



# CLARIN-D-Kurationsprojekt: Linguistische Annotation von Nichtstandardvarietäten Guidelines und „Best Practices“ (F-AG)

Burkhard Dietterle<sup>\*</sup>, Stefanie Dipper<sup>+</sup>,

Anke Lüdeling<sup>\*</sup>, Marc Reznicek<sup>++</sup>,

<sup>+</sup>Ruhr-Universität Bochum, <sup>\*</sup>Humboldt-Universität zu Berlin



# Übersicht

Hintergrund: Projektziele, Korpusdesign

A: Vorverarbeitung und Normalisierung

B: Konversion von TiGer-Konstituenten zu  
NoSta-D-Dependenzen

C: Phänomene

## Hintergrund: Projektziele, Korpusdesign



## Linguistische Annotation von Nichtstandard-Varietäten. Guidelines und „Best Practices“ (F-AG 7)

- Annotationschemata, Guidelines und automatische Tools basieren auf Zeitungssprache (Standardvarietät)
- **Pilotprojekt: Erweiterung existierender Ressourcen für**
  - **5 Nichtstandard-Varietäten (plus Zeitungssprache)**
  - **3 Annotationsebenen**
    - Abhängigkeitsanalyse
    - Named Entities
    - Koreferenz

### L2-Lerneraufsätze:

Falko

6,762 Tokens

**Falko**

- Wortstellungsabweichungen
- kreative Wortbildung
- nicht-kanonische Argumentstruktur
- abweichende morphologische Markierung

### gespr. Map Tasks:

BeMaTac

6,731 Tokens

- Äußerungen ohne Verb
- Unflüssigkeiten wie Selbstkorrekturen, Wiederholungen
- Anakoluthe



### Chat-Protokolle:

DCK – Plauderchat

6,664 Tokens

tu technische universität  
darmstadt  
Dortmunder Chat-Korpus

- nicht-standardisierte Schreibungen
- Inflektive ( $V_{end}$ )
- Asterisk-Ausdrücke
- @-Adressierung
- Emoticons

### Literarische Prosa:

Kafka – Der Prozeß

7,294 Tokens



- mehrfache Argumentbesetzung
- komplexe Parenthesen

### Zeitung:

TüBa-DZ

5,000 Tokens



**Standardvarietät**

### Historische Texte:

DDB & Anselm

2,348 + 4,705 Tokens

- keine Satzsegmentierung
- freiere Wortstellung
- nicht-standardisierte Schreibung



(Dipper et al. 2013) <https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



# Guidelines & Best Practices

- Dokumentation der Datenverarbeitung
- Evaluation bestehender Annotationsrichtlinien
- Erweiterung der Annotationsschemata für NoSta-D-Varietäten

### Named Entity

---

- STTS  
(Schiller et al. 1999)
- TüBa-D/Z  
(Telljohan et al. 2012)
- MUC-6  
(Grishman 1995)

### Dependenz

---

- Constraint-Dependenz-Grammatik  
(Foth 2006)
- TiGer  
(Albert et al. 2003)

### Koreferenz

---

- TüBa-D/Z  
(Naumann 2007)
- PoCoS (erweitertes Schema)  
(Kaupat et al. 2013)

- NoSta-D liegt im Moment in TCF und TSV (CONLL) vor
- Lizenz CC-BY
- CLARIN-Stuttgart finanziert eine SHK für die Programmierung eines Konverters von TCF zu RelAnnis (via SaltNPepper) – dann werden die Daten über Annis durchsuchbar sein

- die Guidelines sind auf der Homepage veröffentlicht

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>

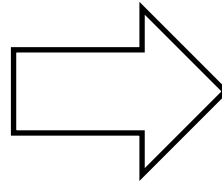


## A: Vorverarbeitung und Normalisierung

- Linearisierung
- Segmentierung
- (Re)tokenisierung
- Normalisierung

## Überlappende Redebeiträge in Map-Tasks

Sprecher 1	Sprecher 2
äh und	
gehst jetzt nach	warte mal Tür is noch nich ganz zu



Sprecher 1	Sprecher 2
Sprecher 1	äh und
Sprecher 1	gehst jetzt nach
Sprecher 2	warte mal Tür is noch nich ganz zu

- Linearisierung
- **Segmentierung**
- (Re)tokenisierung
- Normalisierung

## Fehlende Satzgrenzen in historischen Texten

### ■ Pro Segment: 1 Matrixsatz mit Abhängigen

B1_1v,16	bifchhof tete fente anshel(=)	sente anshelmus bat marien manch iar
B1_1v,17	m <sup>9</sup> bat marien manch	myt heysen trenen · das sy ym offenbarte
B1_1v,18	iar myt heysen trenen·	wy vnser here ihus cristus syne marter
B1_2r,01	das fy ym offenbarte wy	irleden hatte
B1_2r,02	vnser here ihus cristus	
B1_2r,03	fyne marter irleden hatte	
B1_2r,04	do sprach vnse vrouwe Anf=	do sprach vnse vrouwe Anshelme ich sage
B1_2r,05	helme ich fage dir das	dir das myn here ihesus cristus · alzo grose
B1_2r,06	myn here ihesus cristus·	martir irleden hot · das sy nyrkeyn
B1_2r,07	alzo grose martir irleden	mensche usgelegen mak ¶
B1_2r,08	hot· das fy nyrkeyn men=	
B1_2r,09	fche us gelegen mak ¶ Doch	
B1_2r,10	faltu wiffen· daz ich an fo ta=	
B1_2r,11	ne wirdekeit komen byn·	Doch salt u wissen · daz ich an sotane
B1_2r,12	das ich nvmmermer be=	wirdekeit komen byn · das ich
B1_2r,13	trubet mak werden ¶ dar(=)	nvmmermer betrubet mak werden ¶

