

# Korpuslinguistik und Statistik

(050370)

Fabian Barteld, M.A.

Fabian.Barteld@ruhr-uni-bochum.de

8. Sitzung

# Wiederholung

- ▶ **Beobachtungen** sind die Basis der Untersuchung
- ▶ festgehalten als **Ausprägungen** von **Variablen**

## *Beispiel*

Untersuchung des Zusammenhangs von Wortlänge und Wortart

**Beobachtungen:**

z.B. alle Wörter aus einem Korpus (**Belege**)

**Variablen:**

Wortart, Länge, Wortform, Quelle

# Beispiel

Beobachtung	Wortform	Quelle	Wortart	Länge
1	Wortlänge	Wikipedia/Wortlänge, 1	NN	9
2	wird	Wikipedia/Wortlänge, 2	VAFIN	4
3	danach	Wikipedia/Wortlänge, 3	PAV	6
4	bestimmt	Wikipedia/Wortlänge, 4	VVPP	8
5	,	Wikipedia/Wortlänge, 5	\\$,	1

# Daten aufbereiten

Standardform: **Tabelle**

- ▶ pro **Zeile** eine **Beobachtung**
- ▶ pro **Spalte** eine **Variable**

→ **Urliste**

## Datenformat für Tabellen: CSV

```
Beobachtung,Wortform,Quelle,Wortart,Länge  
1,Wortlänge,"Wikipedia/Wortlänge,1",NN,9  
2,wird,"Wikipedia/Wortlänge,2",VAFIN,4  
3,danach,"Wikipedia/Wortlänge,3",PAV,6  
4,bestimmt,"Wikipedia/Wortlänge,4",VVPP,8  
5,"","Wikipedia/Wortlänge,5","\$,",1
```

Spaltentrenner: ,

Texttrenner: "

# Was ist eine Beobachtung? (I)

Satznr.	Satz	Subjektlänge	Objektlänge
1	The younger bachelors ate the nice little cat.	3	4
2	He was looking at the door.	1	2
3	The quick brown fox hit the lazy dog.	4	3

(Gries, 2008, S. 33)

## Was ist eine Beobachtung? (II)

Satznr.	Phrase	Synt. Funktion	Länge
1	The younger bachelors	Subj	3
1	the nice little cat	Obj	4
2	He	Subj	1
2	at the door	Obj	2
3	The quick brown fox	Subj	4
3	the lazy dog	Obj	3

(Gries, 2008, S. 34)

# Literatur

- ▶ *1. Grundlagen empirischer Forschung*  
in Gries, 2008, S. 8–62
- ▶ *2.1 Verschiedene Arten von Daten*  
in Büchter und Henn, 2007, S. 17–27

# Gruppenarbeit

Von der Abfrage zur Tabelle

**Zeit:** bis 9:30

am Ende eine kurze Präsentation:

1. Forschungsfrage
2. Korpus
3. Beschreibung der relevanten Belege
4. Suchanfrage
5. Export der Daten

(außerdem in einer PDF-Datei festhalten)

# Ausblick

Datum	Thema
08.01	Beschreibende Statistik
15.01	Wahrscheinlichkeitstheorie und Schätzer
22.01	Tests
29.01	Abschluss