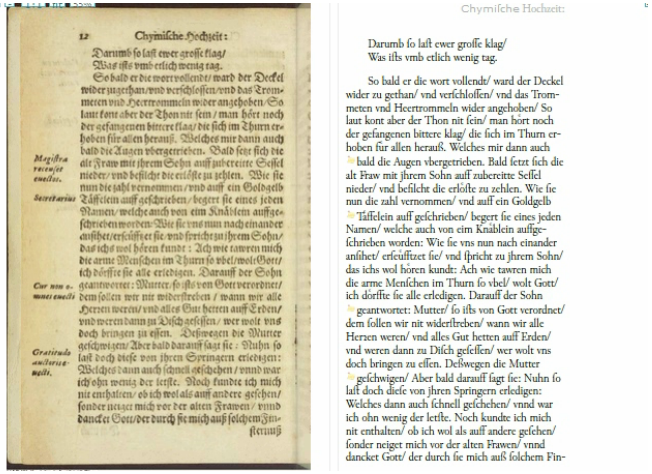
Korpuslinguistik und Morphologie

Zeitmaschinen...

Zeitmaschin

Zeitmaschinen…

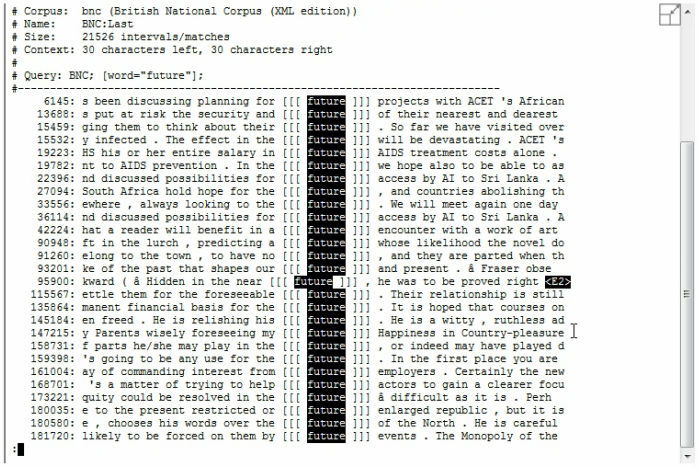


Chymiſche Hochzeit:

Darumb ſo laſt ewer groſſe klag/
Was iſts vmb etlich wenig tag.

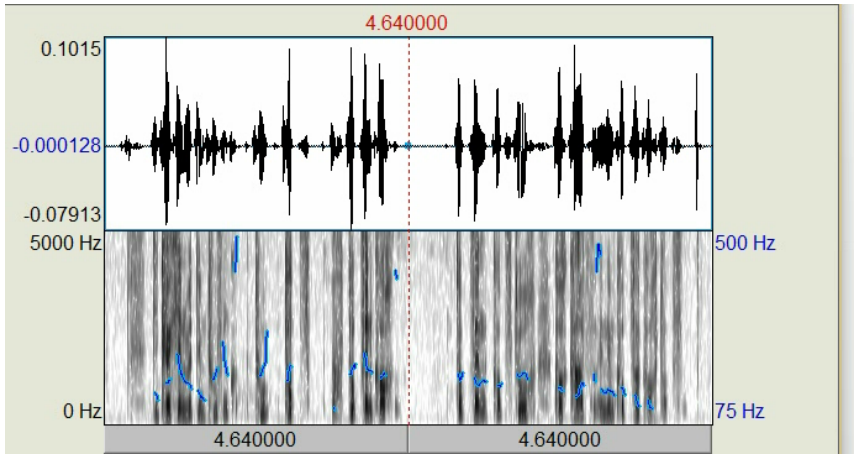
So bald er die wort vollendt/ ward der Deckel wider zu gethan/ vnd verſchloſſen/ vnd das Trom-meten vnd Heertrommeln wider angehoben/ So laut kont aber der Thon nit ſein/ man hört noch der gefangenen bittere klag/ die ſich im Thurn er-hoben für allen herauß. Welches mir dann auch bald die Augen vbergetrieben. Bald ſetzt ſich die alt Fraw mit jhrem Sohn auff zubereitte Seſſel nieder/ vnd beſilcht die erlöſte zu zehlen. Wie ſie nun die zahl vernommen/ vnd auff ein Goldgelb Täfflein auff geſchrieben/ begert ſie eines jeden Namen/ welche auch von eim Knäblein auffge-ſchrieben worden: Wie ſie vns nun nach einander anſihet/ erſeufftzet ſie/ vnd ſpricht zu jhrem Sohn/ das ichs wol hören kundt: Ach wie tawren mich die arme Menſchen im Thurn ſo vbel/ wolt Gott/ ich dörffte ſie alle erledigen. Darauff der Sohn geantwortet: Mutter/ ſo iſts von Gott verordnet/ dem ſollen wir nit widerſtreben/ wann wir alle Herſen weren/ vnd alles Gut hetten auff Erden/ vnd weren dann zu Diſch geſeſſen/ wer wolt vns doch bringen zu eſſen. Deßwegen die Mutter geſchwigen/ Aber bald darauff ſagt ſie: Nuhn ſo laſt doch dieſe von jhren Springern erledigen: Welches dann auch ſchnell geſchehen/ vnnd war ich ohn wenig der letſte. Noch kundte ich mich nit enthalten/ ob ich wol als auff andere geſehen/ ſonder neiget mich vor der alten Frawen/ vnnd dancket Gott/ der durch ſie mich auß ſolchem Fin-

Deutsches Textarchiv



```
# Corpus:  bnc (British National Corpus (XML edition))
# Name:    BNC:Last
# Size:    21526 intervals/matches
# Context: 30 characters left, 30 characters right
#
# Query: BNC; [word="future"];
#-------------------------------------------------------------------------------
     6145: s been discussing planning for [[[ future ]]] projects with ACET 's African
    13688: s put at risk the security and [[[ future ]]] of their nearest and dearest
    15459: ging them to think about their [[[ future ]]] . So far we have visited over
    15532: y infected . The effect in the [[[ future ]]] will be devastating . ACET 's
    19223: HS his or her entire salary in [[[ future ]]] AIDS treatment costs alone .
    19782: nt to AIDS prevention . In the [[[ future ]]] we hope also to be able to as
    22396: nd discussed possibilities for [[[ future ]]] access by AI to Sri Lanka . A
    27094: South Africa hold hope for the [[[ future ]]] , and countries abolishing th
    33556: ewhere , always looking to the [[[ future ]]] . We will meet again one day
    36114: nd discussed possibilities for [[[ future ]]] access by AI to Sri Lanka . A
    42224: hat a reader will benefit in a [[[ future ]]] encounter with a work of art
    90948: ft in the lurch , predicting a [[[ future ]]] whose likelihood the novel do
    91260: elong to the town , to have no [[[ future ]]] , and they are parted when th
    93201: ke of the past that shapes our [[[ future ]]] and present . â Fraser obse
    95900: kward ( â Hidden in the near [[[ future ]]] , he was to be proved right <p2>
   115567: ettle them for the foreseeable [[[ future ]]] . Their relationship is still
   135864: manent financial basis for the [[[ future ]]] . It is hoped that courses on
   145184: en freed . He is relishing his [[[ future ]]] . He is a witty , ruthless ad
   147215: y Parents wisely foreseeing my [[[ future ]]] Happiness in Country-pleasure
   158731: f parts he/she may play in the [[[ future ]]] , or indeed may have played d
   159398: 's going to be any use for the [[[ future ]]] . In the first place you are
   161004: ay of commanding interest from [[[ future ]]] employers . Certainly the new
   168701:  's a matter of trying to help [[[ future ]]] actors to gain a clearer focu
   173221: quity could be resolved in the [[[ future ]]] â difficult as it is . Perh
   180035: e to the present restricted or [[[ future ]]] enlarged republic , but it is
   180580: e , chooses his words over the [[[ future ]]] of the North . He is careful
   181720: likely to be forced on them by [[[ future ]]] events . The Monopoly of the
```

British National Corpus



Alcohol Language Corpus

TV News Archive

1. Korpora sind Zeitmaschinen

Frag. Warumb haben die Weiber laenger Haar/ dann die Maenner?

Antwort. Dieweil die Weiber mehr feuchtiger Natur sind/ dann die Maenner/ sind auch schnupffiger vnd fluessiger/ daher in jhnen mehr Saamens der Haar ist/ vnd folget endlich darauss die Laenge der Haar/ vnd darmit wird auch die Materi des Hirns ueberfluessiger von den jnnerlichen Gliedmassen/ sonderlich aber in der Weiber Vierwochenzeit/ dieweil alssdann das Wesen auffsteiget/ dardurch die Feuchtigkeit der Haar gemehret wird/ wie solches Albertus meldet. Sagt auch/ dass/ wo eines Weibs (die jhre Zeit hat) Haupthaar vnder den Mist gelegt werde/ soll darauss ein gifftige Schlange erwachsen.

Zum andern gibt man diese Antwort. Dieweil die Weiber keine Baert haben/ vnd dass also die Natur vnd Zeug des Barts zu dem Wesen der Haupthaar komme/ auch dieselbige vollstrecke.

1. Korpora sind Zeitmaschinen

Frag. Warumb haben die Weiber laenger Haar/ dann die Maenner?

Antwort. Dieweil die Weiber mehr **feuchtiger** Natur sind/ dann die Maenner/ sind auch **schnupffiger** vnd fluessiger/ daher in jhnen mehr **Saamens** der Haar ist/ vnd folget endlich darauss die Laenge der Haar/ vnd darmit wird auch die Materi des Hirns **ueberfluessiger** von den jnnerlichen Gliedmassen/ sonderlich aber in der Weiber Vierwochenzeit/ dieweil alssdann das Wesen auffsteiget/ dardurch die Feuchtigkeit der Haar gemehret wird/ wie solches Albertus meldet. **Sagt auch**/ dass/ wo eines Weibs (die jhre Zeit hat) Haupthaar vnder den Mist gelegt werde/ soll darauss **ein gifftige Schlange** erwachsen.

Zum andern gibt man diese Antwort. Dieweil die Weiber **keine Baert** haben/ vnd dass also die Natur vnd Zeug des Barts zu dem Wesen der Haupthaar komme/ auch dieselbige vollstrecke.

