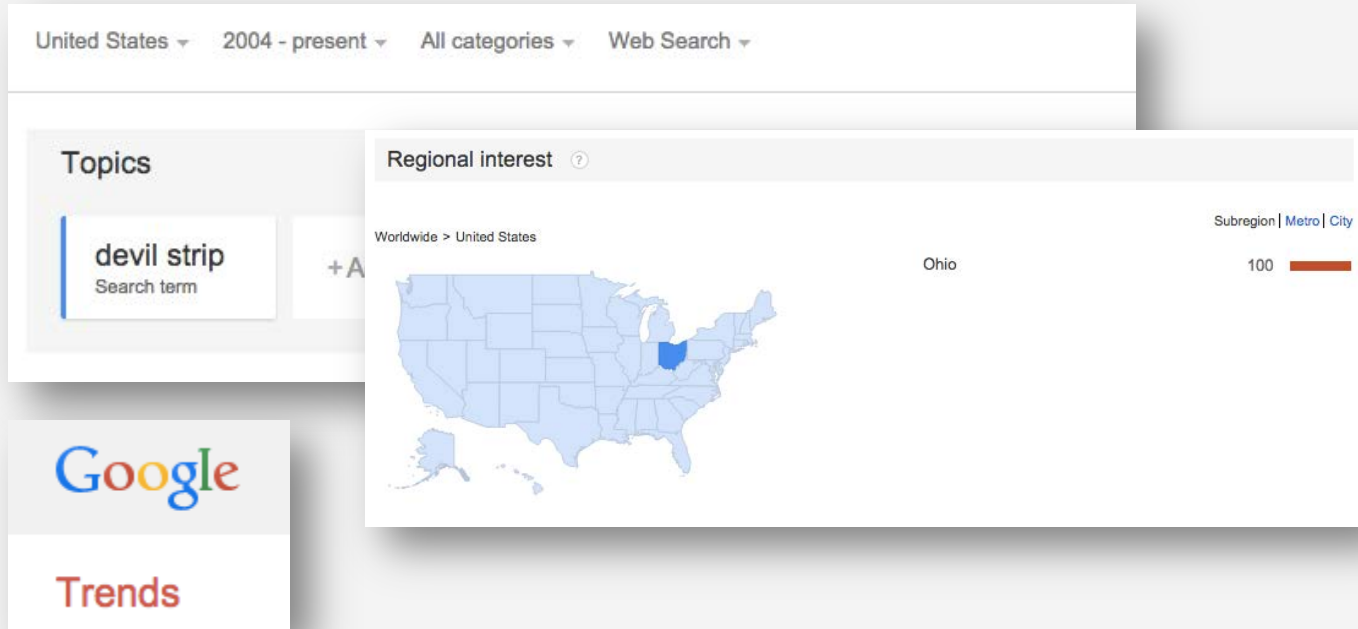
- N-Gram-Analyse:
- Sheldon: The Big Bang Theory 3.14



Concordance				Concordance Plot	File View	Clusters/N-Grams	Collocates
Total No. of N-Gram Types 987				Total No. of N-Gram Tokens 1006			
Rank	Freq	Range	N-gram				
1	3	1	bazinga and again bazinga				
2	3	1	knock knock knock leonard				
3	3	1	knock knock leonard and				
4	3	1	knock leonard and penny				
5	2	1	again bazinga and again				
6	2	1	and again bazinga and				
7	2	1	and penny knock knock				
8	2	1	as if they have				
9	2	1	i don t need				
10	2	1	if they have no				
11	2	1	leonard and penny knock				
12	2	1	penny knock knock knock				



- Google Trends für Dialekte: devil strip (Criminal Minds)





- Metaphern in House of Cards: n-grams



AntConc 3.4.3m (Macintosh OS X) 2014

Corpus Files
HoC1-1.txt

Concordance				Concordance Pio		File View		Clusters/N-Grams		Collocates		Word Lis		Keyword Lis	
Total No. of N-Gram Types 5818				Total No. of N-Gram Tokens 5876											
Rank	Freq	Range	N-gram												
60	1	1	a bird up in												
61	1	1	a bit don t												
62	1	1	a bit e												
63	1	1	a bit i												
64	1	1	a boy s												
65	1	1	a boy s												
66	1	1	a buffe												
67	1	1	a cab i												
68	1	1	a canir												
69	1	1	a chari												
70	1	1	a cheap												
71	1	1	a check												
72	1	1	a chick												
73	1	1	a city												
74	1	1	a commi												