- Linearisierung
- Segmentierung
- **(Re)tokenisierung**
- Normalisierung

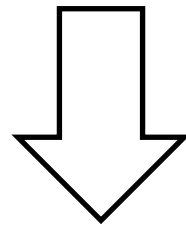
- Tokenisierung aus ursprünglichen Korpora übernommen
- in einigen Fällen Retokenisierung

# Retokenisierung: Konkatenationen

16

- Lexeme werden in einzelne Tokens getrennt
- mit „|“ markiert

165 quaki \*nagut50cmlaufaufleine\*



165 quaki \* na| gut| 50| cm| laufaufleine| \*





# Retokenisierung: @-Adressierung

*mit dem ast in der hand im teich am rumsitzen @stoeps*

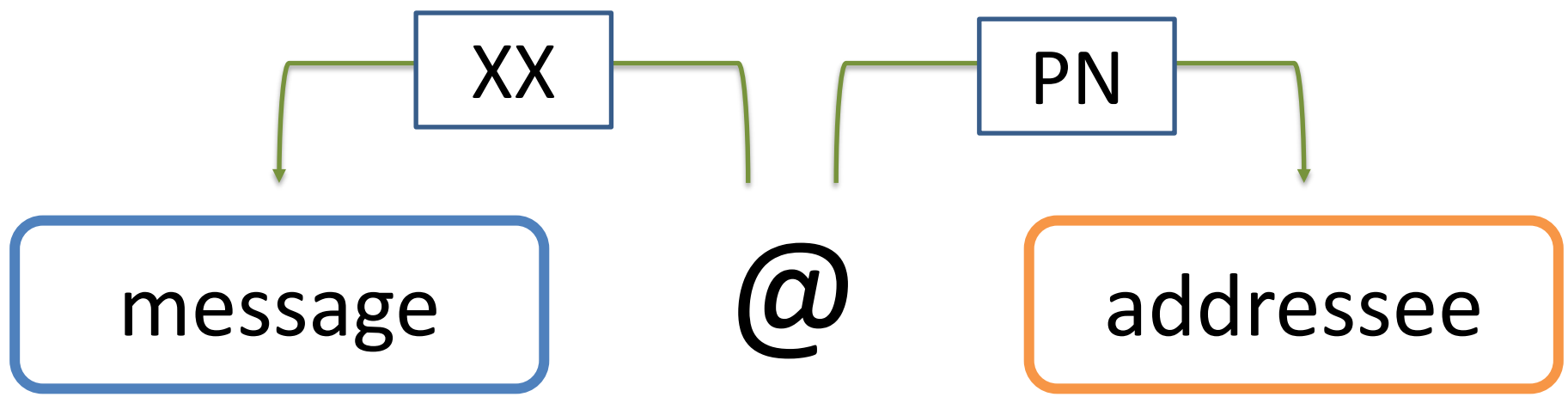
mit dem ast in der hand im teich am rumsitzen @ stoeps



# Retokenisierung : @-Adressierung

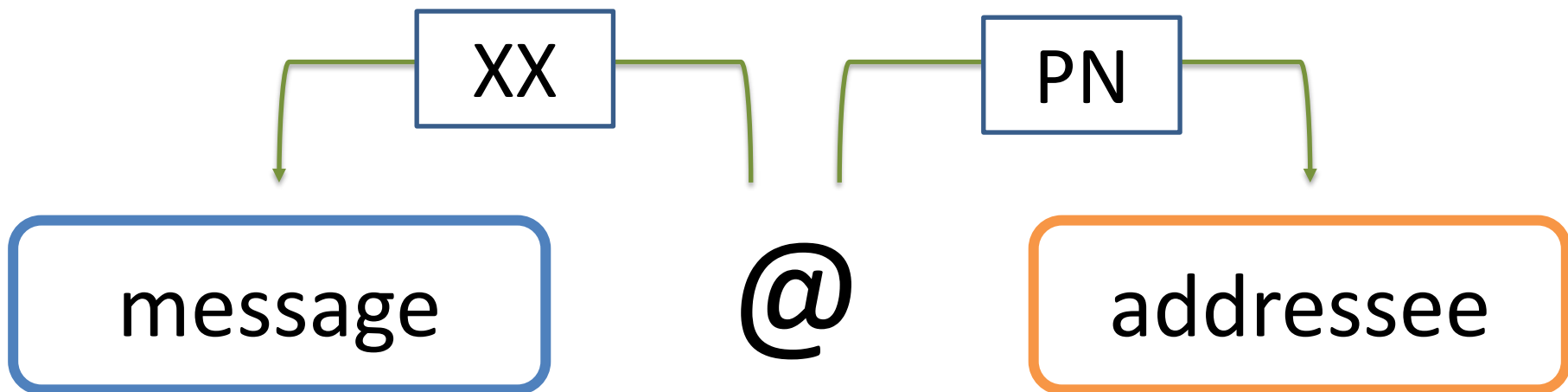
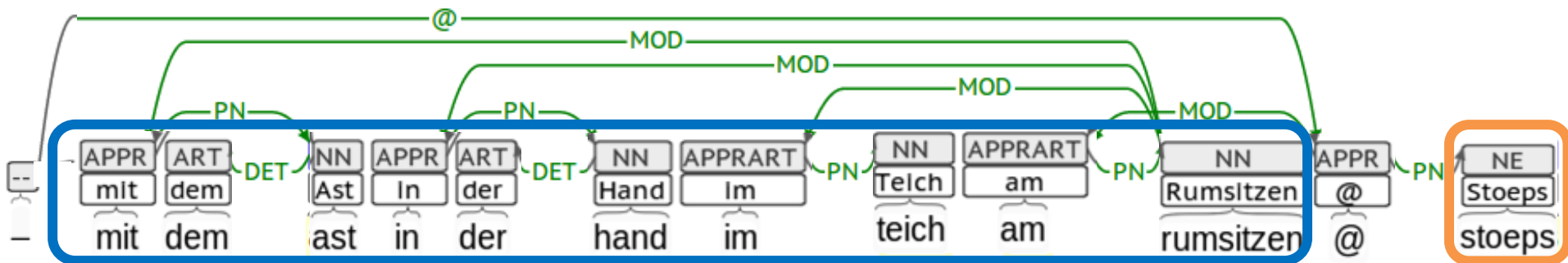
*mit dem ast in der hand im teich am rumsitzen @ stoeps*