1. Korpora sind Zeitmaschinen

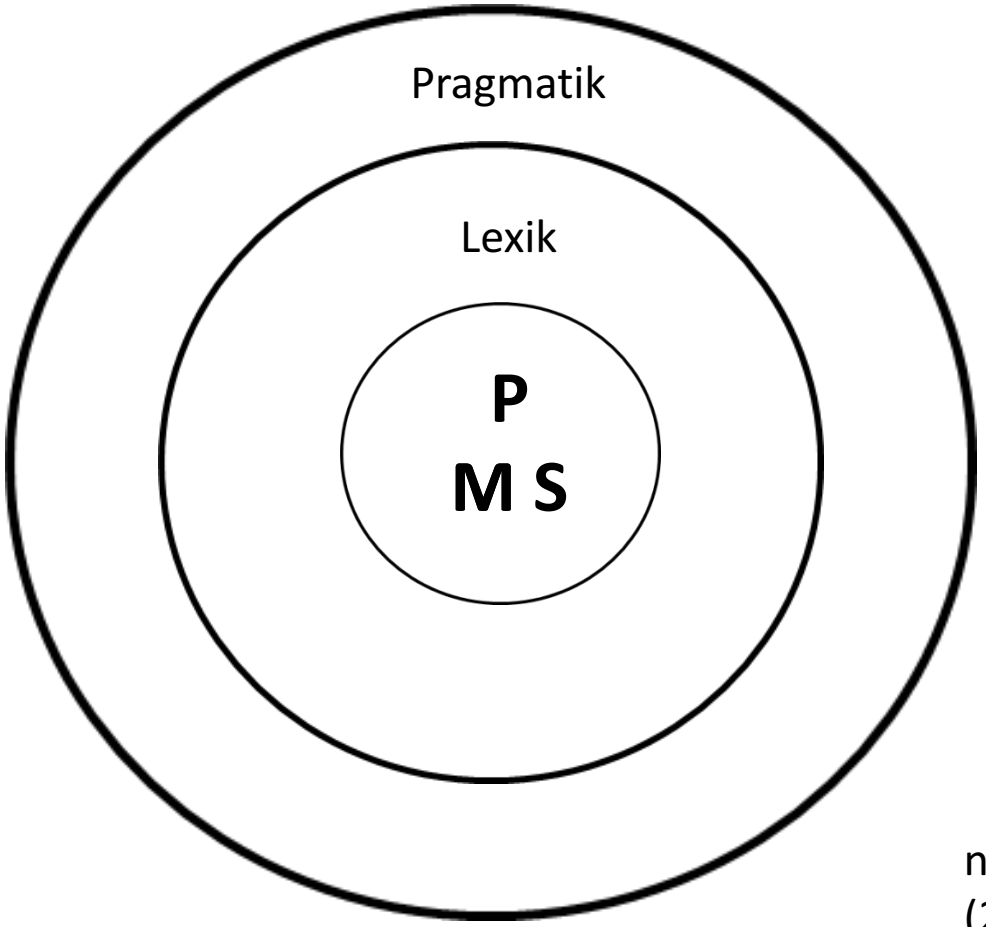
morphologischer Wandel:
feuchtiger Natur, keine Baert

graphematischer Wandel:
mehr Saamens; / (Virgel)

syntaktischer Wandel:
sind auch schnupffiger vnd fluessiger/
Sagt auch …

semantischer Wandel:
vnd darmit wird auch die Materi des Hirns
ueberfluessiger (→ ‚stärker überfließend')

Exkurs: Die „sprachliche Zwiebel“



nach Nübling et al. (2006: 2f.)

Korpora sind Zeitmaschinen

- Korpora ermöglichen Studien über Sprachgebrauch in der Vergangenheit
- Damit ermöglichen sie u.U. auch Voraussagen über mögliche zukünftige Entwicklungen (Beispiel: „flektierende“ Präpositionen)

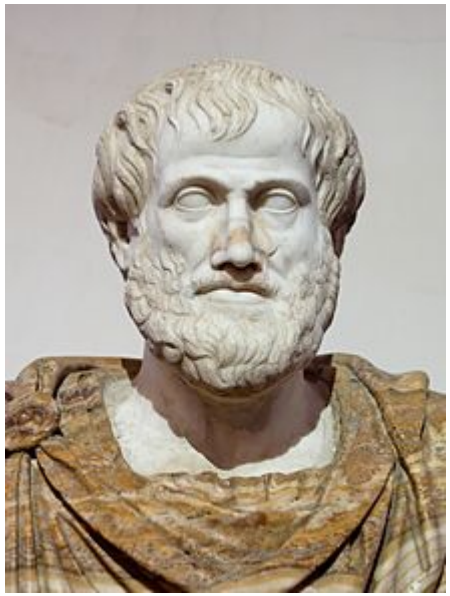
Die Zukunft beginnt heute…

- Korpuslinguistik als Disziplin „im Aufbruch“
- Verfügbarkeit von immer mehr Daten ermöglicht neue Fragestellungen
- Verfügbarkeit neuer methodischer Ansätze ermöglicht komplexere Fragestellungen
- … gerade in der Morphologie, wo die Operationalisierung von Konzepten wie "Produktivität" notorisch schwierig ist!

Wieso, weshalb, warum? Empirische Methoden und Wissenschaftstheorie

(nach Maxwell & Delaney 1999)

Aristoteles



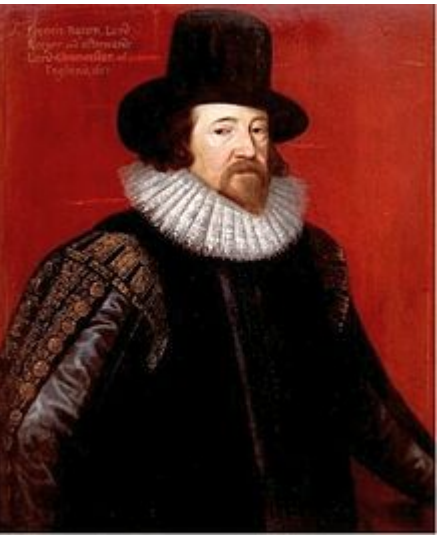
- **deduktive** (ableitende) Methode
- Ideal: Syllogismus

Alle Menschen sind sterblich.

Alle Griechen sind Menschen.

→ *Alle Griechen sind sterblich.*

Sir Francis Bacon



Novum Organon (1620):

- **Induktive** Methode
- aus Beobachtung von Naturphänomenen auf allgemeine Erkenntnis schließen
- Methode: Experimente!
- Ideal: Forscher ist objektiv und rational
- Explorativer Ansatz: Experimente nicht hypothesengeleitet

Vorannahmen

- Wissenschaft ist nie ganz frei von Vorannahmen
- Die wichtigsten davon:

  - Uniformität der Natur
  - Finite Kausalität

Uniformität der Natur

Uniformität der Natur

- Natur folgt gewissen Gesetzmäßigkeiten
- Daher sind Generalisierungen möglich.

Finite Kausalität

- Kausalkette, die zu einem Effekt führt, ist endlich.
- Damit ist der Effekt **replizierbar**.

Positivismus

- Vorläufer: David Hume: Inferenz einer kausalen Relation zwischen Unbeobachtbarem nie gerechtfertigt.
- Comte: Positivismus als ("ultimative") Religion

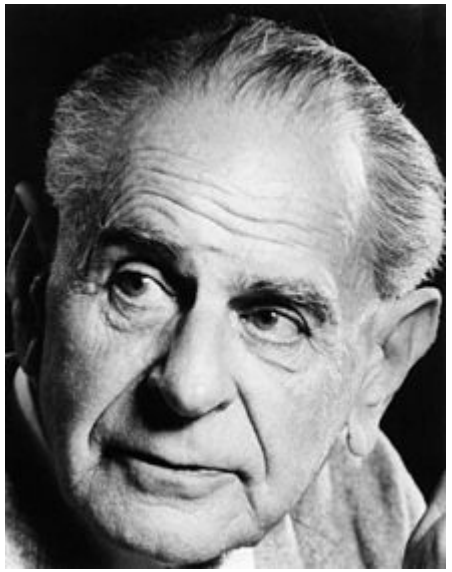
Auguste Comte

Logischer Positivismus

- Wiener Kreis (z.B. Rudolf Carnap, Herbert Feigl)
- Symbolische Logik als wichtigstes Analysewerkzeug
- Verifikationismus: Eine Proposition ist genau dann sinnvoll, wenn es eine empirische Methode gibt, um zu entscheiden, ob sie wahr oder falsch ist.
- Jedoch: Nicht alle wissenschaftlichen Fragestellungen lassen sich als universalgültige Propositionen formulieren.

Falsifikationismus

- Karl Popper: Wissenschaftlicher Fortschritt wird durch **Falsifikation** von Theorien erzielt.
- Rückkehr zur Deduktion, jedoch auf empirischer Grundlage.



Syllogismus der Bestätigung:
Wenn meine Theorie wahr ist, folgen die Daten dem von mir vorausgesagten Muster.
Die Daten folgen dem von mir vorausgesagten Muster.
Deshalb ist meine Theorie wahr.

Syllogismus der Falsifikation:
Wenn meine Theorie wahr ist, folgen die Daten dem von mir vorausgesagten Muster.
Die Daten folgen dem von mir vorausgesagten Muster *nicht*.
Deshalb ist meine Theorie falsch.

Occam's razor

- "*Entia non sunt multiplicanda praeter necessitatem*" (Johannes Clauberg, 17. Jh.)
- benannt nach Wilhelm von Ockham (13. Jh.)

Zusammenhang zur Korpuslinguistik

- Jede Korpusuntersuchung ist im Prinzip ein Experiment.
- Gerade in der **quantativen** Korpuslinguistik geht es v.a. ums Hypothesentesten.
- Dabei wird ein **falsifikationistischer** Ansatz gewählt:
  - Ich formuliere eine Hypothese…
  - …und überprüfe die **Nullhypothese.**

Aber…

- Nicht alle (korpus-)linguistischen Methoden sind falsifikationistisch.
- Verbreitet ist auch **exploratives** Arbeiten: Muster in den Daten erkennen.
- Auch die derzeit sehr gängigen **Bayesschen Ansätze** sind nicht, oder zumindest nicht immer, falsifikationalistisch orientiert.

