

# Relationen zwischen Nomen und ihren Assoziationen



Michael Roth

# Assoziationen ...

- ... sind psychologisch interessant.
  - Wie erfolgt der Zugriff auf sie?
  - Welche Bedeutung haben sie?
- ... erfüllen einen linguistischen Zweck.
  - Möglicherweise linguistisch geprägt
  - Sie repräsentieren „irgendwelche“ Relationen
  - Systematisch analysierbar

# Motivation

- Erweiterung semantischer Netze
  - ... durch so entdeckte fehlende Verknüpfungen
- Mögliche Basisstruktur für Thesauri
  - Strukturierung nach assoziativer Verwandtschaft
- Grundlage zur Disambiguierung von Polysemie
  - ... durch (eindeutige) Assoziationen im Kontext

# Datengewinnung

- Studie mit 409 Nomen
- 300 Versuchsteilnehmer
- 118.028 Antworten  
(=Tokens)
- 31.782 unterschiedliche Assoziationen  
(=Types)

Assoziation	Wortart	#
Musik	Nomen	72
Hören	Verb	22
Nachrichten	Nomen	22
Sender	Nomen	13
Laut	Adjektiv	11
Antenne	Nomen	10

Beispiel: *Radio*

# Ressourcen

- GermaNet
- Statistische Grammatik (SSiW)
- Korpus (FR)
- Wörterbuch (DWDS)