mit dem ast in der hand im teich am rumsitzen @ stoeps



# Retokenisierung : @-Adressierung

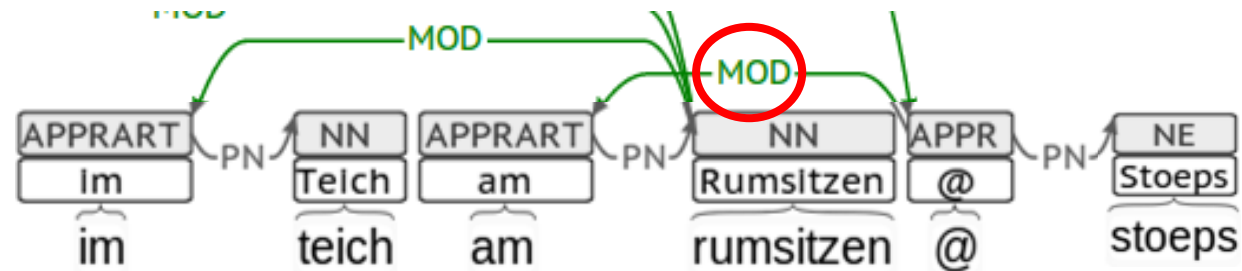
*mit dem ast in der hand im teich am rumsitzen @stoeps*



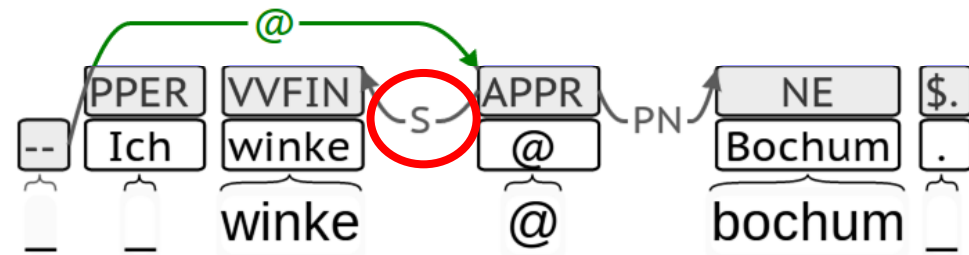
# Retokenisierung : @-Adressierung

## @-adressierte Argumente aus verschiedenen Kategorien

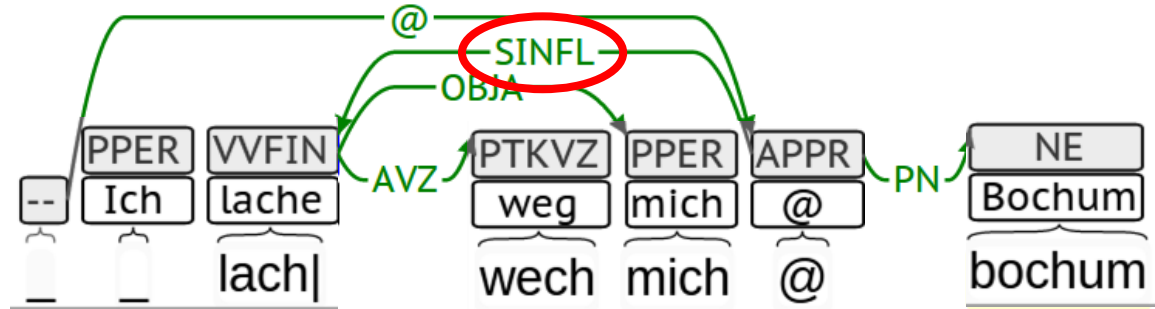
Modifikator: **MOD**



Finiter Satz: **S**



Inflektiver Satz:  
**SINFL**



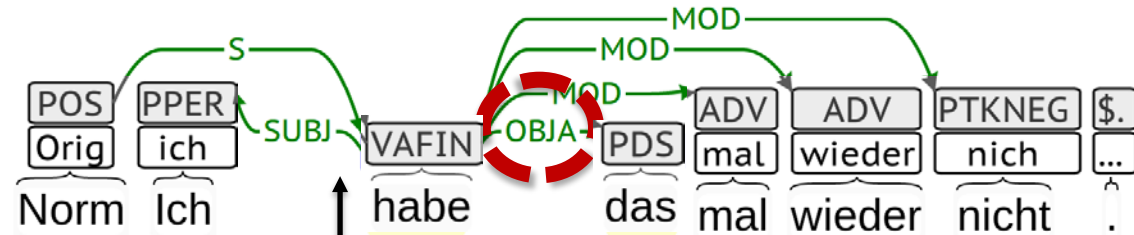
- Linearisierung
- Segmentierung
- (Re)tokenisierung
- **Normalisierung**

# Normalisierung: Perspektiven

## 1 Variationistischer Ansatz:

Wie variiert die Realisierung von Subjekten zwischen Varietäten?

