

# Kontext

Detmar Meurers  
Seminar für Sprachwissenschaft

Ringvorlesung "Text und Sprache: Theorien am Brechtbau (und am Blochbau)"  
organisiert von Frauke Berndt  
31.5.2011

Kontext Detmar Meurers
Überblick
Kontext und gram. Kategorien
Variationsklassifikation
Kontext zur Übersetzung
Ergebnisse für HELI POS
Sprachwissenbeweiz
Von Wörtern zu anderen grammatischen Kategorien
Außersprachlicher Kontext
Übersetzungsbispiele
Sprachliche Regularitäten und Weltwissen
Parsing
Informationsstruktur
Theorie
Spezifitätsbereich
Pragmatik des Artikel
Generalisierung
Praxis
Bedeutungswegleich
Graviness Elimination
Von Given zu Fokus
Schluss
Problemlöse für Lokalität
Adjunktextraktionen

# Überblick: Drei Streiflichter

- I. Kontext und grammatische Kategorien:
  - Distributionelle Evidenz für Wortarten
  - Wo kommen grammatische Kategorien her?
  - Wie lassen sich linguistisch annotierte Korpora automatisch überprüfen?
- II. Sprache und außersprachlicher Kontext
  - Sprache als linguistisches System vs. Weltwissen
    - Beispiel: Statistische maschinelle Übersetzung
    - Beispiel: Parservergleich mit handgeschriebenen vs. aus Korpora induzierten Grammatiken
- III. Kontext in der linguistische Analyse: Informationsstruktur
  - Linguistisches Fallbeispiel: Reduktion von syntaktisch stipulierten Beschränkungen auf Kontexteigenschaften
  - Informationsstruktur im Bedeutungsvergleich von Antworten auf Leseverständnisfragen (SFB 833/A4)

Kontext Detmar Meurers
Überblick
Kontext und gram. Kategorien
Variationsklassifikation
Kontext zur Übersetzung
Ergebnisse für HELI POS
Sprachwissenbeweiz
Von Wörtern zu anderen grammatischen Kategorien
Außersprachlicher Kontext
Übersetzungsbispiele
Sprachliche Regularitäten und Weltwissen
Parsing
Informationsstruktur
Theorie
Spezifitätsbereich
Pragmatik des Artikel
Generalisierung
Praxis
Bedeutungswegleich
Graviness Elimination
Von Given zu Fokus
Schluss
Problemlöse für Lokalität
Adjunktextraktionen

## Kontext und grammatische Kategorien

- Grammatische Kategorien: Wortarten
  - Konstituenten, Dependenzrelationen, ...
- Wortartenzuweisung (POS tagging) reduziert die lexikalisch möglichen Kategorien eines Wortes auf die für ein bestimmtes Vorkommen korrekte.
- Beispiel *can*: Hilfsverb, Nomen, Verb
  - (1) *I can walk.* (Aux)
  - (2) *Pass me a can of beer.* (Noun)
  - (3) *I can tuna for a living.* (Verb)
- **Variation** in Korpusannotation:  
Sprachliches Material das mehrfach in einem Korpus vorkommt und unterschiedlich annotiert ist
- Variation kann resultieren aus
  - genuine **Ambiguität**
  - inkonsistente, **fehlerhafte Annotation**

Kontext Detmar Meurers
Überblick
Kontext und gram. Kategorien
Variationsklassifikation
Kontext zur Übersetzung
Ergebnisse für HELI POS
Sprachwissenbeweiz
Von Wörtern zu anderen grammatischen Kategorien
Außersprachlicher Kontext
Übersetzungsbispiele
Sprachliche Regularitäten und Weltwissen
Parsing
Informationsstruktur
Theorie
Spezifitätsbereich
Pragmatik des Artikel
Generalisierung
Praxis
Bedeutungswegleich
Graviness Elimination
Von Given zu Fokus
Schluss
Problemlöse für Lokalität
Adjunktextraktionen

## Variationsklassifikation

- Wie kann man solche Variation auffinden und klassifizieren, ob die Annotation fehlerhaft ist?
  - Entscheidender Faktor: sprachlicher Kontext
    - Umso ähnlicher der Kontext eines Vorkommens, desto wahrscheinlicher ist Annotationsvariation inkorrekt.
  - Die einfachste Art, Kontext zu konkretisieren beinhaltet:
    - Oberflächenformen der Wörter
    - in direkter Umgebung der Variation
- ⇒ Extrahiere alle Annotationsvariation mit Kontext
- **Variations-Nucleus**: Wiederkehrendes Material (hier: Wort), das unterschiedlich annotiert wurde
  - **Variations-n-Gram**: Variations-Nucleus mit identischem Kontext von insgesamt  $n$  Worten