Search Term ☒ Words ☐ Case ☐ R

a

Start Stop Sort

Sort by ☐ Invert Order Search Terr

Sort by Freq ☒ On Left

60	1	1	a bird up in
61	1	1	a bit don t
62	1	1	a bit early in
63	1	1	a bit irresponsible what
64	1	1	a boy scout nobody
65	1	1	a boy scout not
66	1	1	a buffer you mean
67	1	1	a cab i had
68	1	1	a canine hit and
69	1	1	a charity but not
70	1	1	a cheap ploy it
71	1	1	a check this is
72	1	1	a chicken shit move

Total No.
1
Files Processed



- I vs. you? n-grams

HOUSE
of CARDS



Concordance	Concordance Plo	File View	Clusters/N-Grams	Collocates	Word Lis
Total No. of N-Gram Types 4362		Total No. of N-Gram Tokens 5878			
Rank	Freq	Range	N-gram		
1	42	1	i m		
2	42	1	it s		
3	29	1	don t		
4	25	1	we re		
5	22	1	going to		
6	21	1	you re		
7	19	1	that s		
8	17	1	do you		
9	16	1	i ll		
10	15	1	i can		
11	14	1	i ve		
12	14	1	in the		
13	14	1	you want		
14	13	1	have to		
15	13	1	i don		

Search Term	☒ Words	☐ Case	☐ Regex	☒ N-Grams	N-Gram Size
					Min. 2 Max. 2
Start	Stop	Sort	Advanced		
Sort by	☐ Invert Order	Search Term Position			
Sort by Freq		☒ On Left ☐ On Right			



- I vs. you? concordance plot



HIT FILE: 1 FILE: HoC 1-1.txt



No. of Hits = 263

File Length (in chars) = 29797

No. of Hits = 232

File Length (in chars) = 29797



- Vergleich mit dem SOAP-Corpus



HOUSE of CARDS

CORPUS OF AMERICAN SOAP OPERAS

100 MILLION WORDS, 1990-2012



BRIGHAM YOUNG UNIVERSITY

ENTER



- Structural Analysis: Plot and Acts
 - > AntConc Concordance Plot, korreliert

> Acts (HoC, 1-1)

1 On the line

2 Betrayal

3 The visit

4 Intrigues

5 Scandal

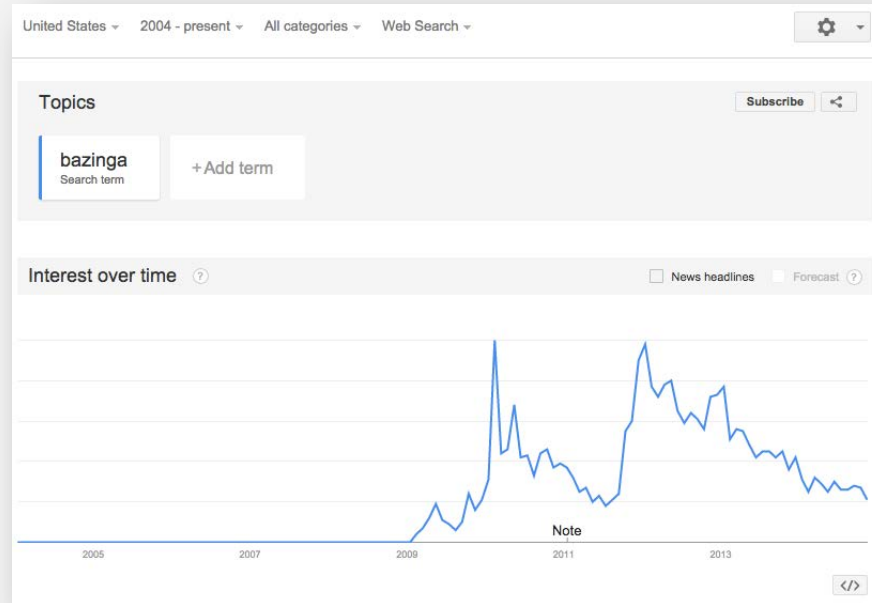
6 The leak

7 The morning





- Ausbreitung von Begriffen in andere Genres: Bazinga!



Google

Trends



- Ausbreitung von Begriffen in andere Genres: Bazinga!