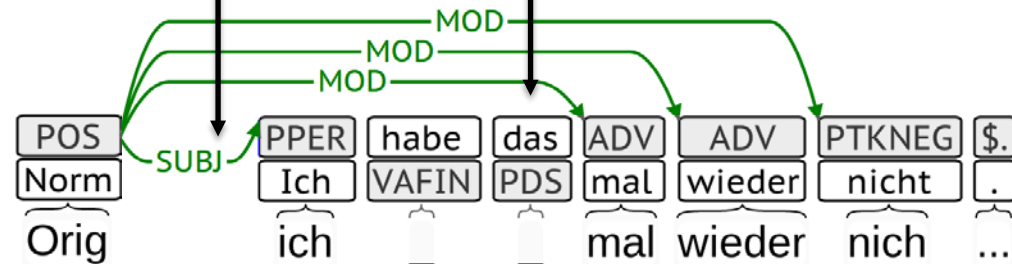
- Unabhängige Annotation von Originaltext und Normalisierung



## 2 Computerlinguistischer Ansatz:

Wie annotiert man Fragmente?

- Annotation des Originaltextes in Abhängigkeit von der Annotation der Normalisierung



# Normalisierung: Perspektiven

	<b>Variationistisch</b>	<b>Computerlinguistisch</b>
Bezeichnung der Textebene	(weite) Normalisierung, Zielhypothese	(enge/technische) Normalisierung
Realisierung in NoSta-D	<b>nein</b>	<b>ja</b>
Zweck	Ableitung vergleichender Aussagen über die Varietäten von Originaltext und Normalisierung	Annotation des Originaltexts
Status der Normalisierung als Varietät	selbst eine Varietät	selbst keine Varietät
Abhängigkeit von Forschungsfrage	ja	nein
Abhängigkeit von zu nutzen- den Annotationsschemata und Korpustools	nein	ja

# Normalisierung: Perspektiven

	<b>Variationistisch</b>	<b>Computerlinguistisch</b>
Bezeichnung der Textebene	(weite) Normalisierung, Zielhypothese	(enge/technische) Normalisierung
Kanonizität in Bezug auf ein zu nutzendes Annotationsschema	nicht notwendigerweise	notwendigerweise
Notwendigkeit sekundärer enger Normalisierung	weite Normalisierung muss selbst ebenfalls annotiert sein und bedarf dazu einer eigenen engen Normalisierung (!)	nein
Ausmaß der Abweichung vom Originaltext	so viel wie nötig	so wenig wie möglich
Überlappung von weiter und enger Normalisierung	genau dann, wenn weite Normalisierung gemäß derjenigen Varietät erfolgte, die als Referenzvarietät des zu nutzenden Annotationsschemas bzw. als Trainingsvarietät der Korpus-tools diene	



## Named Entity:

- **Uneinheitliche Namensschreibung**

Einheitliche Schreibung erleichtert konsistente Klassifizierung.

→ **Normalisierung auf Sprecheralias (Chat), Lexikoneinträge**

## Dependenz:

- **Satzfragmente**

Nur Verben können grammatische Rolle im Satz verteilen.

→ **Explizitmachung von Auslassungen und Ellipsen**

## Koreferenz:

- **Nicht/falsch an der sprachlichen Oberfläche realisierte Referenzen**

Explizitmachung nicht realisierter bzw. Korrektur falsch realisierter Referenten erlaubt Einbindung in referenzielle Kette

→ **Explizitmachung bzw. Korrektur von Referenzausdrücken**

# Normalisierung: Named Entities

## Regel (Chat): Normalisierung = Alias → PER

#	Sprecher	Original	Normalisierung
25	system	[Lantonie]PER betritt den Raum.	[Lantonie]PER betritt den Raum.
26	Lantonie	:)	:)
27	quaki	[lantonieeeee]PER	[Lantonie]PER!
28	Lantonie	Hallo. :)	Hallo :)
29	zora	[LANTOOO]PER :)))	[Lantonie]PER :)))
34	marc30	[Lantöööö]PER :o)	[Lantonie]PER :o)
35	TomcatMJ	hi [lanto]PER	Hi, [Lantonie]PER!

## Normalisierung = Organisation → ORG

#	Sprecher	Original	Normalisierung
429	Emon	boah... ich bekomme echt augenkrebs von bochum...	Boah! Ich bekomme echt Augenkrebs von Bochum.
436	Thor...	ich vom <b>[has]ORG</b> emon *g*	Ich vom <b>[HSV]ORG</b> , Emon *g*
439	Thor...	<b>[hasv]ORG</b>	<b>[HSV]ORG</b>
453	Emon	<b>[hsv]ORG</b> heisst dat	<b>[HSV]ORG</b> heißt das.

221006\_unicum\_21-02-2003\_1

# Normalisierung: Dependenz

Explizitmachung prototypischer Verben mit passender Argumentstruktur in verblosen Sätzen  
→ **Motivation der syntaktischen Funktion von Fragmenten**