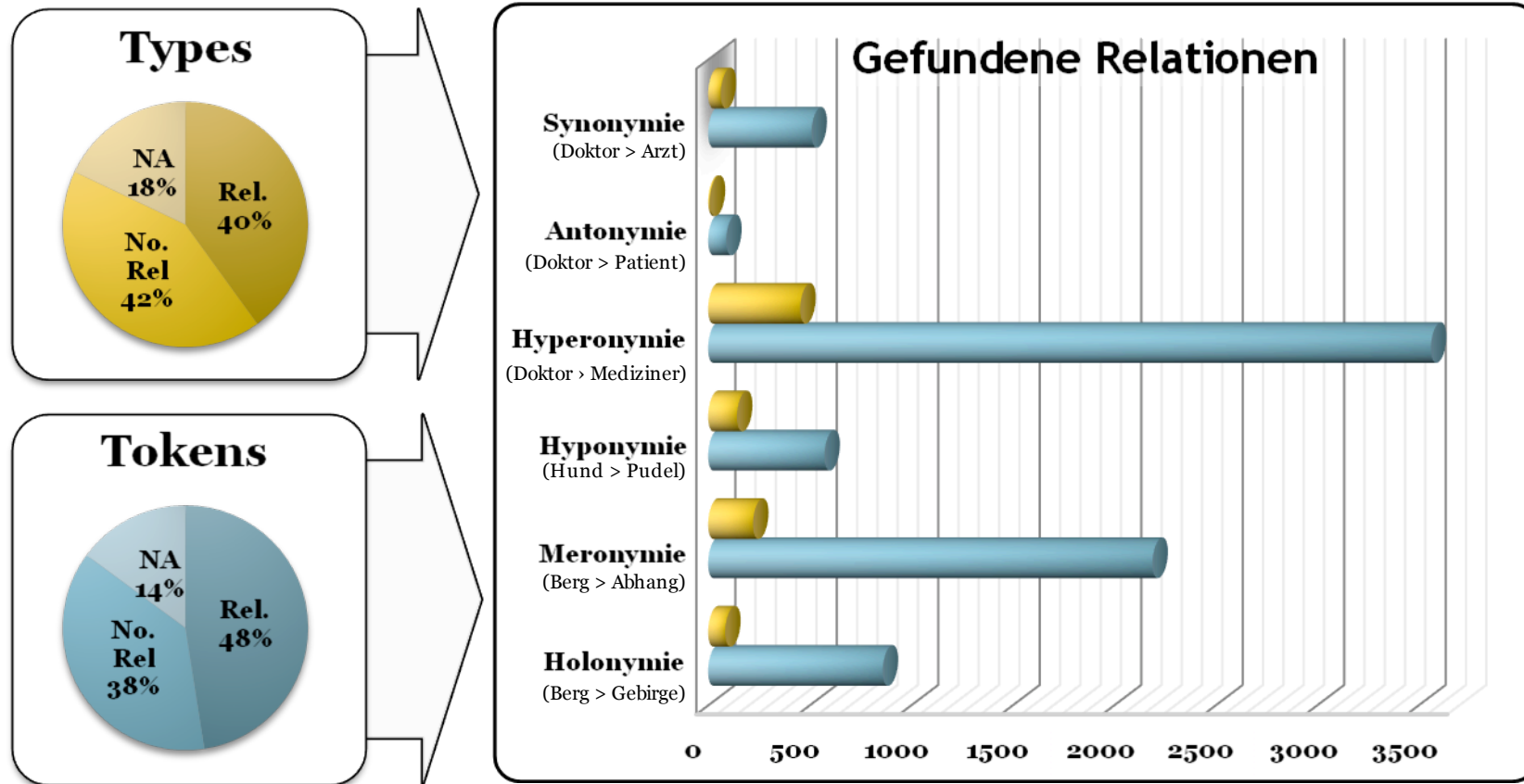
# GermaNet

- Lexikalische Taxonomie
- Synonyme Begriffe als *Synsets*
- Semantische Relationen zwischen Synsets
  - Hyperonymie/Hyponymie
  - Meronymie/Holonymie
  - ...

# GermaNet - Ziele

- Nomenantworten:
  - Finden einer semantischen Relationen zwischen gegebenem Nomen und Assoziation
- Verb- und Adjektivantworten:
  - Finden signifikanter Klassen für das gegebene Nomen anhand der Antwortkategorien

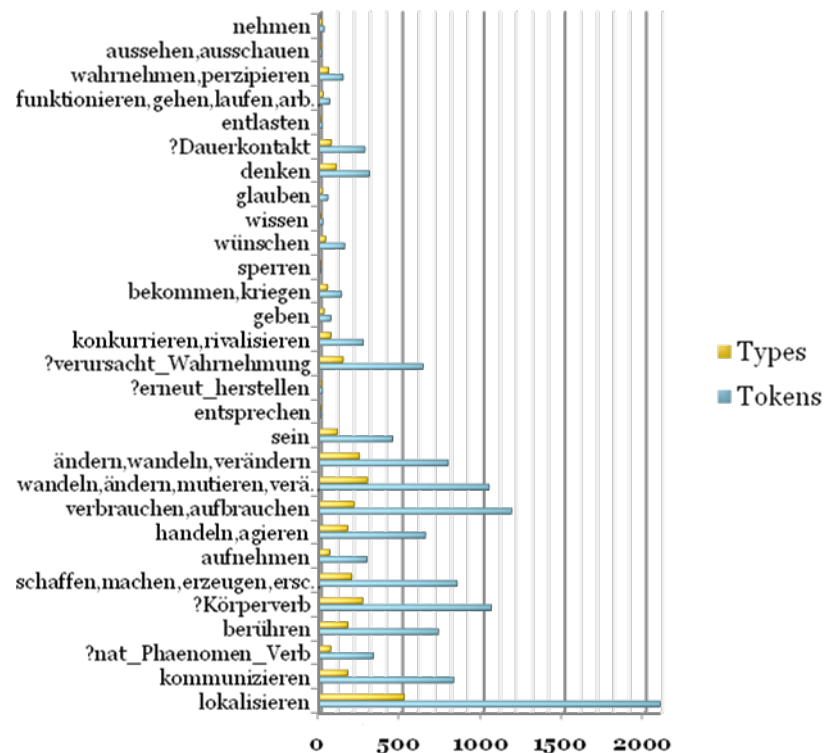
# GermaNet - Ergebnisse (Nomen)