Kontext Detmar Meurers
Überblick
Kontext und gram. Kategorien
Variationsklassifikation
Kontext zur Übersetzung
Ergebnisse für HELI POS
Sprachwissenbeweiz
Von Wörtern zu anderen grammatischen Kategorien
Außersprachlicher Kontext
Übersetzungsbispiele
Sprachliche Regularitäten und Weltwissen
Parsing
Informationsstruktur
Theorie
Spezifitätsbereich
Pragmatik des Artikel
Generalisierung
Praxis
Bedeutungswegleich
Graviness Elimination
Von Given zu Fokus
Schluss
Problemlöse für Lokalität
Adjunktextraktionen

## Variation klassifizieren

Beispiel aus dem Wall Street Journal Corpus (Marcus et al. 1993)

- ▶ Variations-12-Gram mit off als Nucleus

(4) *to ward off a hostile takeover attempt by two European shipping concerns*

- ▶ einmal als Präposition (IN) annotiert und
- ▶ einmal als Partikel (RP).

- ▶ Annotationsvariation lässt sich effizient auch in großen Korpora identifizieren und klassifizieren. → DECCA Projekt

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationskategorien

Kontext zur Disambiguierung  
Eigenschaften für NUC/POC

Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettleszen

Paraphrase

Informationsstruktur

Theorie

Spezifitätseffekt  
Prognostik des Artikels  
Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

5 / 42

## Welcher Kontext zur Disambiguierung?

**Beispiel:** Ein Variations-37-Gram mit Nucleus *joined*:

- (5) a. *John P. Karalis ...*  
b. *John P. Karalis has ...*
- joined the Phoenix , Ariz. , law firm of Brown & Bain . Mr. Karalis , 51 , will specialize in corporate law and international law at the 110-lawyer firm . Before joining Apple in 1986 ,*
- ▶ Der Kontext vor dem 37-Gram zeigt: korrekte Ambiguität
  - a. Vergangenheitsform (vø),
  - b. Partizip (vøv).

**Einsicht:** Morphol. & syntakt. Eigenschaften **lokal** registriert.

⇒ Ein Rahmen um den Nucleus herum ist eine guter disambiguierender Kontext: X \_\_ Y

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationskategorien

Kontext zur Disambiguierung  
Eigenschaften für NUC/POC

Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettleszen

Paraphrase

Informationsstruktur

Theorie

Spezifitätseffekt  
Prognostik des Artikels  
Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

6 / 42

## Wie gut funktioniert das in der Realität?

- ▶ Wall Street Journal Corpus, handannotiert (Marcus et al. 1993)
- ▶ Resultierende Variations-*n*-Grammen mit Kontextrahmen:
  - ▶ zwischen 3 und 224 Wörter Länge
  - ▶ 7141 unterschiedliche Typen mit Rahmenkontext
  - ▶ 6626 echte Fehler (92.8% Präzision)

→ Warum ist der Rahmen als Kontext so effektiv?

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationskategorien

Kontext zur Disambiguierung  
Eigenschaften für NUC/POC

Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettleszen

Paraphrase

Informationsstruktur

Theorie

Spezifitätseffekt  
Prognostik des Artikels  
Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

7 / 42

## Unabhängige Evidenz vom Spracherwerb

- ▶ Mintz (2002): Auftreten eines Elements in Rahmen (*frame*) X \_\_ Y führt zu Kategorienbildung bei Erwachsenen
- ▶ Mintz (2003): häufige Rahmen sind robuste Datenquellen in Erstspracherwerbskorpora (CHILDES, MacWhinney 2000):
  - ▶ you put it
  - ▶ you want it
  - ▶ you see it
  - you \_\_\_ it
- ▶ Rahmen-Konzept auch sprachübergreifend validiert:
  - ▶ Französisch (Chemla et al. 2009)
  - ▶ Mandarin (Xiao et al. 2006)

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationskategorien

Kontext zur Disambiguierung  
Eigenschaften für NUC/POC

Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettleszen

Paraphrase

Informationsstruktur

Theorie

Spezifitätseffekt  
Prognostik des Artikels  
Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

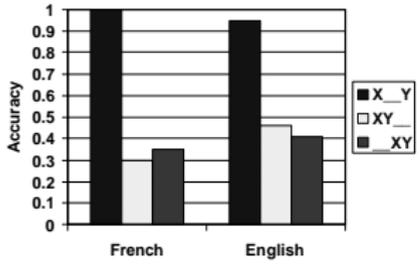
Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

8 / 42

# Kontexte als Trigger für Kategorienbildung

► Chemla et al. (2009): Rahmen um ein Wort besserer Kontext als andere Optionen gleicher Größe



⇒ Automatisierte Fehlersuche in annotierten Korpora verwendet die kognitiven Grundlagen, welche die linguistischen Kategorien überhaupt geschaffen haben.