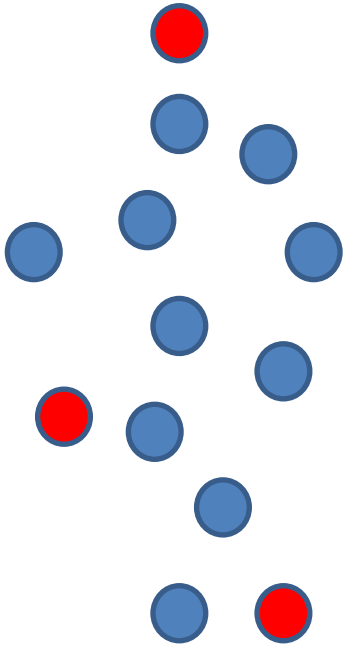
Deduktive vs. induktive Methode

Operationalisierung

Hypothese

Test

Neue / modifizierte / zusätzliche Hypothesen

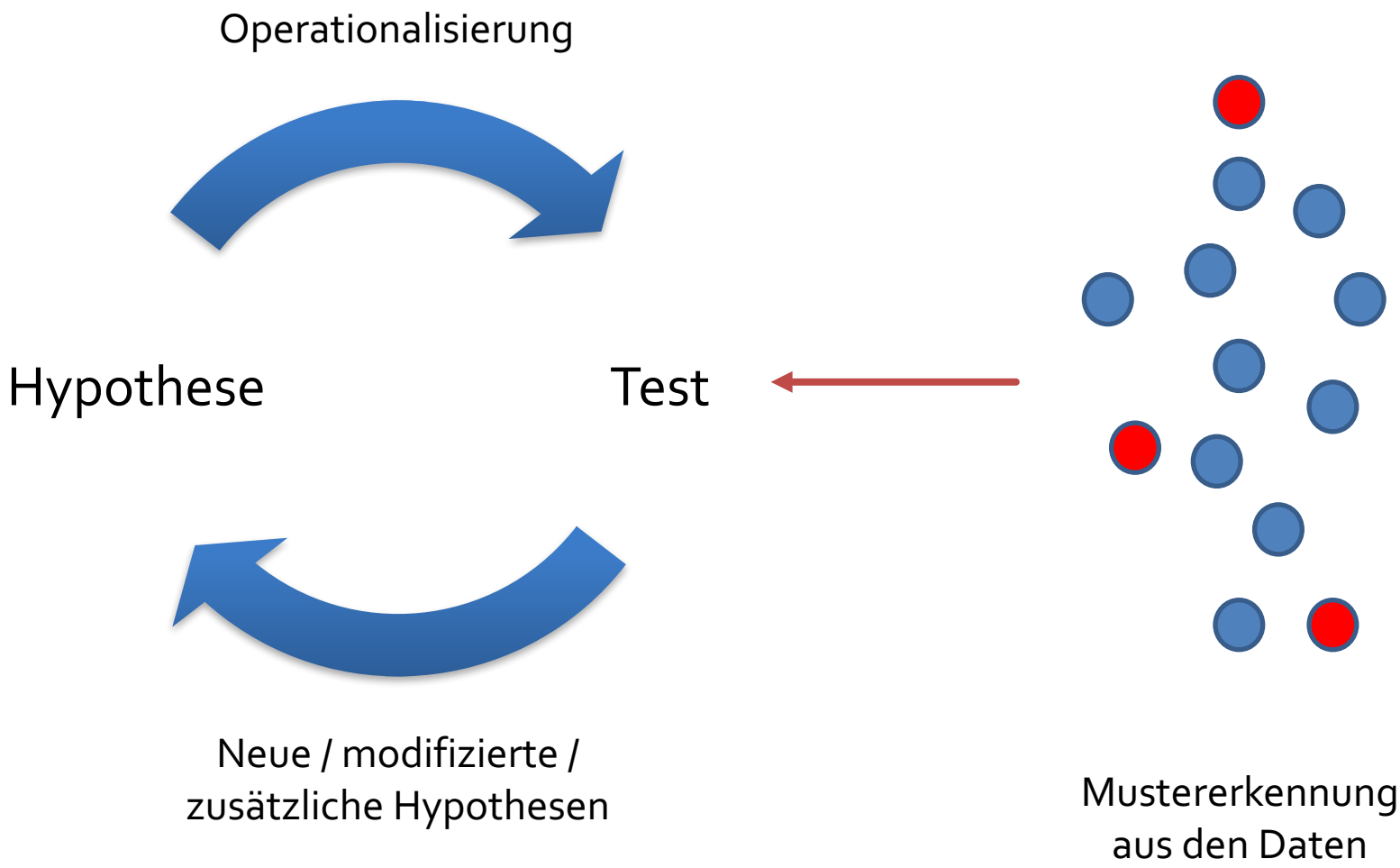


Mustererkennung aus den Daten

(nach Stefanowitsch 2017)


WOLVERINE?....
OR 2 BAT MEN?

Deduktive vs. induktive Methode



(nach Stefanowitsch 2017)

Warum eigentlich Korpuslinguistik?

"Corpus linguistics doesn't mean anything."
(Chomsky 2004)

Wozu?

Ich kann doch Deutsch (Englisch, Französisch, Mandarin……) - warum brauche ich dann ein Korpus?

beizeiten, mitunter

Aus dem Korpus der Wikipedia-Diskussionen (verfügbar über COSMAS II):

- Ich werde **beizeiten** nach Quellen suchen
- Ich werde **beizeiten** die Gliederung noch ein wenig umstellen und mir das ganze nochmal mit etwas Abstand durchlesen,
- Dieser Artikel ist grausamst falsch. ich sollte mich **beizeiten** als Tropenmedizinerin mal selbst dransetzen…
- Vielleicht hat ja jemand ein vollständigeres Bild, das man **beizeiten** hier einfügen kann.

beizeiten, mitunter

Aus dem Korpus der Wikipedia-Diskussionen (verfügbar über COSMAS II):

- Das dritte Zitat ist ja **mitunter** ein Grund für die Namensgebung

- Ich hab dazu nirgends was gefunden. Es sollte **mitunter** auch im Artikel erwähnt werden!

gleichwohl

Aus dem Korpus der Wikipedia-Diskussionen (verfügbar über COSMAS II):

- . ..gleichwohl noch nicht alle bereiche komplett dereguliert sind
- … und andere Themen, gleichwohl sie sich im Kontext des ersten Themas befinden mögen, lieber unter den Tisch fallen lassen.

gleichsam

Internetbelege (z.T. eigene Funde, z.T. DECOW14AX)):

- [Dieses Vorgehen] ist **gleichsam** künstlerisch integer, wie konzernwirtschaftlich gerissen (http://bit.ly/1LYhWka)

- um ähnlich wie in Fassbinders CHINESISCHES ROULETTE (1976) die **gleichsam** dekadente wie misanthropische Upperclass abzubilden (http://bit.ly/1a1Sw86)

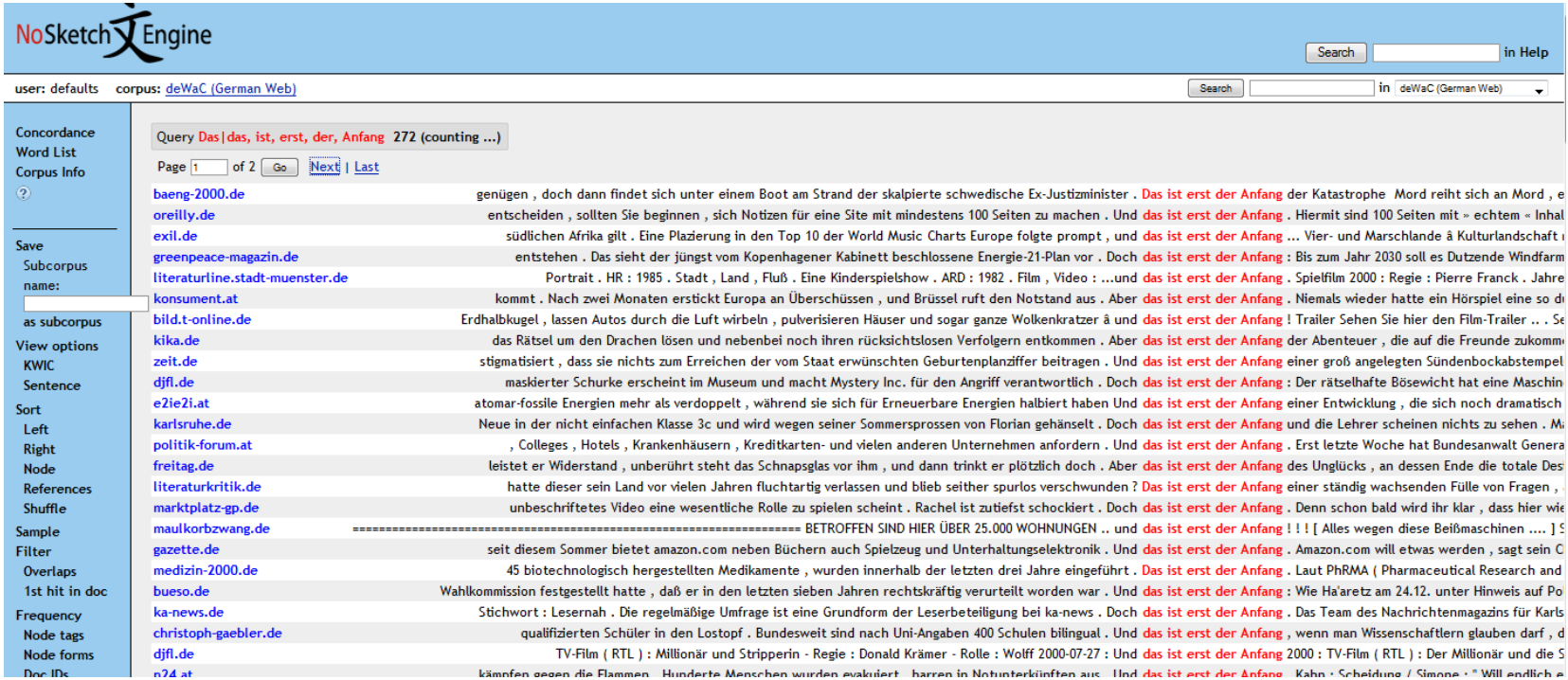
- Demnach ließe sich also leicht die Feststellung treffen, ein kongenialeres und **gleichsam** spannungsreicheres Duo als Joshua Redman und Brad Mehldau ließe sich nur schwerlich im 21. Jahrhundert auf einer Jazzbühne vereinigen. (http://bit.ly/1PQ8m8U)

Können wir unseren Intuitionen trauen?