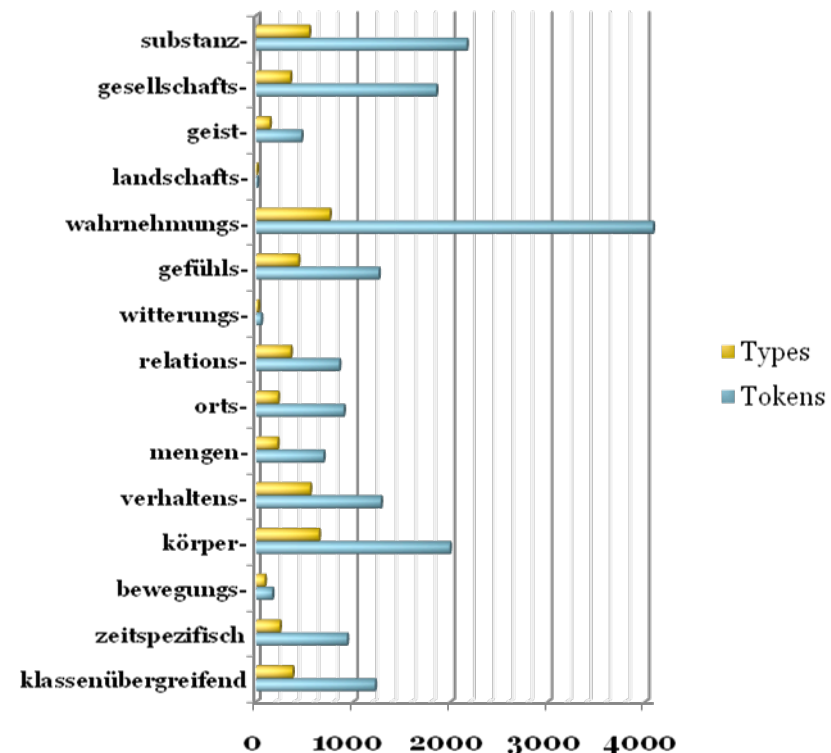


# GermaNet - Ergebnisse (V.&Adj.)

## Signifikante Verbklassen



## Adjektivklassen



# GermaNet - Erkenntnisse

## ... über die Assoziationen

- Hyperonyme sind mit Abstand am häufigsten.
- Meronyme sind häufiger als Holonyme und Hyponyme.
  - **Trotz 4.002 vs. 30.710 Relationen in GermaNet**
- Antonymie und Synonymie sind sehr selten.

## ... über GermaNet

- Teilweise umständliche Beziehungen
  - **z. B. Oma ( > Mutter > Elternteil > Vater) > Opa**
- Unausgewogene Verteilung der Verbklassen
  - **„lokalisieren“ umfasst 1.643 von 8.120 Verben.**

# Statistische Grammatik

- Kontextfreie Grammatik
- Lexikalisiert
- Trainiert auf einem Teil des „HGC“ (35 Mio. Wörter)

## SG - Ziel

- Finden gemeinsamer Auftreten von Nomen und Assoziation in einer syntaktischen Struktur
  - Tauchen die Begriffe gemeinsam auf?
  - Wie wahrscheinlich ist das Auftreten?
- Erwartung:
  - Prototypische Verben und Adjektive werden häufig gefunden (hohe Wahrscheinlichkeit).

# SG - Ergebnisse (1)

## Gefundene Assoziationen:

- 954 (26%)  
Types
- 5834 (42%)  
Tokens

### Verben

(Schraube > drehen,  
Hahn > kraehen,  
Brief > schreiben, ...)

- 937 (16%)  
Types
- 4750 (29%)  
Tokens

### Adjektive

(Salat > gruen,  
Opa > alt,  
Sessel > bequem, ...)

- 973 (5%)  
Types
- 7426 (9%)  
Tokens

### Nomen

(Kaffee > Tasse,  
Kirsche > Kern,  
Bild > Rahmen, ...)

# SG - Ergebnisse (2)

Anteile der Verben mit Rahmen und Nomenrolle

Wahrscheinlichkeit >1%		Types / Tokens			
VPA/Nom	7.5% / 12.9%	VPA/Dat	1.9% / 3.4%	VPP/Akk	2.8% / 2.5%
VPA/Akk	7.1% / 14.3%	VPP/Nom	2.7% / 2.5%	<b>Insgesamt</b>	<b>19.5% / 33.6%</b>

Wahrscheinlichkeit >5%					
VPA/Nom	2.4% / 5.7%	VPA/Dat	1.3% / 2.5%	VPP/Akk	1.8% / 1.7%
VPA/Akk	3.7% / 8.3%	VPP/Nom	1.8% / 1.7%	<b>Insgesamt</b>	<b>9.3% / 18.0%</b>