CORPUS OF GLOBAL WEB-BASED ENGLISH (GloWbE)

20 COUNTRIES / 1.9 BILLION WORDS [DOWNLOAD ALL 1,800,000 TEXTS] -- START -- ? history | lists | profile | logou

DISPLAY 2
☒ LIST ☐ CHART ☒ KWIC ☐ COMPARE

SEARCH STRING 2
WORD(S)
COLLOCATES

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
1	BAZINGA	52	28	3	9		3	3						1		1	2			2		

0.783 second

gentlemen alike : read it ! # Bazinga at 11:44 am on July 5 , 2011 # Anyone aware of
ards Ryan . # Thanks for your BAZINGA back Mona ! I 'm just tired of the whole concept of
he Woeful Countenance . *Joe Bazinga Biden , Court Jester , and Comic Relief Appointee * and
season , complete with a total BaZinga finale ! Also a listener Live Report from the Royal Winter
ce of philosophy that sparks a BAZINGA moment . # Yawwwn . # Anyway . # I 'm off
says--MALARKY ! As you say , BAZINGA ! Have a good day and stay safe . # Sorry ,
ss fan of Dr. Sheldon Cooper . Bazinga ! Learn more # Exactly Fletcher ! A lot of people do



- Kontrastiver Ansatz (Übersetzung)

OPEN IN THE TOWN SQUARE

[Lorelai and Rory are walking down the sidewalk.]

RORY: It's culs-de-sac.

LORELAI: No way!

RORY: It is.

LORELAI: The plural of cul-de-sac is culs-de-sac?

RORY: Yes.

LORELAI: That doesn't even sound like English.

RORY: That's because it's French.

LORELAI: You know what I mean.

RORY: I hate to be the bearer of bad news.

LORELAI: Words should sound right to be right.

RORY: That's not how it works.

LORELAI: So, what, the plural of yo-yo is yos-yo?

RORY: Yeah, 'cause that sounds so natural.

LORELAI: As natural as culs-de-sac.

(Gilmore Girls, 4.10)

RORY: Es heißt culs-de-sac.

LORELAI: Ummöglich!

RORY: It is.

LORELAI: Der Plural von cul-de-sac is culs-de-sac?

RORY: Jep.

LORELAI: Das klingt nicht nach unserer Sprache.

RORY: Das ist Französisch.

LORELAI: Du weißt, was ich meine.

RORY: Tut mir leid, aber so ist es.

LORELAI: Richtige Wörter sollten richtig klingen.

RORY: Das kannst du aber nicht erwarten.

LORELAI: Gut. Und ist der Plural von yo-yo etwa yos-yo?

RORY: Nein, das klingt doch nicht natürlich.

LORELAI: Natürlicher als culs-de-sac.

(Gilmore Girls, 4.10)



<https://www.youtube.com/watch?v=6cWGzd21ENI>



- Syntax und Lexik – tagged

HOUSE 
of CARDS

I_PNP 'm_VBB feeling_VVG

hungry_AJ0

today_SENT ._PUN



Word Types: 1346		Word To
Rank	Freq	Word
1	1310	pun
2	868	sent
3	858	pnp
4	786	nn
5	361	av
6	328	at
7	320	prp
8	281	vvi
9	275	aj
10	264	to
11	263	i
12	243	vvb
13	232	you
14	229	np
15	180	dt
16	178	the



- Discourse markers, interjections

HOUSE of CARDS

Concordance Hits 89		
Hit	KWIC	File
1	-----_PUN Hey_ ITJ !_SENT -----_PUN What_	HoC 1-1 tagg
2	og_NN1 ._SENT -----_PUN Aw_ ITJ ,_PUN man_NN1 ._SENT ---	HoC 1-1 tagg
3	e_AV0 ._SENT -----_PUN Shh_ ITJ ._SENT -----_PUN Shh_IT	HoC 1-1 tagg
4	h_ITJ ._SENT -----_PUN Shh_ ITJ ._SENT -----_PUN It_PNF	HoC 1-1 tagg
5	r_NN1 !_SENT -----_PUN Oh_ ITJ ._SENT -----_PUN Presid	HoC 1-1 tagg
6	im_PNP ?_SENT -----_PUN No_ ITJ ._SENT -----_PUN Do_VDE	HoC 1-1 tagg
7	ng_VVG ._SENT -----_PUN Oh_ ITJ ,_PUN Jim_NP0 Matthews_N	HoC 1-1 tagg
8	r_NN1 ._SENT -----_PUN Huh_ ITJ ._SENT -----_PUN Linda_	HoC 1-1 tagg
9	nn_NN1 ._SENT -----_PUN No_ ITJ ._SENT -----_PUN We_PNF	HoC 1-1 tagg
10	NN1 ,_PUN I_PNP said_VVD no_ ITJ calls_NN2 ,_PUN not_XX0	HoC 1-1 tagg
11	nd_VVI ?_SENT -----_PUN No_ ITJ ._SENT -----_PUN Go_VVE	HoC 1-1 tagg
12	se_AV0 ._SENT -----_PUN Uh_ ITJ ,_PUN put_VVD him_PNP th	HoC 1-1 tagg
13	_NP0 ._SENT -----_PUN Ahem_ ITJ ._SENT -----_PUN Mr._NF	HoC 1-1 tagg
14	ou_PNP ._SENT -----_PUN Uh_ ITJ ,_PUN yes_ITJ ,_PUN we_F	HoC 1-1 tagg
15	-----_PUN Uh_ITJ ,_PUN yes_ ITJ ,_PUN we_PNP were_VBD ,	HoC 1-1 tagg
16	JN we_PNP were_VBD ,_PUN uh_ ITJ We_PNP were_VBD pleased	HoC 1-1 tagg