- Intuiton als notwendiger erster Schritt…
- …aber Intuition ist erst der Anfang!
- …oder *nur* der Anfang?

erst der Anfang vs. nur der Anfang

- deWaC: *erst* 1374, *nur* 1144; ganzer Satz: 278 *erst* vs. 138 *nur*



Anwendungsbereiche…

- Zweifelsfälle
- …was noch?

Anwendungsbereiche…

- Zweifelsfälle
- Historische Wandelprozesse
- Varietätenlinguistik und Dialektologie
- graphematischer Wandel
- Multimodalität und Interaktionsstudien
- Phonetik
- ……

Was ist Korpuslinguistik?

Corpus linguistics is the investigation of linguistic research questions that have been framed in terms of the conditional distribution of linguistic phenomena in a linguistic corpus.
(Stefanowitsch 2017)

Was ist Korpuslinguistik?

Corpus linguistics is the investigation of linguistic research questions that have been framed in terms of the **conditional** distribution of linguistic phenomena in a linguistic corpus.
(Stefanowitsch 2017)

- "Der Genitiv taucht in älteren Texten **häufiger** auf als in neueren Texten."
- "Ältere Sprecherinnen benutzen **seltener** Fremdwörter als jüngere."
- "Frauen benutzen **mehr** Diskurspartikeln als Männer."
- "Der Ausdruck *parkieren* wird **nur** im Schweizerdeutschen gebraucht."

- "Anglizismen werden im Deutschen **häufig** gebraucht."

Was ist Korpuslinguistik?

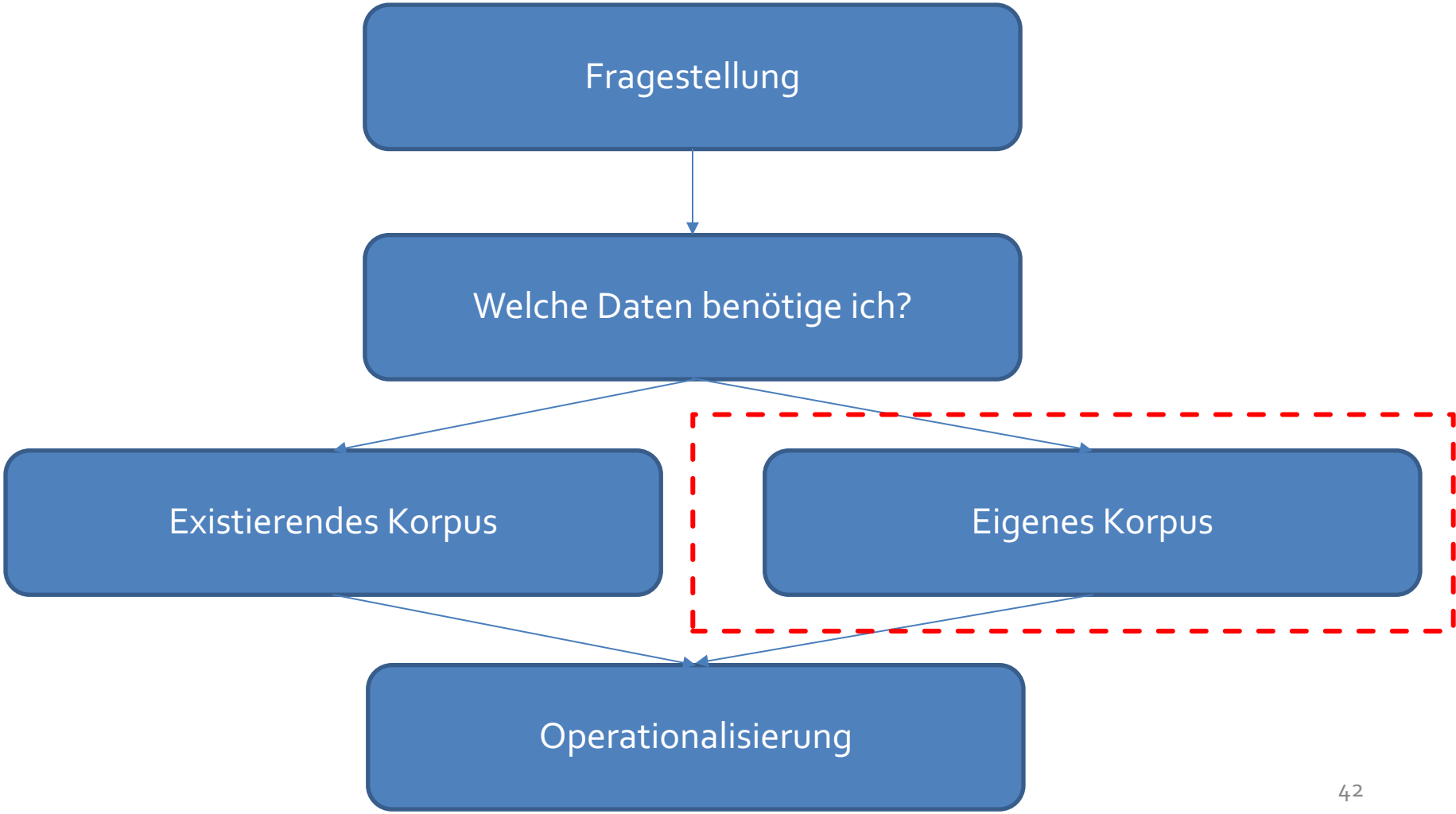
Korpusbasierte vs. korpus-illustrierte Ansätze

- "korpus-illustrierte" Ansätze sind qualitativ, benutzen aber selektiv ausgewählte Korpusbelege (z.B. viele Arbeiten von Bybee, Traugott, Trousdale)
- korpusbasierte Ansätze können rein **quantitativ** sein oder aber "**quantitativ-qualitativ"** (Lemnitzer & Zinsmeister 2015)

Was ist Korpuslinguistik

- Rein quantitative Ansätze stützen sich **ausschließlich** auf die Korpusdaten (z.B. n-Gramme, Latent-semantische Analyse…)
- Quantitativ-qualitative Ansätze stützen sich auf die Analyse und Interpretation der Daten (**Annotation**)

Arbeitsschritte in der Korpuslinguistik



Korpusdesign

Ein Wissenschaftler vom Mars bittet Sie darum, ein Korpus zusammenzustellen, das möglichst genau abbildet, wie die Leute in Hamburg sprechen.

Wie gehen Sie vor?

Korpusdesign

- Repräsentativität
- Ausgewogenheit
- Größe
- Angemessenheit für die jeweilige Forschungsfrage

bei transliterierten Texten:

- Qualität der Transliteration

Korpusdesign

Grundsätzliche Fragen:

- Was genau möchte ich untersuchen?
- Welche Art von Daten brauche ich dafür?
- Gibt es ein solches Korpus schon?
- In welcher Hinsicht muss das Korpus besonders akkurat sein?
  - z.B. bei graphematischen Untersuchungen: Graphie des Originals genau abbilden etc.

Korpusdesign

Falls ich ein eigenes Korpus zusammenstelle:

- Wie komme ich an Daten?
- Gibt es urheberrechtliche Bedenken?
- Gibt es sonstige moralische / ethische Bedenken?

Korpuserstellung

- Datensammlung und -aufbereitung
- Tokenisierung
- Lemmatisierung und POS-Tagging (z.B. TreeTagger)
- ggf. weitere Annotation

Exkurs: Wir basteln uns ein Korpus

NEWSTICKER +++ Neue Satzzeichen sollen Inflation von Frage- und Ausrufezeichen eindämmen!!! +++

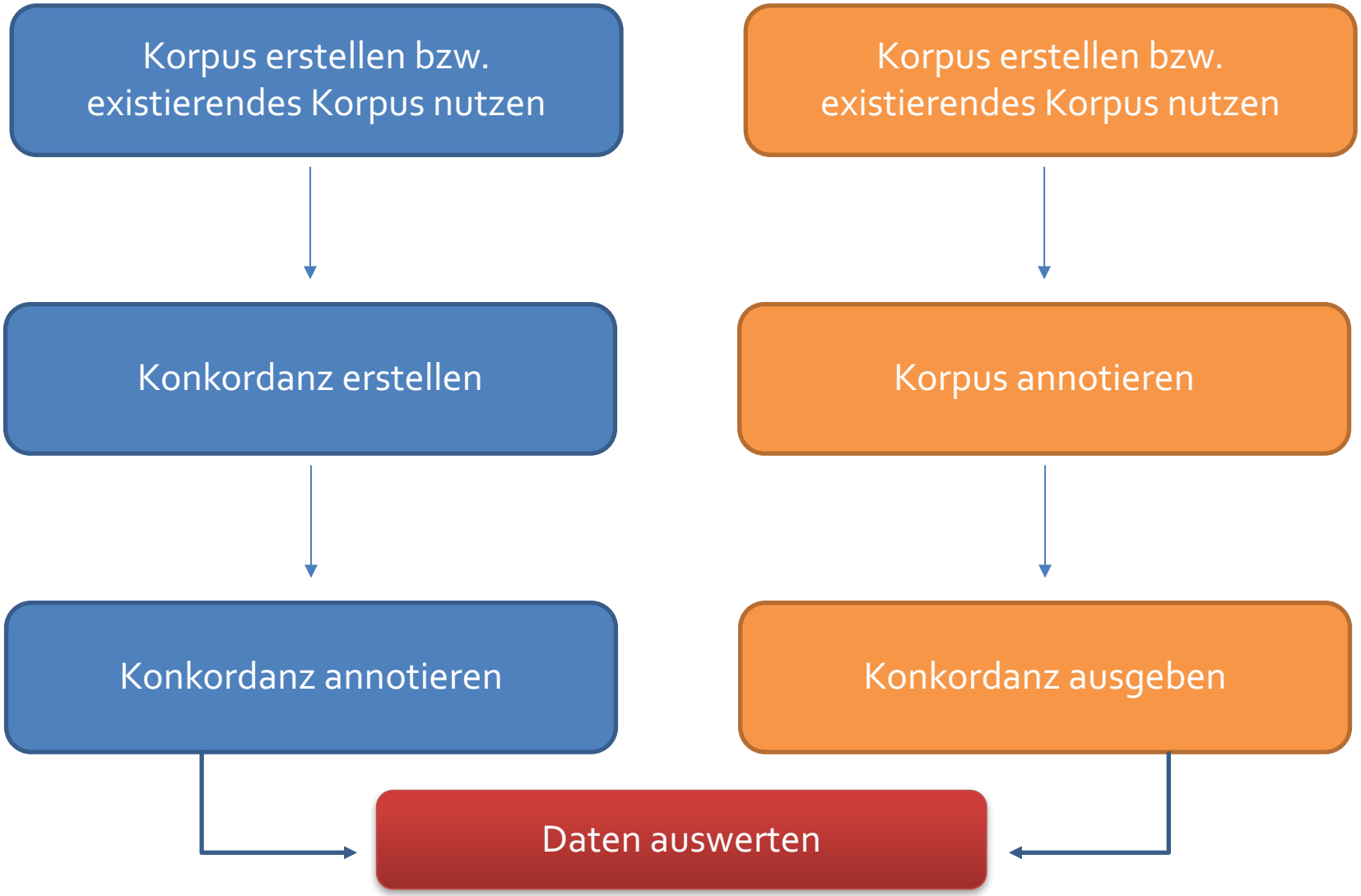
Sonntagsfrage: Wen würden Sie zum Hundespräsidenten wählen?