Wahrscheinlichkeit >10%					
VPA/Nom	0.9% / 2.4%	VPA/Dat	1.2% / 2.2%	VPP/Akk	1.2% / 0.9%
VPA/Akk	2.1% / 4.9%	VPP/Nom	1.2% / 0.9%	<b>Insgesamt</b>	<b>5.2% / 10.4%</b>

# FR-Korpus

- Zeitungskorpus der Frankfurter Rundschau
- Enthält über 300 Mio. Wörter
- Lemmatisiert

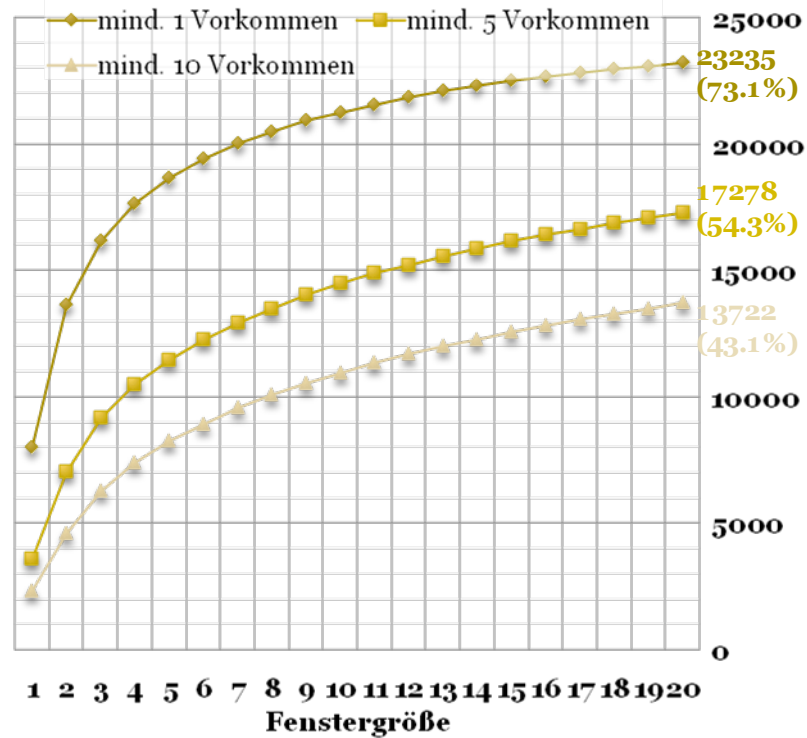
# FR-Korpus - Ziel

- Finden von Kookkurrenzen zwischen gegebenem Nomen und Assoziationen.
- Berechnung der Kollokationsstärke
- Erwartung:
  - Häufige Kookkurrenzen und signifikante Kollokationsstärke für Assoziationen