HOUSE of CARDS

- grammatical features, e.g. preposition stranding

1	ker_NP0 What_DTQ 's_VDZ he_PNP like_	PRP ?_SENT -----_PUN	Mr._NP0 preside
2	ey_NN1 I_PNP 'm_VBB upset_AJ0 about_	PRP ._SENT -----_PUN	It_PNP 's_VBZ t
3	c_DTQ you_PNP 're_VBB getting_VVG at_	PRP ._SENT -----_PUN	Kern_NP0 first
4	AVQ is_VBZ this_DT0 coming_VVG from_	PRP ?_SENT -----_PUN	You_PNP 're_VBB
5	nebody_PNI I_PNP can_VM0 talk_VVI to_	PRP ._SENT -----_PUN	We_PNP 're_VBB
6	DTQ we_PNP 're_VBB talking_VVG about_	PRP ._SENT -----_PUN	I_PNP protect_V
7	VB what_DTQ that_DT0 feels_VVZ like_	PRP ._SENT -----_PUN	_-PUN Do_VDB I_
8	hing_PNI I_PNP can_VM0 work_VVI with_	PRP ._SENT -----_PUN	Okay_AV0 ,_PUN
9	DTQ are_VBB we_PNP talking_VVG about_	PRP ?_SENT -----_PUN	Just_AV0 asking



- AutorInnen verfolgen: Jane Espenson (Buffy, West Wing, Gilmore Girls, Game of Thrones)



Concordance Hits 8	
Hit	KWIC
1	'clock for an 8 o'clock date? RORY: I don't know, maybe a Chilton teacher
2	s ready. How are the feet? LORELAI: I don't know. They stopped talking to
3	: What's been keeping you? LORELAI: I don't know - things. MAX: Uh huh.
4	s there something going on here that I don't know about? LORELAI: Yeah, u
5	ing up with me aren't you? LORELAI: I don't know. MAX: Not only are you
6	he hell is wrong with you? LORELAI: I don't know. RORY: You said you wer
7	is the first time I'd seen him and I don't know, there was just somethir
8	this was the man for you. LORELAI: I don't know. He might have been. Exc

2	e it's not your House. Where's Bran? I don't know. It's not my House. Robb
3	he gold is still mine. Where? Where? I don't know where, but when they fre
4	and of the King. Oh, I'm very sorry. I don't know anything about all that.



- What feature would you like to look at?