Ja, wer ist ein guter Hund? Ja, wer ist ein braver Hund? Wer ist der richtige Hund für Deutschland? Ja, wer? Ja, wer? Genau das will der *Gassillon* in dieser Woche (unterstützt vom Meinungsforschungsinstitut *Bone*

Wir basteln uns ein Postillon-Korpus

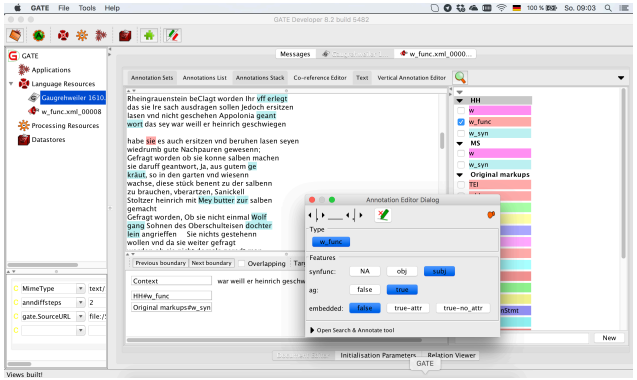
- Vorteil: Postillon-Texte unter Creative-Commons-Lizenz
- Text lässt sich recht einfach extrahieren
- Man muss nur den Seitenquelltext analysieren…

Mögliche Workflows



Annotation: Zwei Wege

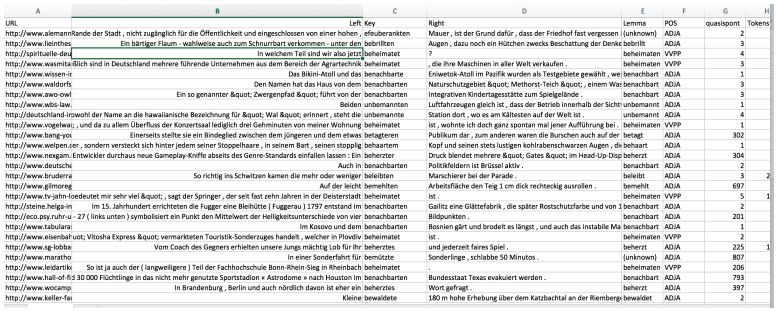
Annotation
direkt im Korpus



Export

Export

Annotation
außerhalb des Korpus



Korpus- oder Konkordanzannotation?

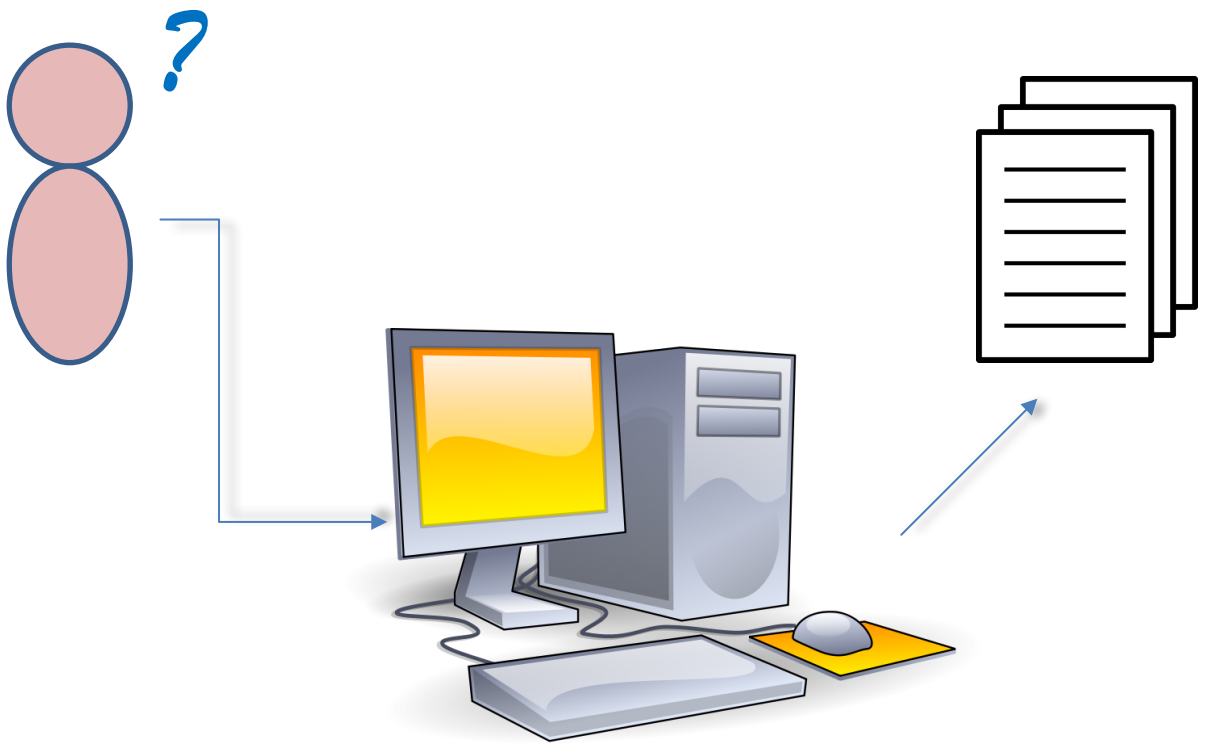
- **Korpusannotation**: Die Annotationen werden für das gesamte Korpus direkt in den Korpusdateien vorgenommen.
- **Konkordanzannotation**: Die Annotation erfolgt nach Ausgabe der Konkordanz in einem Tabellenkalkulationsprogramm.

Korpusauswertung

- In diesem Seminar beschränken wir uns auf die Arbeit mit **existierenden** Korpora.

- Daher werden wir uns auf den zweiten Annotationsweg beschränken (Annotation in Spreadsheet-Programmen).

Vorgehen



Fragestellung Suchabfrage Konkordanz

Korpuslinguistische Grundbegriffe

POS-Tagging & Lemmatisierung

Dieweil	ADV	dieweil
die	ART	die
Weiber	NN	Weib
mehr	ADV	mehr
feuchtiger	ADJA	feuchtiger
Natur	NN	Natur
sind/	VVFIN	sind/
dann	ADV	dann
die	ART	die
Maenner/	ADJA	Maenner/
sind	VAFIN	sein
auch	ADV	auch
schnupffiger	ADJA	schnupffiger
vnd	NN	vnd
fluessiger/	VVFIN	fluessiger/
daher	PAV	daher
in	APPR	in
jhnen	ADJA	jhnen
mehr	PIAT	mehr
Saamens	NN	Saamens
der	ART	die
Haar	NN	Haar
ist/	ADJA	ist/

- oft automatisch, z.B. mit TreeTagger
- Vorteil: extrem schnell und effizient
- Nachteil: ungenau
- für historische Daten z.T. eigene Tagger verfügbar
- z.B. eigenes TreeTagger Parameter File für Mhd.

Tagsets

- unterschiedliche Tagsets für POS
- am verbreitetsten jedoch: Stuttgart-Tübingen Tagset (STTS)
- Übersicht: http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html

Types und Tokens

Wort	Freq
die	9
der	5
vnd	5
Weiber	4
auch	3
Antwort.	2
dann	2
darauss	2
den	2
des	2
Dieweil	2
Haar	2
Haar/	2
Haupthaar	2
in	2

Types vs. Tokens






Types vs. Tokens






Types vs. Tokens



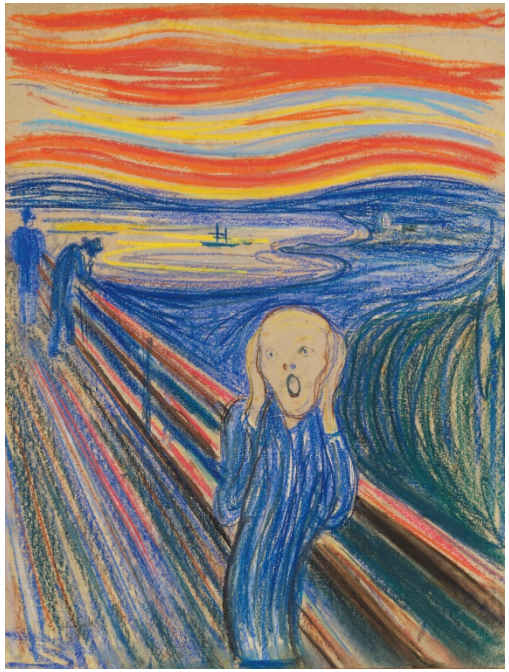


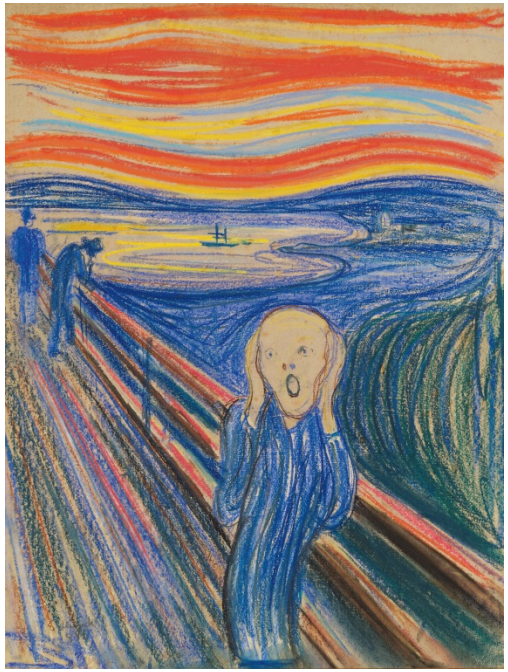




Types vs. Tokens

Types vs. Tokens





Types vs. Tokens






Wie viele Types…?