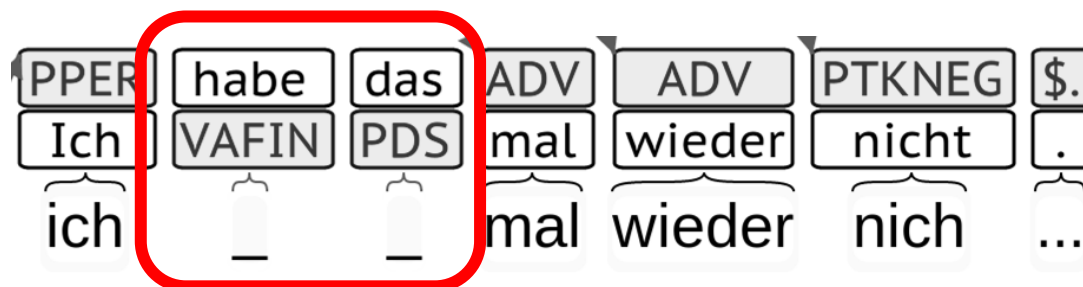
## Original



# Normalisierung: Dependenz

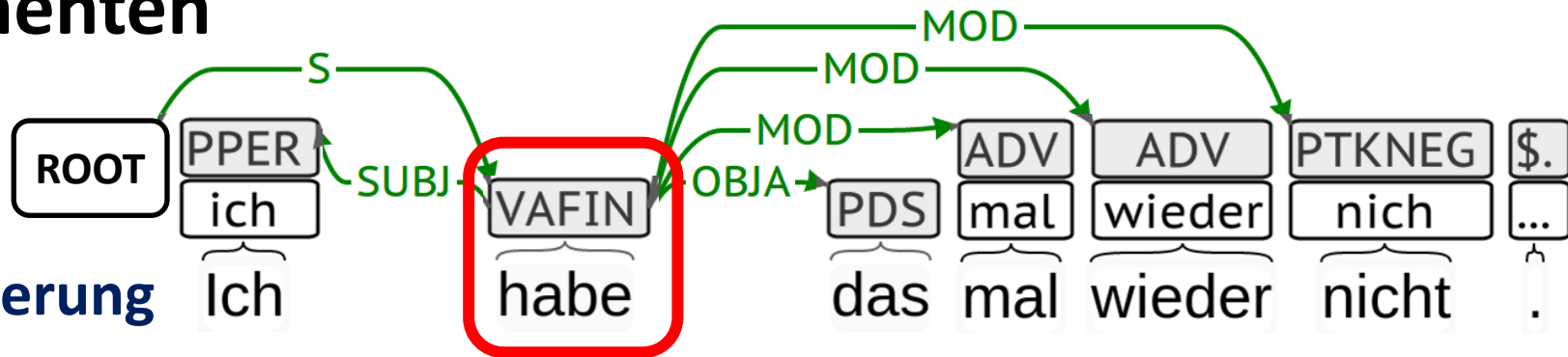
Explizitmachung prototypischer Verben mit passender Argumentstruktur in verblosen Sätzen  
→ **Motivation der syntaktischen Funktion von Fragmenten**

## Original

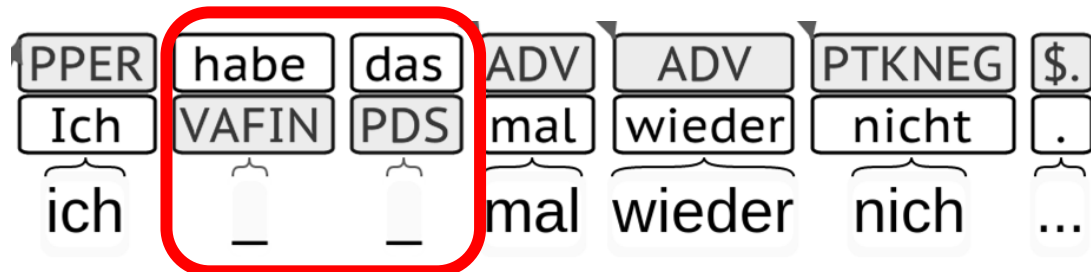


# Normalisierung: Dependenz

Explizitmachung prototypischer Verben mit passender Argumentstruktur in verblosen Sätzen  
 → **Motivation der syntaktischen Funktionen von Fragmenten**