THE
Sopranos

HOUSE 
of CARDS


the **BiG**
BANG
THEORY

THE
WIRE

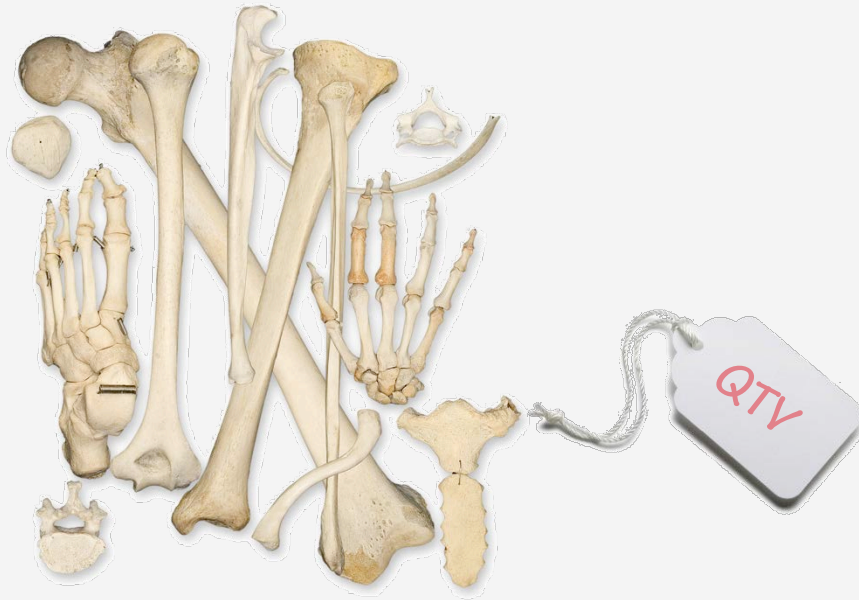
 **Br**eaking
 **Ba**d

SHERLOCK

Gilmore girls



Quality TV: Das Ergebnis der Autopsie





Merkmale:



- Catchphrases
- Authentische Sprache
- Variation
- Sprachliche Komplexität
- Wortdichte
- Inhaltliche Schwerpunkte
- Identität





To be continued...



s.diemer@mx.uni-saarland.de

Universität des Saarlandes, FR 4.3 Anglistik, Englische Sprachwissenschaft, Postfach 15 11 50, 66041 Saarbrücken



- Biber, Douglas. 1991. Variation across speech and writing. Cambridge University Press.
- Chafe, Wallace, and Jane Danielewicz. 1987. Properties of spoken and written language. Academic Press.
- Granger, Sylviane, Gaetanelle Gilquin, and Fanny Meunier. 2013. Twenty Years of Learner Corpus Research-Looking Back, Moving Ahead. Louvain: Presses universitaires de Louvain.
- Granger, Sylviane, and Martin Wynne. 2000. Optimising Measures of Lexical Variation in EFL Learner Corpora. In: Corpora Galore: Analyses and Techniques in Describing English: Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora (ICAME 1998). Rodopi: 249.
- Stubbs, Michael. 1996. Text and corpus analysis: Computer-assisted studies of language and culture. Oxford: Blackwell.
- Westin, Ingrid. 2002. Language change in English newspaper editorials (No. 44). Amsterdam: Rodopi.
- Google Trends at <https://www.google.de/trends/explore>
- Anthony, L. 2014. AntConc version 3.4.3w. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>



Illustrationen

Surgical instruments

<http://www.elroimedical.com/wp-content/gallery/surgical-instruments-and-supplies/T7-Surgical-Instruments.jpg>

Gestures

http://www.bbc.co.uk/schools/gcsebitesize/english/images/sp_gestures.jpg

Tape

<http://www.spotsify.com/wp-content/uploads/2013/07/sentence-tape1.jpg>

Stegosaurus

http://brantworks.com/resources/Stegosaurus/Shigeru_recast_kit/Stegosaurus-kit-skelparts-2000.jpg

Forensic linguistics

<http://lidahibu.com/wp-content/uploads/2012/03/forensic-linguistics.jpg>

Mortuary

http://3.imimg.com/data3/OO/QR/MY-196149/kwikot_commercial_hospital_mortuary_kmt3-4-250x250.jpg

Formula

<http://www.appmarketingagentur.de/wp-content/uploads/2013/02/formula-app-downloads.jpg>

QTV Logos

<https://stuartreviewsstuff.files.wordpress.com/2013/05/bbtlogo.jpg>

<http://www.trademarkia.com/logo-images/567990-company/the-sopranos-85804945.jpg>

<https://fanart.tv/fanart/tv/76568/clearlogo/gilmoregirls-76568.png>

http://img3.wikia.nocookie.net/_cb20140429172045/logopedia/images/f/f1/House-of-cards-2013-510db9807623c.png

http://upload.wikimedia.org/wikipedia/commons/4/48/Breaking_Bad_logo.png

http://upload.wikimedia.org/wikipedia/commons/8/8e/The_Wire_logo.png

http://upload.wikimedia.org/wikipedia/commons/c/c7/Sherlock_Logo.svg

Forensiv linguistics 2

http://www.all-about-forensic-science.com/images/forensic_linguist.jpg

Language

<http://www.iafl.org/images/home3.jpg>

Illustrationen



TV series

<http://zinemaniacos.files.wordpress.com/2012/06/tv-shows-completed1.jpg>

Tag

<https://mikebrandlyauctioneer.files.wordpress.com/2010/06/price-tag.jpg>

Bones

http://www.skullsunlimited.com/userfiles/image/category3_family_241_large.jpg

Lexis

http://wordstodeeds.files.wordpress.com/2012/08/istock_000015756375xsmall-copy.jpg