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Verstärkungsmerkmale

Kontext zur Identifizierung

Ergebnisse für HELI-PCOS

Sprachartenbestimmtheit

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegleitet

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Gewinnens Elimination

Von Given zu Fokus

Schluss

Problemtipps für Lokalität

Adjunktextraktionen

HELI-PCOS UNIVERSITÄT TUBINGEN

9 / 42

# Von Wortarten zu grammatischen Kategorien

DECCA Projekt zur Fehlersuche in "gold standard" Korpora:

- Wortarten (Dickinson & Meurers 2003a)
- Konstituenten (Dickinson & Meurers 2003b; Boyd, Dickinson & Meurers 2007)
- diskontinuierliche Konstituenten (Dickinson & Meurers 2005b)
- transkribierte gesprochene Sprache (Dickinson & Meurers 2005a)
- Dependenzannotation (Boyd, Dickinson & Meurers 2008)

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Verstärkungsmerkmale

Kontext zur Identifizierung

Ergebnisse für HELI-PCOS

Sprachartenbestimmtheit

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegleitet

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Gewinnens Elimination

Von Given zu Fokus

Schluss

Problemtipps für Lokalität

Adjunktextraktionen

HELI-PCOS UNIVERSITÄT TUBINGEN

10 / 42

# Syntaktische Annotation

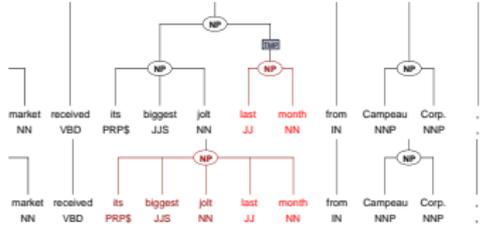
Beispiel aus Penn Treebank 3 (Marcus et al. 1993)

► Variation zwischen zwei syntaktischen Kategorien:

(6) *maturity* next Tuesday

labeled as **NP** twice  
**PP** once

► Variation in der Konstituentenstruktur:



**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Verstärkungsmerkmale

Kontext zur Identifizierung

Ergebnisse für HELI-PCOS

Sprachartenbestimmtheit

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegleitet

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Gewinnens Elimination

Von Given zu Fokus

Schluss

Problemtipps für Lokalität

Adjunktextraktionen

HELI-PCOS UNIVERSITÄT TUBINGEN

11 / 42

# Sprache und außersprachlicher Kontext

- Zur Relation von Bedeutung und Verwendung im Kontext
- "You shall know a word by the company it keeps." (Firth 1957)
- Einsicht ist Basis der distributionellen Semantik
  - Automatische Identifikation von Synonymen (Turney 2001; Turney & Pantel 2010)
  - Essay Bewertung mit Latent Semantic Analysis (Landauer et al. 1998)
  - ...
- Caveat: Korpora erlauben keine klare Unterscheidung zwischen sprachlichen Regularien und Weltwissen!
- Beispiel: statistische maschinelle Übersetzung in Google
  - Verfahren basieren auf Distribution in alignierten Korpora

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Verstärkungsmerkmale

Kontext zur Identifizierung

Ergebnisse für HELI-PCOS

Sprachartenbestimmtheit

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegleitet

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Gewinnens Elimination

Von Given zu Fokus

Schluss

Problemtipps für Lokalität

Adjunktextraktionen

HELI-PCOS UNIVERSITÄT TUBINGEN

12 / 42

Google translate

From: English ▼ To: German ▼ Translate

What do we want in Tübingen?  
Any pedestrians, any bicycles, and  
any cars are welcome.

Google translate

From: English ▼ To: German ▼ Translate

What do we want in Tübingen?  
Any pedestrians, any bicycles, and  
any cars are welcome.

English to German translation

Was wollen wir in Tübingen?  
Jeder Fußgänger, alle Fahrräder und  
keine Autos sind willkommen.

Google translate

From: English ▼ To: German ▼ Translate

any pedestrians  
any cars  
any bicycles

English to German translation

jeder Fußgänger  
keine Autos  
alle Fahrräder

Listen

Google translate

From: English ▼ To: German ▼ Translate

If it is any help, you can rely on me!

Google translate

From: English To: German Translate

English to German translation

If it is any help, you can rely on me!

Wenn es keine Hilfe ist, können Sie sich auf mich verlassen!

[Listen](#)

Google translate

From: English To: German Translate

The Republicans surely do well!  
The Democrats surely do well!  
The Greens surely do well!

Google translate

From: English To: German Translate

English to German translation

The Republicans surely do well!  
The Democrats surely do well!  
The Greens surely do well!

Die Republikaner sicherlich nicht gut!  
Die Demokraten wollen, sicherlich auch!  
Die Grünen wollen sicherlich gut!

Google translate

From: English To: German Translate

Clearly on the web anything can be found.

Google translate

From: English To: German Translate

English to German translation

Clearly on the web anything can be found.

Deutlich zu finden nichts gefunden werden kann.

Google translate

From: English To: German Translate

English to German translation

the web  
on the web  
to find

finden  
zu finden  
zu finden

Google translate

From: German To: English Translate

Ich dich auch.