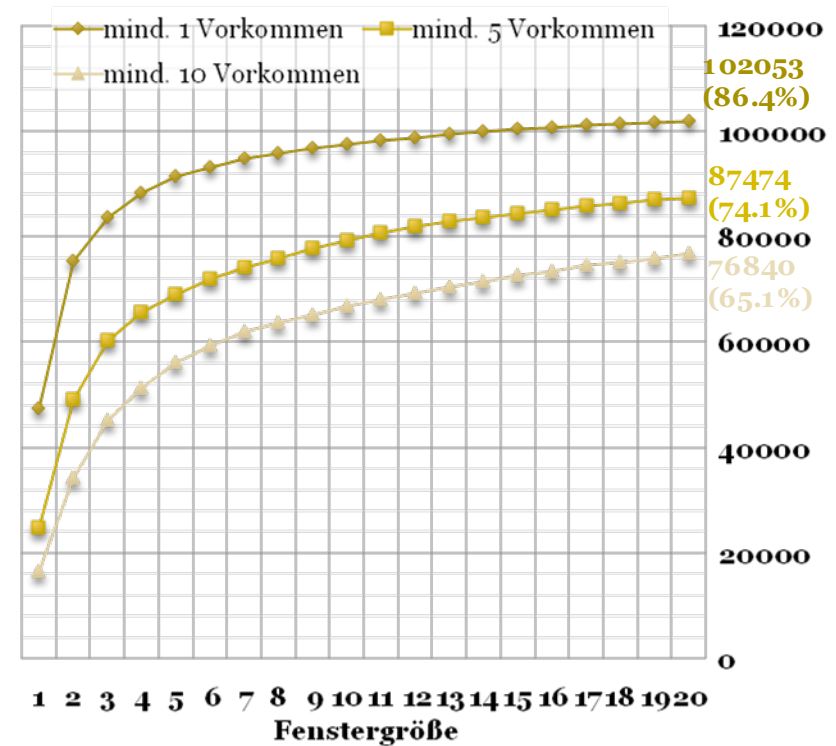


# FR-Korpus - Ergebnisse (1)

## Types



## Tokens



# FR-Korpus - Ergebnisse (1)

		Fenstergröße				
Mindestanzahl von Vorkommen	Types	1	2	5	10	20
	1	8003 (25.1%)	13645 (42.9%)	18662 (58.7%)	21254 (66.8%)	23235 (73.1%)
	5	3606 (11.3%)	7032 (22.1%)	11437 (35.9%)	14502 (45.6%)	17278 (54.3%)
	10	2308 (7.26%)	4631 (14.5%)	8275 (26.0%)	10944 (34.4%)	13722 (43.1%)
	Tokens	1	2	5	10	20
	1	47425 (40.1%)	75528 (63.9%)	91533 (77.5%)	97682 (82.7%)	102053 (86.4%)
	5	24721 (20.9%)	49147 (41.6%)	69107 (58.5%)		
	10	16624 (14.0%)				

# FR-Korpus - Ergebnisse (2)

## Log-Likelihood-Werte

- Median (Types):
  - 3.542
- Median (Tokens):
  - 29.496
- „Top 10“:
  - Frau -> alt
  - Lehrerin -> Aufgabe
  - Kirche -> evangelisch
  - Kirche -> katholisch
  - Schwein -> klug
  - Lampe -> Winter
  - Frau -> Mann
  - Engel -> fliegen
  - Mann -> Frau
  - Mantel -> Hemd

# Wörterbuch

- Digitales Online-Wörterbuch ([www.dwds.de](http://www.dwds.de))
- Großzügige Bedeutungserklärungen
  - Etwa 150 Wörter pro relevantem Eintrag
  - Ausführliche Beschreibungen
  - Sinnvolle Beispielsätze

# Wörterbuch - Ziel

- Finden von genannten Assoziationen im jeweiligen Wörterbucheintrag
- Erwartung:
  - Typische Eigenschaften und Tätigkeiten sollten im Wörterbuch zu finden sein.
  - Ebenso einzelne Hyperonyme (*Genera Proxima*) und einige „Beispiel-Hyponyme“.

# Wörterbuch - Ergebnis

## Gefundene Assoziationen:

- 543 (16%)  
Types
- 4352 (34%)  
Tokens

### Verben

(Besen > kehren,  
Frosch > quaken,  
Mund > essen, ...)

- 765 (15%)  
Types
- 4655 (27%)  
Tokens

### Adjektive

(Schnecke > schleimig,  
Pudel > klein,  
Sonne > gelb, ...)

- 2340 (12%)  
Types
- 22078 (30%)  
Tokens

### Nomen

(Schale > Obst,  
Athlet > Muskel,  
Hose > Bein, ...)

# Zusammenfassung

- Für die meisten Assoziationen können Relationen gefunden werden.
- Einige Assoziationen bleiben linguistisch ungeklärt – Weltwissen?
- Ergebnisse können in der Computerlinguistik Verwendung finden.

# Literatur

- H. H. Clark (1970) – Word Associations and Linguistic Theory. In J. Lyons – *New Horizons in Linguistics*
- S. Schulte im Walde & A. Melinger (2005) – *Identifying Semantic Relations of Functional Properties of Human Verb Associations*
- A. Melinger, S. Schulte im Walde & A. Weber (2006) – *Characterizing Response Types and Revealing Noun Ambiguity in German Association Norms*
- A. Sinopalnikova (2004) – *Word Association Thesaurus as a Resource for Building Wordnet*