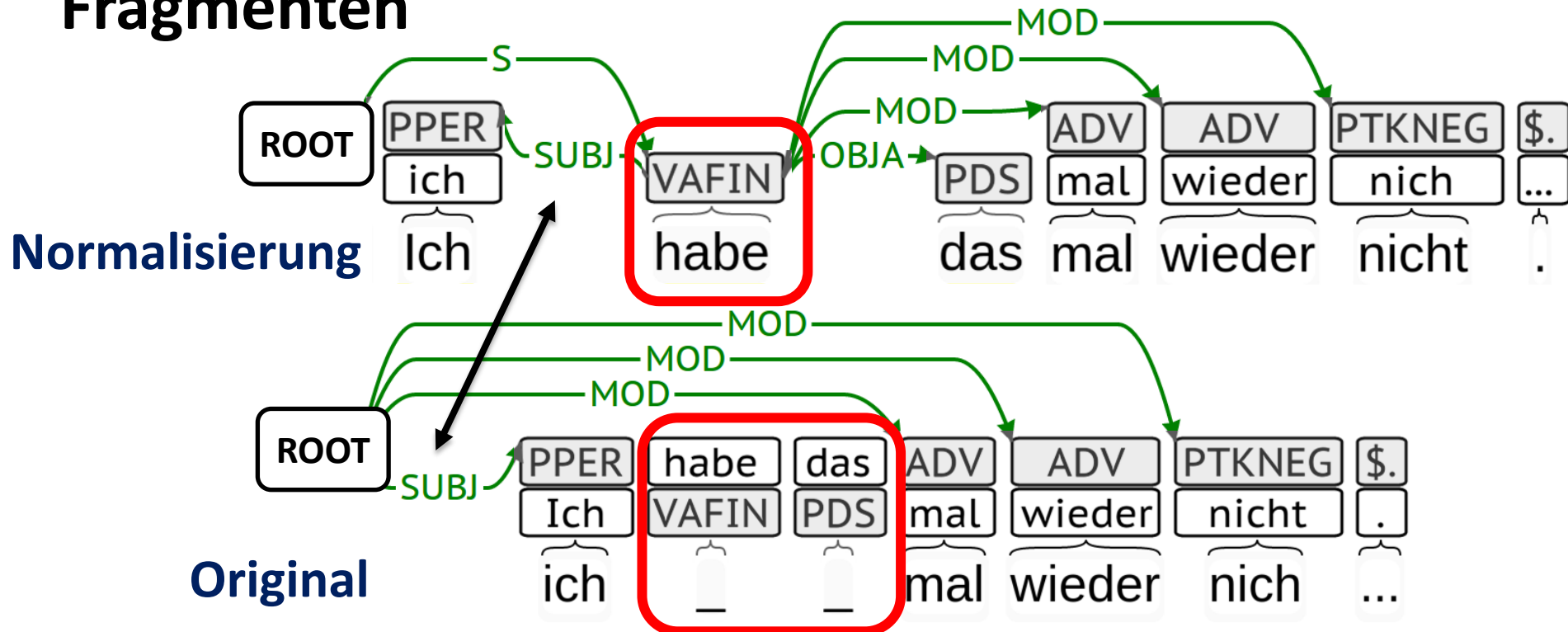
**Original**



# Normalisierung: Dependenz

Explizitmachung prototypischer Verben mit passender Argumentstruktur in verblosen Sätzen

→ **Motivation der syntaktischen Funktion von Fragmenten**



# Normalisierung: Koreferenz

**NORM hat Antezedens → ORIG hat gleichen Antezedens**

**NORM-Regel:** Ersetze in Kombinationen von „man“ und „er/sie“  
Letzteres durch Ersteres, wenn Letzteres im Kontext keinen  
Antezedenten besitzt!

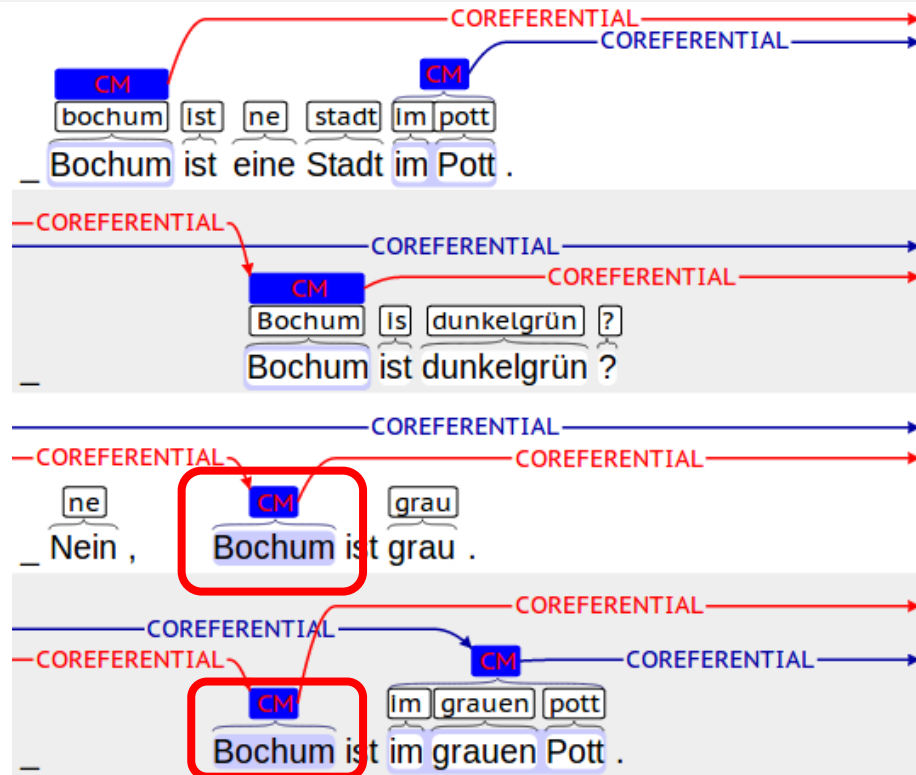
*Wenn **man** sich mit dieser Frage im Rahmen der Ethik beschäftigt, wird  
(**er/ man**) fast auf jeden Fall sagen dass Kriminalität sich nicht auszahlt.*





# Normalisierung: Koreferenz

516	Erdbeer\$	bochum ist ne stadt im pott	Bochum ist eine Stadt im Pott.
519	quaki	Bochum is dunkelgrün???	Bochum ist dunkelgrün?
524	Erdbeere\$	ne grau	Nein, <b>Bochum</b> ist grau.
528	Erdbeere\$	im grauen pott	<b>Bochum</b> ist im grauen Pott.





## B: Übersetzung von TiGer-Konstituenten in NoSta-D-Dependenzen

# Abhängigkeiten (Guidelines)

**Du** hast gegessen und geraucht

# Abhängigkeiten (Guidelines)

## TiGer Annotationsschema (Albert et al. 2003)

### Koordination

- 9.1 Grundstruktur der NP-, AP-, PP-Koordination . . . . .
  - 9.1.1 Koordinierende Konjunktionen . . . . .
  - 9.1.2 Binäre koordinierende Konjunktionen . . . . .
- 9.2 Koordination von satzeinleitenden Konjunktionen (CPs) . . . . .
- 9.3 Koordination von Nominal- und Präpositionalphrasen . . . . .
- 9.4 Koordinierte Adjektive . . . . . **Du**
- 9.5 Koordinierte Präpositionen . . . . .
- 9.6 Koordination von Verbalphrasen und Sätzen . . . . .