Google translate

From: German To: English Translate

German to English translation

Ich dich auch.



Love you too.

Listen

- Korpora kodieren
  - sprachliche Regularitäten
  - außersprachliche Regularitäten (Weltwissen)
- Korpusbasiertes maschinelles Lernen
  - lernt Weltwissen automatisch mit
  - unterscheidet nicht zwischen sprachlichem & Weltwissen
- Unterschied zwischen linguistischer Generalisierung und üblicher Verwendung bereits im Parsing sichtbar.

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassifikation

Kontext zur Disambiguierung

Ergebnisse für MaltParser

Sprachwissenbeizenz

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Givenness Elimination

Von Given zu Fokus

Schluss

Problemtfälle für Lokalisierung

Adjunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

25 / 42

- Vergleich von zwei Parsern (Krivanek & Meurers 2011):
  - handgeschriebenes, regelbasiertes WCDG (Foth et al. 2004)
  - korpusinduzierter, statistischer MaltParser (Nivre et al. 2007)

- Ergebnisse auf deutschem CREG-109 Korpus:

Label	#	MaltParser		WCDG	
		Recall	Prec.	Recall	Prec.
Argument-Relationen:					
SUBJ	88	84.09	78.72	86.36	85.39
OBJA	49	65.31	69.57	75.51	74.00
PRED	25	60.00	68.18	60.00	83.33
AUX	22	59.09	86.67	72.73	94.12
Adjunkt-Relationen:					
ADV	40	67.50	57.45	65.00	48.15
PP	27	74.07	52.63	66.67	38.30

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassifikation

Kontext zur Disambiguierung

Ergebnisse für MaltParser

Sprachwissenbeizenz

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Givenness Elimination

Von Given zu Fokus

Schluss

Problemtfälle für Lokalisierung

Adjunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

26 / 42

Am Syntax-Pragmatik Interface: Informationsstruktur

- Die Integration eines Satzes in den Kontext wird anhand der Informationsstruktur untersucht (Fokus-Hintergrund, Thema-Rhema, Given-New, ...).
- Die explizite Repräsentation & Untersuchung potentieller Kontexte eines Satzes birgt neue Erklärungsmöglichkeiten.
- Bislang in der Grammatik stipulierte Eigenschaften können aus der Art der Diskurseinbettung folgen.
- Beispiel: Partielle Konstituenten und ihre Beschränkungen (De Kuthy 2002; De Kuthy & Meurers 2003)

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassifikation

Kontext zur Disambiguierung

Ergebnisse für MaltParser

Sprachwissenbeizenz

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Givenness Elimination

Von Given zu Fokus

Schluss

Problemtfälle für Lokalisierung

Adjunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

27 / 42

- PP Extraktion aus NP zeigt **Spezifitätseffekt**, eine klassische Extraktionsrestriktion (G. Müller 1996):
 

(7) [Über Syntax], hat Karl [**ein**/**das** Buch ...] gelesen.
- Paefel (1993) weist auf akzeptable Gegenbeispiele hin:
 

(8) Über Syntax hat Karl nur **dieses**, aber nicht **jenes** Buch gelesen.
- Idee (De Kuthy 2002): Reduktion des Spezifitätseffekt auf informationsstrukturelle Notwendigkeiten.

**Kontext**  
Diemar-Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassifikation

Kontext zur Disambiguierung

Ergebnisse für MaltParser

Sprachwissenbeizenz

Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätseffekt

Pragmatik des Artikels

Generalisierung

Praxis

Bedeutungsvergleich

Givenness Elimination

Von Given zu Fokus

Schluss

Problemtfälle für Lokalisierung

Adjunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

28 / 42

## Zur Pragmatik definiter Artikel

Zwei Klassen definiter NPs müssen unterschieden werden:

- a) Definite NPs, die als Antezedenz einen Diskursreferenten im vorhergehenden Diskurs haben und somit im Diskurs gegeben sind, cf. *strongly familiar* (Roberts 2003).  
→ sind immer im Hintergrund des Satzes
- b) Definite NPs, die deiktisch, endophorisch, oder semantisch definit sind, cf. *weakly familiar* (Roberts 2003).  
→ können auch im Fokus des Satzes sein (die Frage beantworten)

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassen  
Kontext zur Identifizierung  
Ergebnisse für HEL/PCIS  
Sprachverwendbarkeit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettbewerber  
Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Pragmatik des Artikel

Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

29 / 42

## a) Definite NPs mit Antezedenz im Diskurs

- (9) *Gestern sah ich ein tolles Buch über Syntax in der UB.*
  - a. *Ich habe mir [das Buch über Syntax] heute ausgeliehen.*
  - ▶ Die gesamte definite NP (inklusive der PP) ist gegeben, also Teil des Hintergrundes.
- (10) *Gestern sah ich ein tolles Buch über Syntax in der UB.*
  - a. # *Über Syntax habe ich mir [das Buch] heute ausgeliehen.*
  - ▶ Die PP kann nicht aus der NP extrahiert werden, wenn beide im Hintergrund sind.