Es kommt drauf an…

Types und Tokens

Wenn Fliegen neben Fliegen fliegen, fliegen Fliegen neben Fliegen.

Lemma	**Tokens**
Fliege	4
fliegen	2
wenn	1
neben	2

Methoden der Korpusanalyse

Korpusauswertung

qualitative Analyse:

- Beobachtungen auf Grundlage einzelner Belege
- kann sich auf alle Aspekte von der Semantik über die Morphologie bis hin zur Syntax beziehen
- gerade für semantische und pragmatische Analysen geeignet

Korpusauswertung

quantitative Analyse:

- Einbezug zahlreicher Belege statt Einzelbeobachtungen
- Quantifizierung z.B. durch
  - Zählen von Wörtern, Wortarten, grammatischen Mustern usw.
  - statistische Methoden (z.B. Kollokationsmaße)

Korpusauswertung

Frag. Warumb haben die Weiber laenger Haar/ dann die Maenner?
Antwort. Dieweil die Weiber mehr feuchtiger Natur sind/ dann die Maenner/ sind auch schnupffiger vnd fluessiger/ daher in jhnen mehr Saamens der Haar ist/ vnd folget endlich darauss die Laenge der Haar/ vnd darmit wird auch die Materi des Hirns ueberfluessiger von den jnnerlichen Gliedmassen/ sonderlich aber in der Weiber Vierwochenzeit/ dieweil alssdann das Wesen auffsteiget/ dardurch die Feuchtigkeit der Haar gemehret wird/ wie solches Albertus meldet. Sagt auch/ dass/ wo eines Weibs (die jhre Zeit hat) Haupthaar vnder den Mist gelegt werde/ soll darauss ein gifftige Schlange erwachsen.
Zum andern gibt man diese Antwort. Dieweil die Weiber keine Baert haben/ vnd dass also die Natur vnd Zeug des Barts zu dem Wesen der Haupthaar komme/ auch dieselbige vollstrecke.

Wie können wir diesen Text untersuchen?

Qualitativ vs. quantitativ

Bitte überlegen Sie: Welche Vor- und Nachteile haben **qualitative** bzw. **quantitative** Ansätze?

Wofür würden Sie welchen Ansatz wählen?

1. Wandel der Genitivstellung (*des Vaters Haus* > *das Haus des Vaters*)
2. Rassismus in Leserbriefen
3. Semantischer Wandel von *geil*

Qualitativ *und* quantitativ vs. quantitativ

- Die meisten korpuslinguistischen Ansätze sind zugleich qualitativ und quantitativ
- Operationalisierung einzelner (z.B. semantischer) Variablen erfordert in der Regel eine (qualitative) **Interpretation** der einzelnen Belege
- Beispiel: Belebtheitsannotation

Qualitativ *und* quantitativ

- Sogar syntaktische Annotation erfordert oft Interpretation der Daten

Gefragt worden,

Ob sie nicht einmal Wolfgang Sohnen des Oberschulteisen dochterlein angrieffen

Sie nichts gestehenn wollen

vnd da sie weiter gefragt worden

ob sie nicht domals geredt man könne dem kindt nicht wieder noch wol helffenes

sey den weil der 9 te noch nit furuber

diesesauchnicht gestehen wollen

(SiGS-Korpus, Gaugrehweiler 1610)

Von der Konkordanz zur Analyse

- Operationalisierung von Hypothesen
- --> klare und nachvollziehbare Annotationskriterien!

Produktivität

Nr.	Ansatz	Definition
I		Produktivität einer Wortbildungsregel ist eine Funktion der Menge der zu einem bestimmten Zeitpunkt nach dieser Wortbildungsregel **realisierten** Bildungen (direkte Proportionalität)
II		Produktivität einer Wortbildungsregel ist die Funktion der Menge ihrer **möglichen** Basen (direkte Proportionalität)
III		Produktivität als Verhältnis von **usuellen** zu **möglichen** Bildungen
IV		Produktivität als **Möglichkeit** von Neubildungen
V		Produktivität als **Wahrscheinlichkeit** von Neubildungen
VI		Produktivität als **Anzahl** der Neubildungen in einem bestimmten Zeitraum

(Rainer 1987)

Produktivität

- Ist die Definition **synchron**, **diachron** oder **unbestimmt**?
- Ist die Definition **quantitativ** oder **qualitativ** orientiert?
- Welche Definition kommt dem **Scherer (2006)** zugrundeliegenden Produktivitätsbegriff am nächsten?

Produktivität

Nr.	Ansatz	Definition
I		Produktivität einer Wortbildungsregel ist eine Funktion der Menge der zu einem bestimmten Zeitpunkt nach dieser Wortbildungsregel **realisierten** Bildungen (direkte Proportionalität)
II		Produktivität einer Wortbildungsregel ist die Funktion der Menge ihrer **möglichen** Basen (direkte Proportionalität)
III		Produktivität als Verhältnis von **usuellen** zu **möglichen** Bildungen
IV		Produktivität als **Möglichkeit** von Neubildungen
V		Produktivität als **Wahrscheinlichkeit** von Neubildungen
VI		Produktivität als **Anzahl** der Neubildungen in einem bestimmten Zeitraum

(Rainer 1987)

Produktivität

Nr.	Ansatz	Definition
I	synchron	Produktivität einer Wortbildungsregel ist eine Funktion der Menge der zu einem bestimmten Zeitpunkt nach dieser Wortbildungsregel **realisierten** Bildungen (direkte Proportionalität)
II	synchron	Produktivität einer Wortbildungsregel ist die Funktion der Menge ihrer **möglichen** Basen (direkte Proportionalität)
III	synchron	Produktivität als Verhältnis von **usuellen** zu **möglichen** Bildungen
IV	unbestimmt	Produktivität als **Möglichkeit** von Neubildungen
V	unbestimmt	Produktivität als **Wahrscheinlichkeit** von Neubildungen
VI	diachron	Produktivität als **Anzahl** der Neubildungen in einem bestimmten Zeitraum

(Rainer 1987)

Usuelle vs. mögliche Wörter

- „Mögliche" Wörter: z.B. *Durstigkeit, Fleißigkeit*
- Usuelles Wort: „Wort, das zu einem bestimmten Zeitpunkt zum Wortschatz eines bestimmten Sprechers gehört" (Rainer 1987)
- (Wie) Lassen sich diese Begriffe **empirisch** operationalisieren?

Produktivität

Nr.	Ansatz	Definition
I	synchron	Produktivität einer Wortbildungsregel ist eine Funktion der Menge der zu einem bestimmten Zeitpunkt nach dieser Wortbildungsregel **realisierten** Bildungen (direkte Proportionalität)
II	synchron	Produktivität einer Wortbildungsregel ist die Funktion der Menge ihrer **möglichen** Basen (direkte Proportionalität)
III	synchron	Produktivität als Verhältnis von **usuellen** zu **möglichen** Bildungen
IV	unbestimmt	Produktivität als **Möglichkeit** von Neubildungen
V	unbestimmt	Produktivität als **Wahrscheinlichkeit** von Neubildungen
VI	diachron	Produktivität als **Anzahl** der Neubildungen in einem bestimmten Zeitraum

(Rainer 1987)

Produktivität

- synchron: Fähigkeit eines Musters, Neubildungen hervorzubringen
- diachron: Fähigkeits eines Musters zum Zeitpunkt t_i, Neubildungen hervorzubringen, im Vergleich zum Zeitpunkt t_{i-1}, t_{i-2}, …, t_{i-n}.

Produktivität

- binär: ein Wortbildungsmuster ist entweder produktiv oder unproduktiv
- graduell: ein Wortbildungsmuster kann in verschiedenem Maße produktiv sein

Produktivität

- „Doppelexistenz“: Jedes Wortbildungsprodukt ist **zugleich** ein eigenes Wort **und** Instantiation eines Wortbildungsmusters
- Wortbildungsprodukte sind in unterschiedlichem Maße **transparent**
- vgl. *Jungtier* ‚junges Tier‘ vs. *Junggeselle* *‚junger Geselle‘; *Landung* vs. *Zeitung*

Produktivität

- Voraussetzung dafür, dass ein Wortbildungsmuster produktiv ist, ist, dass es **erkennbar** ist
- z.T. jedoch phonologischer Schwund (z.B. *vrouw-**ida*** > Freu***de***)
- z.T. noch erkennbar, aber sehr infrequent und (wohl deshalb) unproduktiv: *Kehr**icht***, *Dick**icht***.

„Grammatikalisierung“?

- Grammatikalisierung: Lexikalisch(er)e Einheiten werden zu grammatisch(er)en Einheiten.
- Grammatikalisierung i.e.S.: Entstehung von Einheiten, die grammatische **Kategorien** kodieren (Kriterium der Obligatorizität!)
- Aber: Auch WB-Affixe können „im weiteren Sinne als grammatische Morpheme gelten“ (Munske 2002: 28)

Was ist eigentlich Grammatikalisierung?

Elizabeth Closs Traugott

Paul J. Hopper

- the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical function

Hopper & Thompson (2003): Grammaticalization. 2nd ed. Cambridge.

Was ist eigentlich Grammatikalisierung?

Ich **habe** ein Smartphone.

Ich **habe** Blumenkohl gegessen.

Was ist eigentlich Grammatikalisierung?

Elizabeth Closs Traugott

Paul J. Hopper

- the change whereby lexical items and constructions **come** in certain linguistic contexts **to serve grammatical functions** and, once grammaticalized, continue to develop new grammatical function

Hopper & Thompson (2003): Grammaticalization. 2nd ed. Cambridge.

Was ist eigentlich Grammatikalisierung?

Elizabeth Closs Traugott

Paul J. Hopper

- the change whereby lexical items and constructions **come** in certain linguistic contexts **to serve grammatical functions** and, once grammaticalized, continue to develop new grammatical function

Hopper & Thompson (2003): Grammaticalization. 2nd ed. Cambridge.

Was ist eigentlich Grammatikalisierung?