**hast gegessen und geraucht**



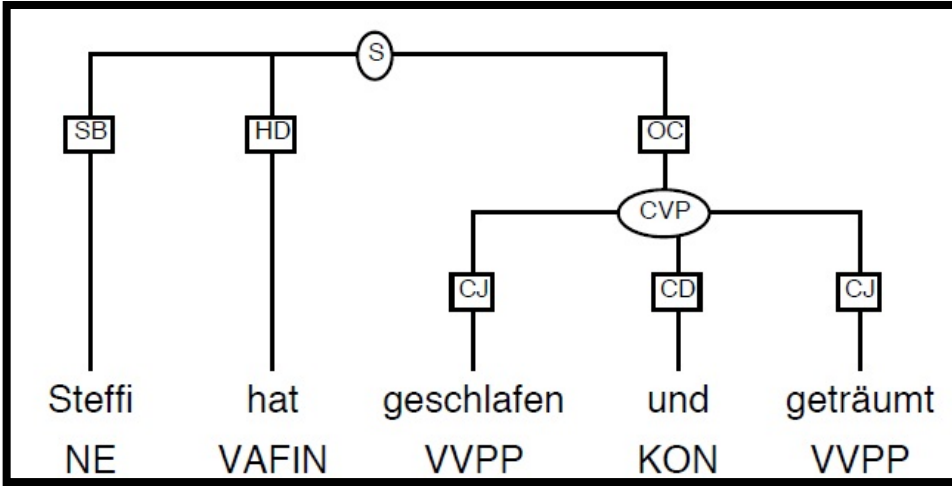
# Abhängigkeiten (Guidelines)

TiGer 2003

S.117, Bsp. 2

## Koordination von Verbalphrasen

NoSta-D 37



**Du hast geschlafen und geträumt**

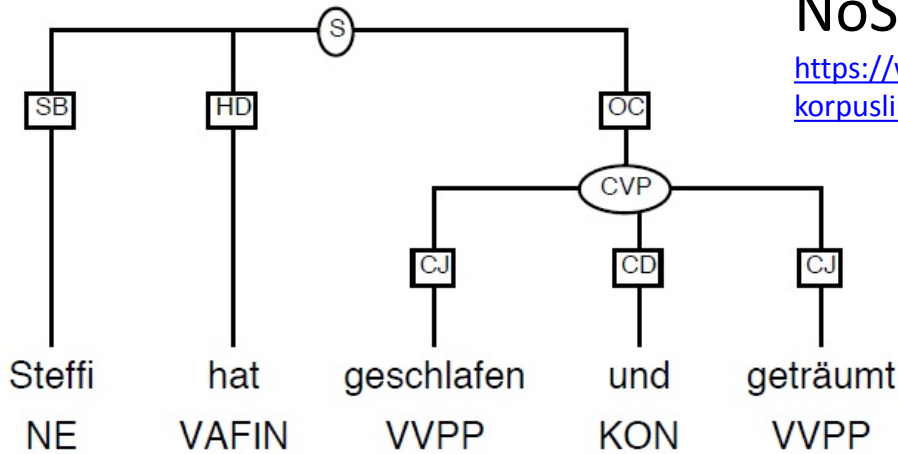
# Abhängigkeiten (Richtlinien)

TiGer 2003

S.117, Bsp. 2

## NoSta-D Richtlinien für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Du

hast gegessen **und** geraucht

Dependens

(C...)-[CD]

Was

# Abhängigkeiten (Guidelines)

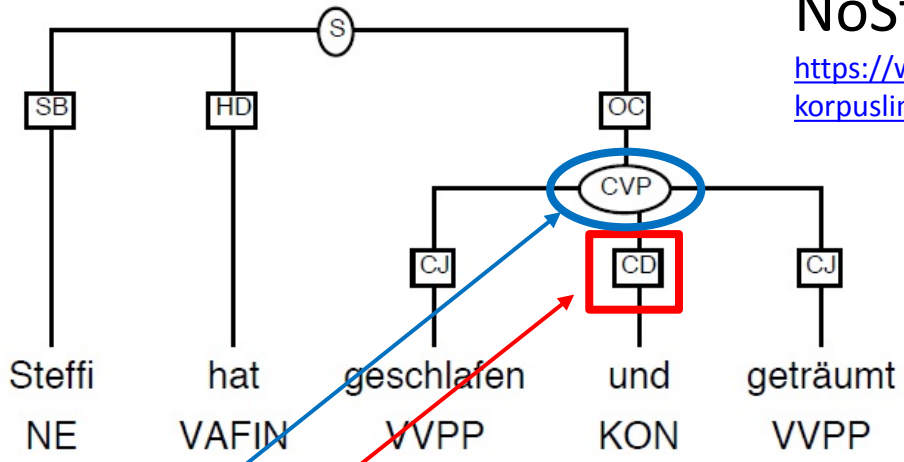
NoSta-D

TiGer 2003

S.117, Bsp. 2

## NoSta-D Guidelines für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Dependens
(C...)-[CD]
Was

# Abhängigkeiten (Guidelines)

TiGer 2003

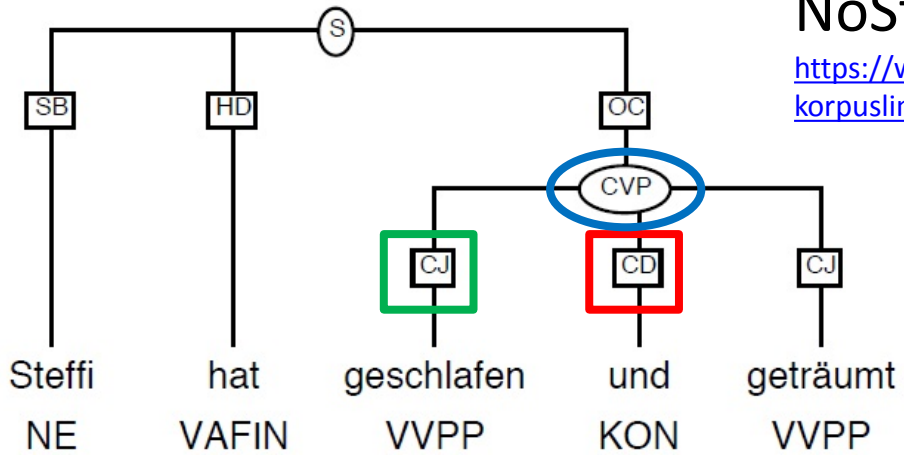
NoSta-D

40

S.117, Bsp. 2

## NoSta-D Guidelines für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Dependens	Regens
(C...)-[CD]	das innerhalb derselben (C...) unmittelbar vorangehende (C...)-[CJ], es sei denn ...
Was	ist Tochter von