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassen  
Kontext zur Identifizierung  
Ergebnisse für HEL/PCIS  
Sprachverwendbarkeit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettbewerber  
Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Pragmatik des Artikel

Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

30 / 42

## b) Definite NPs ohne Antezedenz im Diskurs

- (11) *Was hast Du in der UB ausgeliehen?*
  - a. *Ich habe mir [das Buch über Syntax] ausgeliehen, das Du mir letztlich empfohlen hast.*
  - ▶ Die definite NP trägt als Antwort auf die Frage neue Information bei, steht also im Fokus der Äußerung.
- (12) *Was hast Du über Syntax in der UB ausgeliehen?*
  - a. *Über Syntax habe ich mir [das Buch] ausgeliehen, das Du mir empfohlen hast.*
  - ▶ Definite NP kann im Fokus stehen, mit gegebener PP, und erlaubt dann Extraktion der PP aus der definiten NP!

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassen  
Kontext zur Identifizierung  
Ergebnisse für HEL/PCIS  
Sprachverwendbarkeit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettbewerber  
Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Pragmatik des Artikel

Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

31 / 42

## Generalisierung dank Informationsstruktur

- ▶ **Generalisierung:** Die Extraktion einer PP aus einer NP ist nur möglich, wenn sie zu unterschiedlichen Teilen der Informationsstruktur gehören.
  - ▶ Ausweitung auf andere partielle Konstituenten:
    - ▶ partielle VPs (De Kuthy & Meurers 2003)
- ⇒ Konstituenten kommen nur partiell vor, wenn dies informationsstrukturell motiviert ist.
- ▶ Eine explizite Integration des Kontextes in die linguistische Modellierung bietet interessante neue Erklärungsmöglichkeiten.

Kontext  
Detmar Meurers

Überblick

Kontext und gram. Kategorien

Variationsklassen  
Kontext zur Identifizierung  
Ergebnisse für HEL/PCIS  
Sprachverwendbarkeit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext

Übersetzungshilfen  
Sprachliche Regularitäten und Wettbewerber  
Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Pragmatik des Artikel

Generalisierung

Praxis

Bedeutungsvergleich  
Grunderkenntnis  
Von Given zu Fokus

Schluss

Problemlöse für Lokalisierung  
Adjunktextraktionen

HERIOT-WATSON  
UNIVERSITÄT  
TUBINGEN

32 / 42

# Praktische Relevanz von Informationsstruktur

► **Aufgabe:** Wie kann in realistischen Situationen die Bedeutung von Sätzen und Textfragmenten analysiert und verglichen werden?

- Realistische Situationen:
  - Sprache nicht notwendigerweise wohlgeformt
  - Unterschiede im situativen Wissen oder Weltwissen
  - erschweren vollständige, tiefe Analyse

⇒ Bedeutungsvergleich im Kontext: Komponenten einer flachen semantischen Analyse (SFB 833, Projekt A4)

- Welche Repräsentationen lassen sich robust für eine computerlinguistische Approximation identifizieren?
- Wie lässt sich die Rolle des Kontextes integrieren?

Kontext  
Diemar Meurers

Überblick

Kontext und gram. Kategorien

Verbleiklassifikation

Kontext zur Disambiguierung

Spezifische für NLU / POS

Sprachenbewusstheit

Von Wörtern zu anderen grammatischen Kategorien

Äußersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Programm. des. Artikel

Generalisierung

Praxis

Bedeutungsvergleich

Guinness Elimination

Von Given zu Fokus

Schluss

Problemtypen für Lokalität

Ajunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

33 / 42

# Beispiel für Bedeutungsvergleich

**Q:** Was sind die Kritikpunkte, die Leute über Hamburg äußern?

'What are the objections people have about Hamburg?'

**TA:** Der Gestank von Fisch und Schiffsdiesel an den Kais .

The stink of fish and fuel at the quays .



**SA:** Der Geruch von Fish und Schiffsdiesel beim Hafen .

The smell of fish and fuel at the port .

Kontext  
Diemar Meurers

Überblick

Kontext und gram. Kategorien

Verbleiklassifikation

Kontext zur Disambiguierung

Spezifische für NLU / POS

Sprachenbewusstheit

Von Wörtern zu anderen grammatischen Kategorien

Äußersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Programm. des. Artikel

Generalisierung

Praxis

Bedeutungsvergleich

Guinness Elimination

Von Given zu Fokus

Schluss

Problemtypen für Lokalität

Ajunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

34 / 42

# Elimination von gegebener Information

(13) Welches Beispiel für mögliche Verletzungen der Privatsphäre im Internet erwähnt der Autor?

**TA** Der Autor erwähnt, dass Webseiten, die man besucht hat, von Dritten überwacht werden können.

**LA** Wenn Du im Internet surfst, so können alle besuchten Webseiten durch andere gesehen werden.

- In der Frage gegebene Information sollte bei einem Bedeutungsvergleich nicht berücksichtigt werden.