Scriptwriting

<http://www.davidshillitoe.co.uk/wp-content/uploads/2012/11/Quickfire-script-web2.jpg>

Meaning

<https://quantumshifting.files.wordpress.com/2012/04/meaning.gif>

Discourse bubbles

http://www.jweekly.com/images/uploads/w_photos_2012/zx01_06_12/Jcivilconversationbubbles.jpg

Breaking bad Blu-Ray box

<http://zeitgeist247.de/wp-content/uploads/2013/12/breaking-bad-komplette-serie-box-blu-ray.jpg>

Ruler

<http://www.blogtips.org/wp-content/uploads/2009/05/ruler-2.jpg>

Caliper

http://i00.i.aliimg.com/photo/v2/287687698/OUTSIDE_CALIPER.jpg

Sopranos cover season 1

<https://webshop.cashconverters.com.au/thumbnail/850x520/1611897-the-sopranos-season-1-0.jpg>

Difficulty

<http://jamiethedentist.com/images/difficulty.jpg>

Lupenbrille

http://www.lupenbrille.de/media/image/balupo3_schwarz.jpg

Illustrationen



Grünes Häkchen

<http://awo-bv-hannover.de/typo3temp/pics/fac21b2fee.jpg>

Gilmore Girls Paris and Rory

http://wwwcdn.channel5.com/assets/images/000/027/672/large_size_GilmoreGirls_S3ep01_1_640x360.jpg?1339408171

Tony Soprano cursing

http://1.bp.blogspot.com/-5PwWudQJUSY/UUj5aH8AOKI/AAAAAAAAAfl/tCwcd-tMM90/s1600/tony-soprano_2463138b.jpg

HIMYM image

<http://i2.cdn.turner.com/cnn/dam/assets/140327194124-himym-cast-horizontal-gallery.jpg>

TBBT Sheldon and Leonard

http://img3.wikia.nocookie.net/_cb20120227132707/bigbangtheory/de/images/5/51/Sittzplatz.jpg

HIMYM logo

http://img3.wikia.nocookie.net/_cb20131022234047/logopedia/images/b/b5/How-I-Met-Your-Mother-how-i-met-your-mother-2034398-2400-1800.jpg

Quality

http://www.nusconnect.org.uk/pageassets/campaigns/highereducation/student-engagement/annualqualityreports/header_quality.jpg

Yard Sale & Garage Sale

<http://1.bp.blogspot.com/-O1RzZKkOIPA/UBfP61R7kNI/AAAAAAAAAXk/dC4S46zcyKw/s1600/yard-sale-sign.jpg>

<http://blogs.ocweekly.com/navelgazing/yard-sale-sign.jpg>

West Wing logo

<http://www.innovationpov.com/wp-content/uploads/2011/07/West-Wing.jpg>

Game of Thrones logo

http://www.titancon.com/images/game_of_thrones_logo.jpg

Jane Espenson

http://img3.wikia.nocookie.net/_cb20120220050403/onceuponatime8042/images/5/58/InfoboxJE.jpg

Sheldon

http://img2.wikia.nocookie.net/_cb20120116013516/bigbangtheory/images/3/31/Sheldon.jpg

Illustrationen



Penny

http://img4.wikia.nocookie.net/_cb20100911101236/bigbangtheory/images/3/31/Penny.jpg

Chemie

<http://2.bp.blogspot.com/-VSMT6syCZPw/UKjJNy3N0hI/AAAAAAAAAcs/fqFSjksRFKo/s1600/glass+of+chemistry.JPG>

Gilmore Girls Skript 4.10

<http://www.crazy-internet-people.com/site/gilmoregirls/pages/s4/s4s/75.html>

Lorelai und Rory

<http://i.ytimg.com/vi/EnilueLPXhc/0.jpg>

House of Cards Episode 1

http://www.springfieldspringfield.co.uk/view_episode_scripts.php?tv-show=house-of-cards-2013&episode=s01e01

Frank Underwood

<http://theconversation.com/you-talking-to-me-house-of-cards-and-breaking-the-fourth-wall-23919>

formula

<http://artnaz.com/wp-content/uploads/2014/05/25-secrets-of-breaking-bad-1.jpg>

House of cards actors

<http://obsessiveviewer.files.wordpress.com/2013/05/the-cast-of-house-of-cards-netflix.jpg>