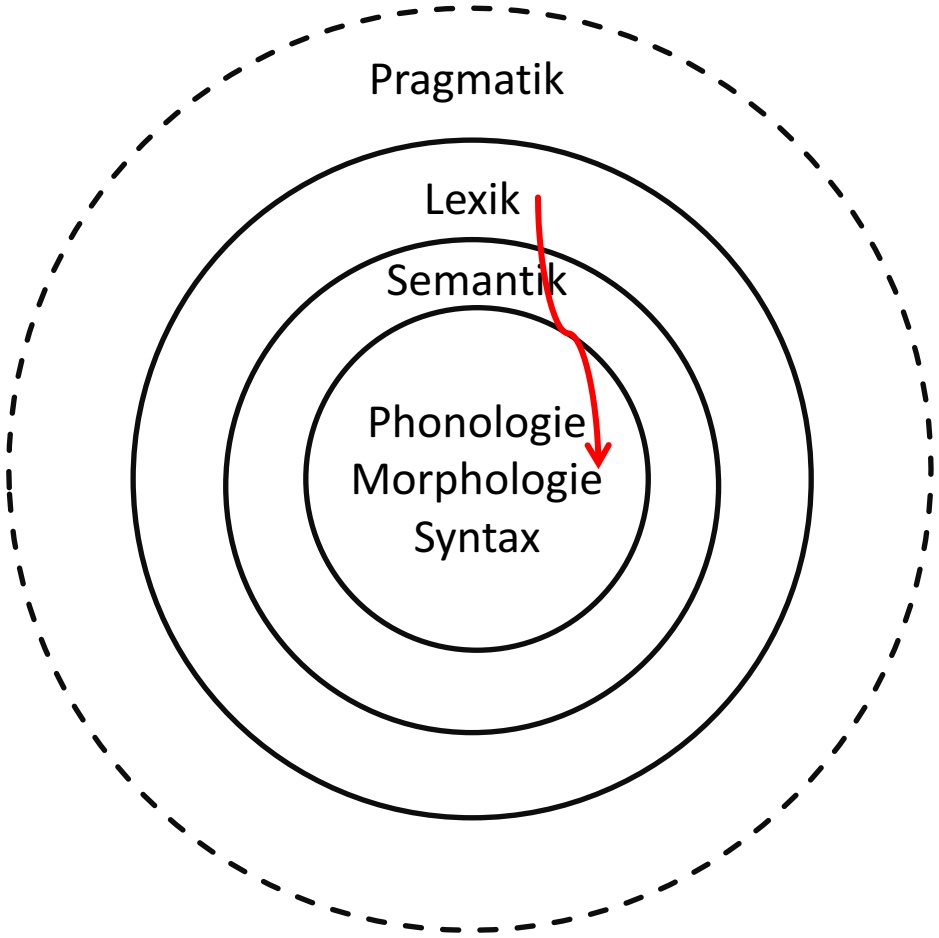
Elizabeth Closs Traugott

Paul J. Hopper

- the change whereby lexical items and constructions **come** in certain linguistic contexts **to serve grammatical functions** and, once grammaticalized, continue to develop new grammatical function

Hopper & Thompson (2003): Grammaticalization. 2nd ed. Cambridge.

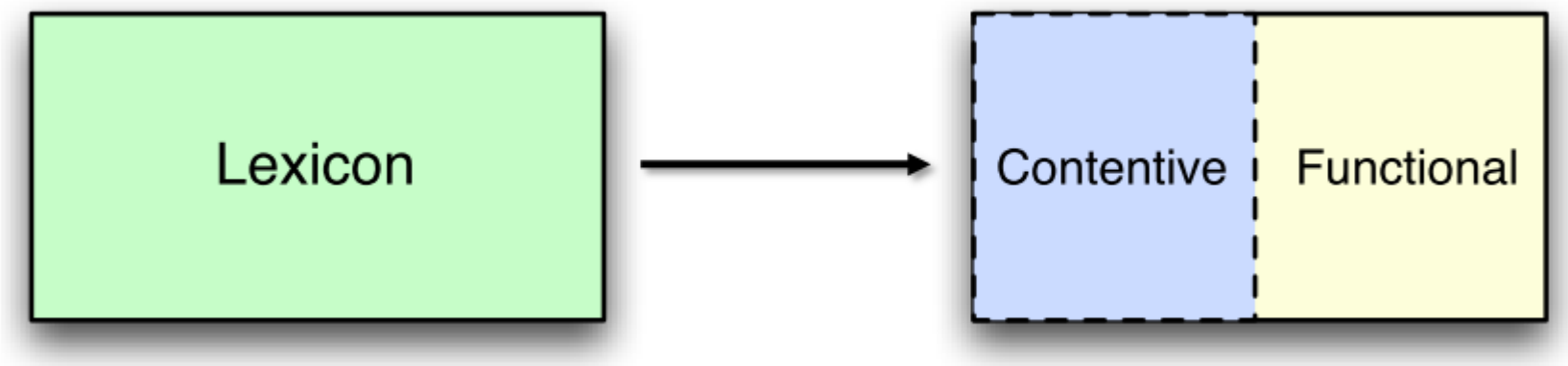
Was ist eigentlich Grammatikalisierung?



nach Nübling et al. [3]2010:2; dort nach Debus [2]1980:188

Was ist eigentlich Grammatikalisierung?

- Inhalts- vs. Funktionswörter / -einheiten
- z.B.: Hilfsverb *werden* (*Wir werden heute über Grammatikalisierung sprechen*) vs. Vollverb werden (*Ich werde Lehrerin*)



Quelle: http://replicatedtypo.com/evolving-linguistic-replicators-major-transitions-and-grammaticalisation/3097.html

Was ist eigentlich Grammatikalisierung?

- Vor vielen Jahren kaufte Petra ein Fahrrad. Jetzt hat sie ihr Fahrrad wieder verkauft.

Was ist eigentlich Grammatikalisierung?

- Vor vielen Jahren kauf**te** Petra ein Fahrrad. Jetzt **hat** sie ihr Fahrrad wieder verkauf**t**.

Was ist eigentlich Grammatikalisierung?

Desemantisierung

Verlust an semantischem Gehalt

Extension

Verwendung in neuen Kontexten (z.B. morphologischer Umlaut)

Dekategorisierung

Verlust des Status als unabhängiges Wort (Verlust der morphosyntaktischen Eigenschaften des Ursprungswortes)

Erosion

Reduktion phonologischer Substanz

Grammatikalisierungszyklus

- Hypothese: unidirektionaler / zyklischer Wandel
- Diskurs > Syntax > Morphologie > Morphonologie > Schwund (nach Givón 1979 und Lehmann 1995)

Konstruktionalisierung

Elizabeth Closs Traugott

Graeme Trousdale

Konstruktionalisierung

Konstruktionsgrammatik:

- Sprache als Netzwerk von **Konstruktionen** (Form-Bedeutungs-Paaren)

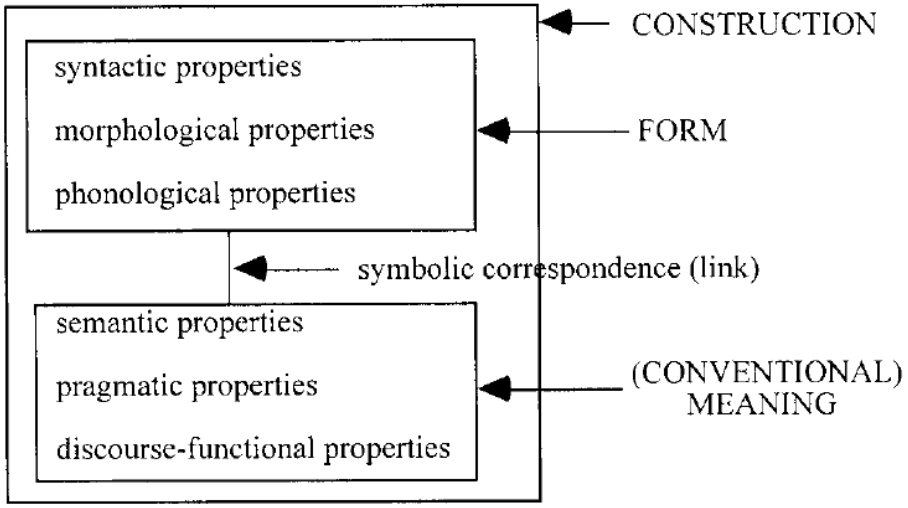


Figure 18.2. The symbolic structure of a construction

(aus Croft 2007)

Konstruktionalisierung

- Lexikon-Syntax-Kontinuum (aus Croft 2001):

Construction type	traditional name	example
complex and (mostly) schematic	syntax	Passivkonstruktion
complex and (mostly) specific	idiom	jn. auf den Arm nehmen, je X-er desto Y-er
complex but bound	morphology	Land-ung; Fisch-e
atomic and schematic	syntactic category	[DEM], [ADJ]
atomic and specifcic	word / lexicon	*Hund, Katze*

Konstruktionalisierung

Traugott & Trousdale (2013)

- Konstruktionalisierung: Entstehung eines neuen Knotens im Konstruktikon
- Beispiel: *Was macht* XY? - Konstruktion (vgl. auch Bybee 2010)

- Was machst du da? fuhr Kronbecher sie an. (Jentzsch, Kerstin: Ankunft der Pandora | DWDS)
- Was macht der Blüm jetzt schon wieder? (Der Spiegel, 05.10.1987 | DWDS)