Kontext  
Diemar Meurers

Überblick

Kontext und gram. Kategorien

Verbleiklassifikation

Kontext zur Disambiguierung

Spezifische für NLU / POS

Sprachenbewusstheit

Von Wörtern zu anderen grammatischen Kategorien

Äußersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Programm. des. Artikel

Generalisierung

Praxis

Bedeutungsvergleich

Guinness Elimination

Von Given zu Fokus

Schluss

Problemtypen für Lokalität

Ajunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

35 / 42

# Von gegebener Information zum Fokus

(14) Ist die Wohnung in einem Neubau oder einem Altbau?

**TA** Die Wohnung ist in einem Neubau .

**SA** Die Wohnung ist in einem Neubau

- Problem: Alle Antwortteile sind in der Frage gegeben!
- Wechsel der Perspektive von Given zu Fokus:
  - Statt gegebene Einheiten nicht zu vergleichen, determinieren was von der Frage erfragt wird (Fokus)!
- Einbettung des Satzes in Kontext (Informationsstruktur) ist wichtig für einen effektiven Bedeutungsvergleich.

Kontext  
Diemar Meurers

Überblick

Kontext und gram. Kategorien

Verbleiklassifikation

Kontext zur Disambiguierung

Spezifische für NLU / POS

Sprachenbewusstheit

Von Wörtern zu anderen grammatischen Kategorien

Äußersprachlicher Kontext

Übersetzungsbegriffe

Sprachliche Regularitäten und Weltwissen

Parsing

Informationsstruktur

Theorie

Spezifitätstheorie

Programm. des. Artikel

Generalisierung

Praxis

Bedeutungsvergleich

Guinness Elimination

Von Given zu Fokus

Schluss

Problemtypen für Lokalität

Ajunktextraktionen

HEINRICH HEINE UNIVERSITÄT TUBINGEN

36 / 42

# Schluss

- Sprachlicher und außersprachlicher Kontext spielt eine zentrale Rolle für linguistische Analyse und Interpretation
- Drei Streiflichter:
  1. distributionelle Induktion grammatischer Kategorien und die Überprüfungsmöglichkeit von Korpusannotation
  2. sprachliches System vs. Weltwissen & was passiert wenn datengetriebene Verfahren den Unterschied verwischen
    - statistische maschinelle Übersetzung
    - Vergleich Ergebnisse regelbasierte und statistische Parser
  3. Informationsstruktur in der linguistischen Analyse
    - Explizite Modellierung der Einbettung eines Satzes in den Kontext bietet neue Erklärungsmöglichkeiten.
    - Relevant für computerlinguistischen Bedeutungsvergleich

**Kontext**  
Dietmar Meurers

**Überblick**

Kontext und gram. Kategorien  
Vertriebsklassifikation  
Kontext zur Disambiguierung  
Ergebnisse für NLU/POS  
Sprachenwettbewerb  
Von Wörtern zu anderen grammatischen Kategorien

**Außersprachlicher Kontext**

Übersetzungsbegleitende  
Sprachliche Regulatorien  
und Weltwissen  
Parsing

**Informationsstruktur**

Theorie  
Spezifikationsfeld  
Programm. det. Artikel  
Generierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

**Schluss**

Probleme für Lokalität  
Adjunktstrukturen

UNIVERSITÄT TüBINGEN

37 / 42

# Literaturverzeichnis

- Boyd, A., M. Dickinson & D. Meurers (2007). Increasing the Recall of Corpus Annotation Error Detection. In *Proceedings of the Sixth Workshop on Treebanks and Linguistic Theories (TLT-07)*. Bergen, Norway. URL <http://purl.org/dm/papers/boyd-et-al-07b.html>.
- Boyd, A., M. Dickinson & D. Meurers (2008). On Detecting Errors in Dependency Treebanks. *Research on Language and Computation* 6(2), 113–137. URL <http://purl.org/dm/papers/boyd-et-al-08.html>.
- Chemia, E., T. H. Mintz, S. Bernal & A. Christophe (2009). Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science* 12(3). URL <http://dx.doi.org/10.1111/j.1467-7687.2009.00825.x>.
- De Kuthy, K. (2002). *Discontinuous NPs in German — A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. Stanford, CA: CSLI Publications. URL <http://ling.osu.edu/~kdk/papers/dekuthy00.html>.
- De Kuthy, K. & W. D. Meurers (2003). The secret life of focus exponents, and what it tells us about fronted verb projections. In S. Müller (ed.), *Proceedings of the Tenth Intl. Conference on HPSG*. Stanford, CA: CSLI Publications, pp. 97–110. URL <http://purl.org/dm/papers/dekuthy-meurers-hpsg03.html>.
- Dickinson, M. & W. D. Meurers (2003a). Detecting Errors in Part-of-Speech Annotation. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03)*. Budapest, Hungary, pp. 107–114. URL <http://purl.org/dm/papers/dickinson-meurers-03.html>.