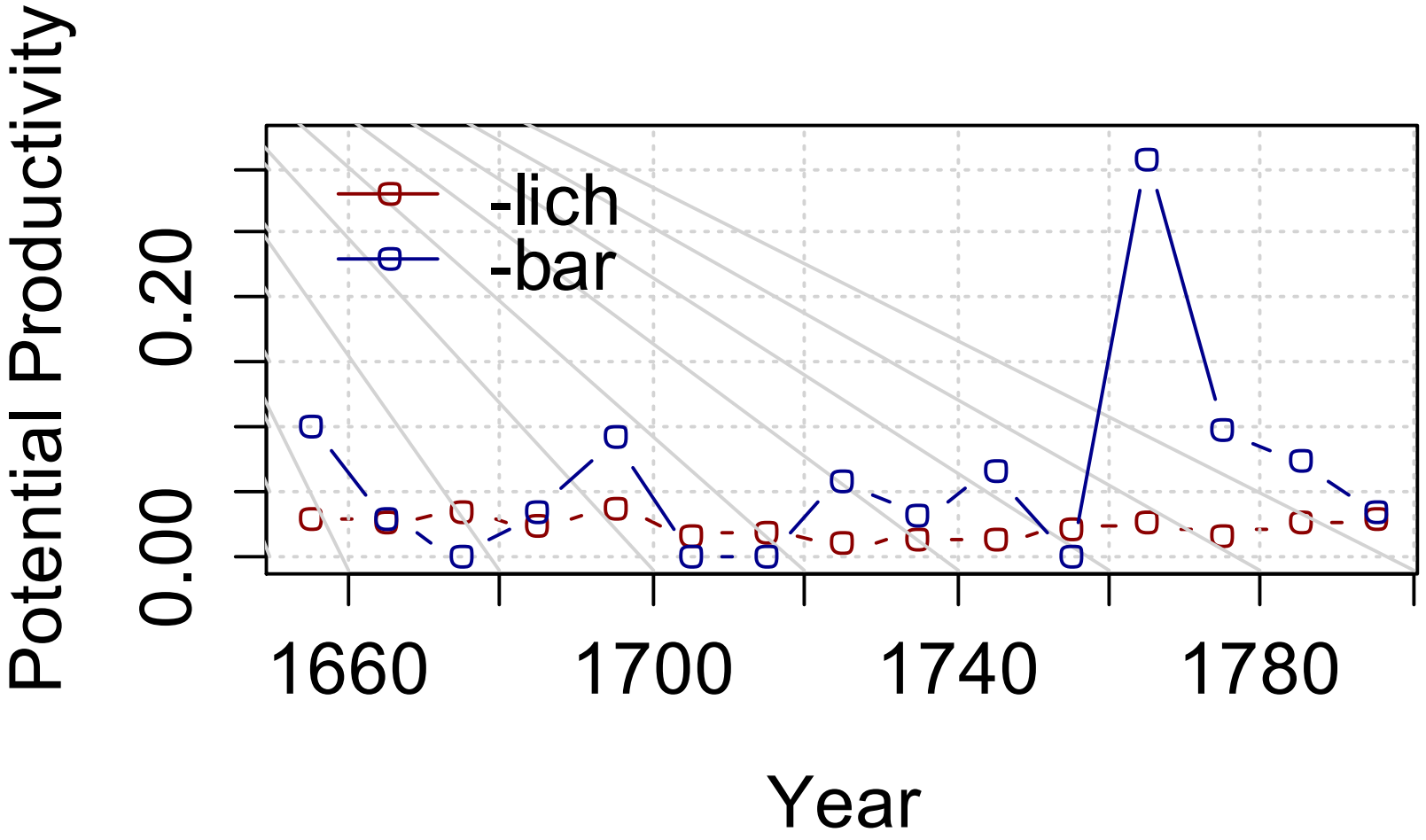
Konstruktionalisierung

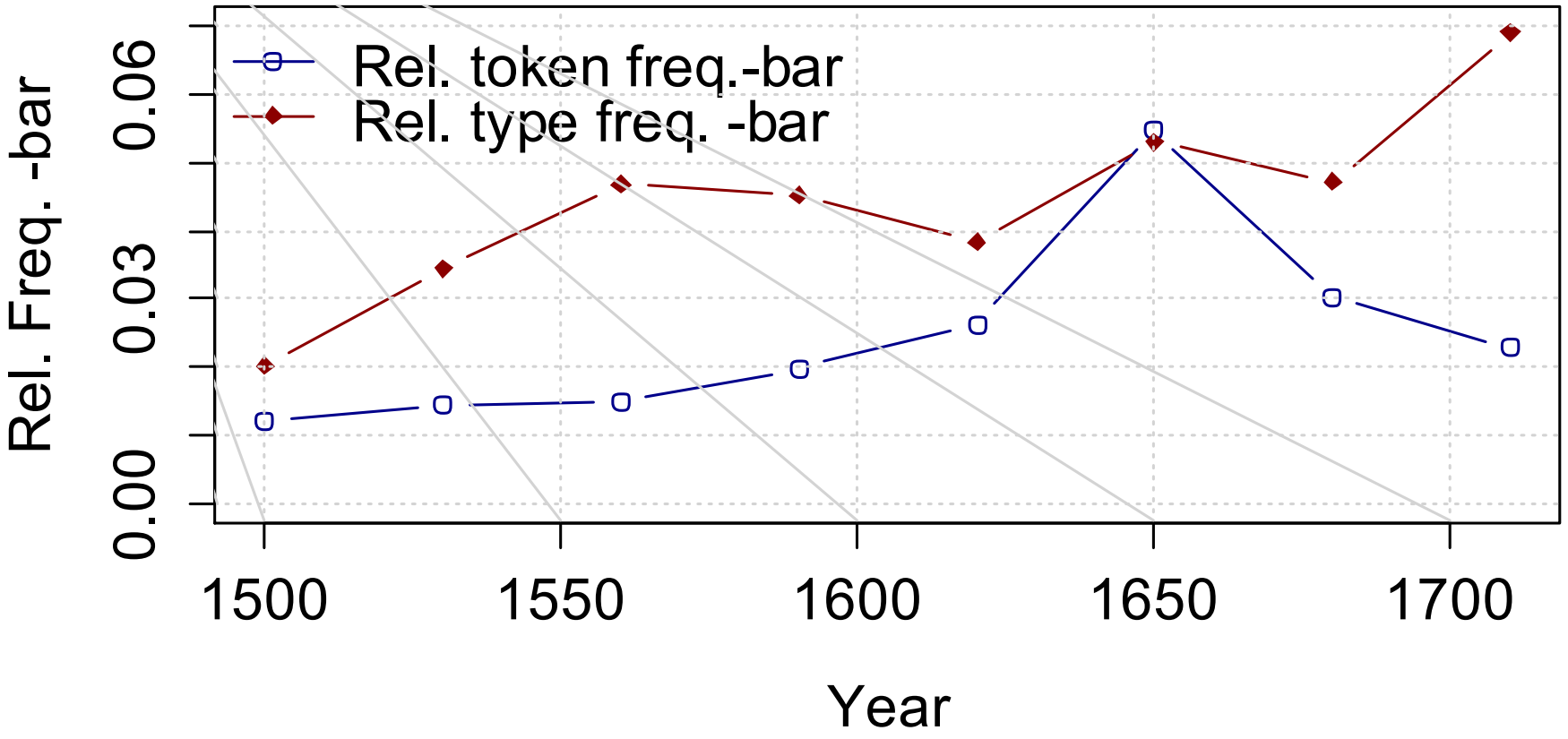
- Konstruktionalisierung in der Wortbildung:
  - Entstehung neuer WB-Konstruktionen
  - Lexikalisierung bestehender WB-Produkte
- Morphologische vs. lexikalische Konstruktionalisierung
- In beiden Fällen entstehen neue Form-Bedeutungs-Paare

Produktivität messen

- Baayen (2009: 902):
- **realized productivity** P = V (C,N)
- **expanding productivity** P = V (1,C,N) / V(1,N)
- **potential productivity** P = V (1,C,N) / N (C)

- V (C,N): Zahl der Types, die zur morphologischen Kategorie C gehören, in einem Korpus mit N Tokens
- V (1,C,N): Hapax Legomena, die zur morphologischen Kategorie C gehören
- V (1,N): Gesamtzahl der Hapax Legomena im Korpus
- N (C): Zahl der Tokens, die zur morphologischen Kategorie C gehören


Potential Productivity
0.00
0.20
-lich
-bar
1660
1700
1740
1780
Year

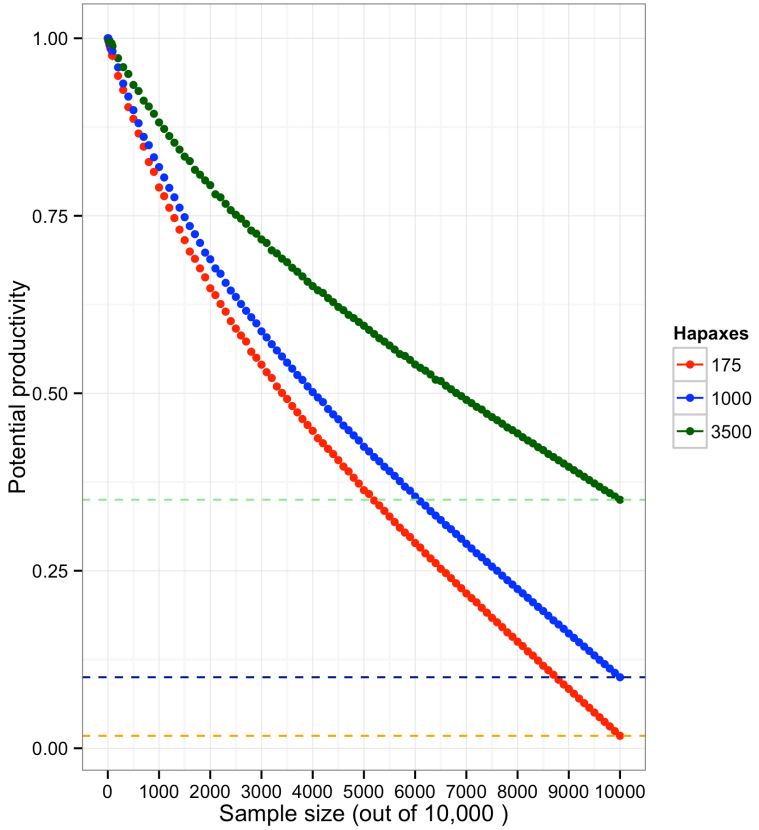

Rel. token freq.-bar
Rel. type freq. -bar
Rel. Freq. -bar
0.00
0.03
0.06
1500
1550
1600
1650
1700
Year

Probleme der Produktivitätsanalyse

- Maß der potentiellen Produktivität ist von der Korpusgröße und von der Anzahl der Tokens abhängig (Baayen 2001; Gaeta & Ricca 2006; Hilpert 2013)
- wenn sich Anzahl der Tokens stark verändert, kann dies den Produktivitätswert verzerren.

Probleme der Produktivitätsanalyse

- Beispiel: Modell mit 10000 fiktiven Wortbildungsprodukten
- Algorithmus zieht Stichproben von sukzessive zunehmender Größe und berechnet potentielle Produktivität
- drei verschiedene Modelle mit unterschiedlicher Anzahl an Hapaxen



Finites Zipf-Mandelbrot-Modell

- Zipfsches Gesetz: Frequenz von Wörtern ist invers proportional zu ihrer Position in Frequenzrangliste
- Mandelbrots Generalisierung des Zipfschen Gesetzes:

$$\pi_z = \frac{C}{(z+b)^a}$$

- Evert (2004), Evert & Baroni (2007): Zipf-Mandelbrot-Modell zur Extrapolation von Produktivitätswerten für größere Datenmengen (vgl. auch Schneider-Wiejowski 2011; Kempf 2016)