**Kontext**  
Dietmar Meurers

**Überblick**

Kontext und gram. Kategorien  
Vertriebsklassifikation  
Kontext zur Disambiguierung  
Ergebnisse für NLU/POS  
Sprachenwettbewerb  
Von Wörtern zu anderen grammatischen Kategorien

**Außersprachlicher Kontext**

Übersetzungsbegleitende  
Sprachliche Regulatorien  
und Weltwissen  
Parsing

**Informationsstruktur**

Theorie  
Spezifikationsfeld  
Programm. det. Artikel  
Generierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

**Schluss**

Probleme für Lokalität  
Adjunktstrukturen

UNIVERSITÄT TüBINGEN

37 / 42

- Dickinson, M. & W. D. Meurers (2003b). Detecting Inconsistencies in Treebanks. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT-03)*. Växjö, Sweden, pp. 45–56. URL <http://purl.org/dm/papers/dickinson-meurers-ilt03.html>.
- Dickinson, M. & W. D. Meurers (2005a). Detecting Annotation Errors in Spoken Language Corpora. In *The 10th Special Session on treebanks for spoken language and discourse at NODALIDA-05*. Joensuu, Finland. URL <http://purl.org/dm/papers/dickinson-meurers-nodalida05.html>.
- Dickinson, M. & W. D. Meurers (2005b). Detecting Errors in Discontinuous Structural Annotation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL05)*. pp. 322–329. URL <http://aclweb.org/anthology/P05-1040>.
- Foth, K. A., M. Daum & W. Menzel (2004). A broad-coverage parser for German based on defeasible constraints. In *In KONVENS 2004, Beiträge zur 7. Konferenz zur Verarbeitung natürlicher Sprache*, pp. 45–52. URL <http://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.59.6728&rep=rep1&type=pdf>.
- Krivanek, J. & D. Meurers (2011). Comparing Rule-Based and Data-Driven Dependency Parsing of Learner Language. In *Proceedings of the Intern. Conference on Dependency Linguistics (DEPLING 2011)*. Barcelona. Submitted.
- Laudauer, T., P. Foltz & D. Laham (1998). An Introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259–284. URL <http://lsa.colorado.edu/papers/dp1.LSAintro.pdf>.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Vol 1: The Format and Programs, Vol 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates, 3rd ed.

**Kontext**  
Dietmar Meurers

**Überblick**

Kontext und gram. Kategorien  
Vertriebsklassifikation  
Kontext zur Disambiguierung  
Ergebnisse für NLU/POS  
Sprachenwettbewerb  
Von Wörtern zu anderen grammatischen Kategorien

**Außersprachlicher Kontext**

Übersetzungsbegleitende  
Sprachliche Regulatorien  
und Weltwissen  
Parsing

**Informationsstruktur**

Theorie  
Spezifikationsfeld  
Programm. det. Artikel  
Generierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

**Schluss**

Probleme für Lokalität  
Adjunktstrukturen

UNIVERSITÄT TüBINGEN

37 / 42

- Marcus, M. P., M. A. Marcinkiewicz & B. Santorini (1993). Building a Large Annotated Corpus of English: the Penn Treebank. *Computational Linguistics* 19(2), 273–290. URL <http://aclweb.org/anthology/J93-2004>.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition* 30, 678–686.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91–117.
- Müller, G. (1996). *Incomplete Category Fronting*. Habilitationsschrift, Universität Tübingen, Tübingen. Published as SIS-Report 1-96.
- Nivre, J., J. Nilsson, J. Hall, A. Chanev, G. Eryigit, S. Kübler, S. Marinov & E. Marsi (2007). MailParser: A Language-Independent System for Data-Driven Dependency Parsing. *Natural Language Engineering* 13(1), 1–41. URL <http://w3.msi.vxu.se/~nivre/papers/nle07.pdf>.
- Patel, J. (1993). Ein Überblick über die Extraktion aus Nominalphrasen im Deutschen. In F.-J. d'Ávis, S. Beck, U. Lutz, J. Patel & S. Trisler (eds.), *SFX 340 Nr. 34*, pp. 191–245.
- Roberts, C. (2003). Uniqueness in Definite Noun Phrases. *Linguistics and Philosophy* 26, 287–350.
- Turney, P. (2001). Mining the Web for Synonyms: PMI-Ir Versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001)*. Freiburg, Germany, pp. 491–502.
- Turney, P. D. & P. Pantel (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* 37, 141–188.

**Kontext**  
Dietmar Meurers

**Überblick**

Kontext und gram. Kategorien  
Vertriebsklassifikation  
Kontext zur Disambiguierung  
Ergebnisse für NLU/POS  
Sprachenwettbewerb  
Von Wörtern zu anderen grammatischen Kategorien

**Außersprachlicher Kontext**

Übersetzungsbegleitende  
Sprachliche Regulatorien  
und Weltwissen  
Parsing

**Informationsstruktur**

Theorie  
Spezifikationsfeld  
Programm. det. Artikel  
Generierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

**Schluss**

Probleme für Lokalität  
Adjunktstrukturen

UNIVERSITÄT TüBINGEN

37 / 42

Überblick

Kontext und gram. Kategorien  
Verbletklassifikation  
Kontext zur Identifizierung  
Ergebnisse für H&L F&S  
Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext  
Übersetzungshilfen  
Sprachliche Regularitäten und Weltwissen  
Parsing

Informationsstruktur  
Theorie  
Spezifitätsfeld  
Prognost. det. Artikel  
Generalisierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

Schluss

Prozessliste für Lokalen  
Adjunktextraktionen

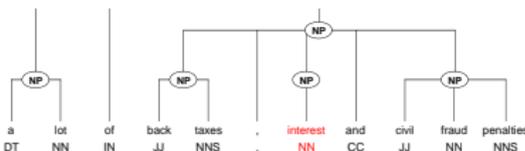
Problemfälle

Theoriegetriebene Unterscheidungen

interest in einer flachen koordinierten Struktur:



interest in einer Koordination mit komplexen Schwestern:



Überblick

Kontext und gram. Kategorien  
Verbletklassifikation  
Kontext zur Identifizierung  
Ergebnisse für H&L F&S  
Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext  
Übersetzungshilfen  
Sprachliche Regularitäten und Weltwissen  
Parsing

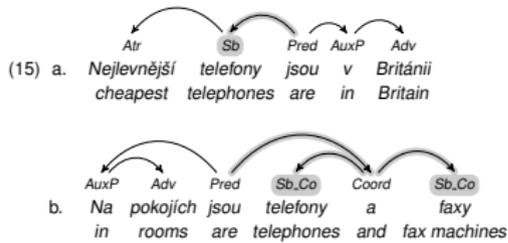
Informationsstruktur  
Theorie  
Spezifitätsfeld  
Prognost. det. Artikel  
Generalisierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

Schluss

Prozessliste für Lokalen  
Adjunktextraktionen

Problemfälle

Nicht lokal motivierte Annotation: Koordination



Überblick

Kontext und gram. Kategorien  
Verbletklassifikation  
Kontext zur Identifizierung  
Ergebnisse für H&L F&S  
Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext  
Übersetzungshilfen  
Sprachliche Regularitäten und Weltwissen  
Parsing

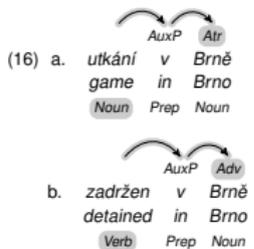
Informationsstruktur  
Theorie  
Spezifitätsfeld  
Prognost. det. Artikel  
Generalisierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

Schluss

Prozessliste für Lokalen  
Adjunktextraktionen

Problemfälle

Nicht lokal motivierte Annotation: Präpositionen



Überblick

Kontext und gram. Kategorien  
Verbletklassifikation  
Kontext zur Identifizierung  
Ergebnisse für H&L F&S  
Sprachbewusstheit  
Von Wörtern zu anderen grammatischen Kategorien

Außersprachlicher Kontext  
Übersetzungshilfen  
Sprachliche Regularitäten und Weltwissen  
Parsing

Informationsstruktur  
Theorie  
Spezifitätsfeld  
Prognost. det. Artikel  
Generalisierung  
Praxis  
Bedeutungsvergleich  
Grueness Elimination  
Von Given zu Fokus

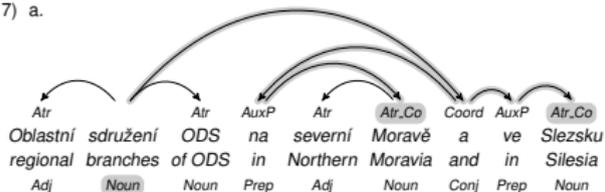
Schluss

Prozessliste für Lokalen  
Adjunktextraktionen

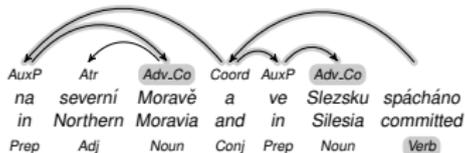
# Problemfälle

## Signifikante Distanz bei Nicht-Lokalität

(17) a.



b.



# Korpusbeispiele für Extraktion aus NP

- (18) Aus dem English Theater stehen zwei Modelle in den Vitrinen.
- (19) Aus dem 17. Jahrhundert erklingen in dynamisch differenziertem Spiel und mit weich gestaltendem Ansatz Tanzsätze von JCP und MP
- (20) In Cockpit und Kabine wurden neue Inhaltsstrukturen mit "marktkonformen" Anfangsgehältern vereinbart.