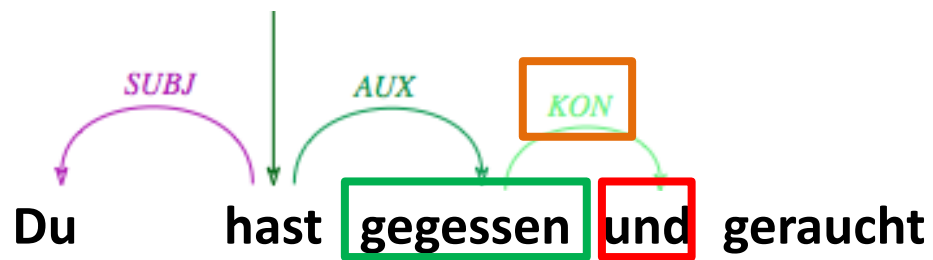
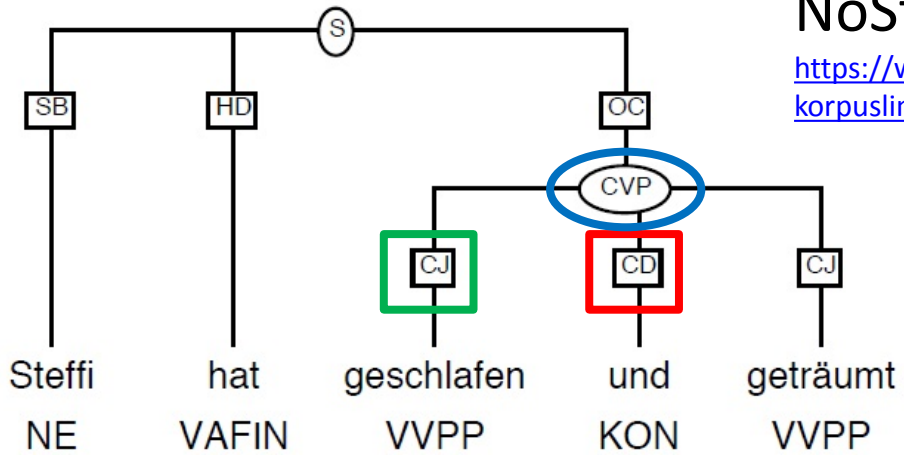
# Abhängigkeiten (Richtlinien)

TiGer 2003

S.117, Bsp. 2

## NoSta-D Richtlinien für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Abhängens	Regens	Label
(C...)-[CD]	das innerhalb derselben (C...) unmittelbar vorangehende (C...)-[CJ], es sei denn ...	KON
Was	ist Tochter von	wie

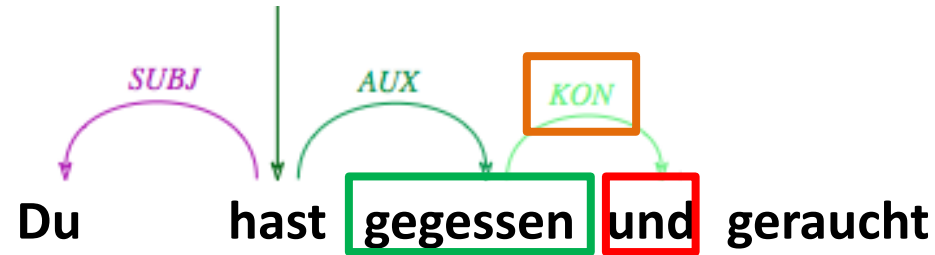
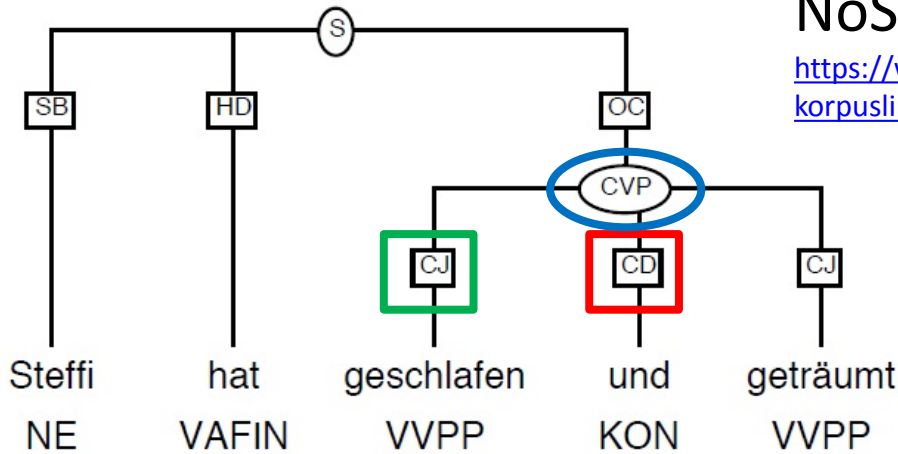
# Abhängigkeiten (Richtlinien)

TiGer 2003

S.117, Bsp. 2

## NoSta-D Richtlinien für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Abhängens	Regens	Label	Beispiel
(C...)-[CD]	das innerhalb derselben (C...) unmittelbar vorangehende (C...)-[CJ], es sei denn ...	KON	S.117, Bsp. 2
Was	ist Tochter von	wie	?

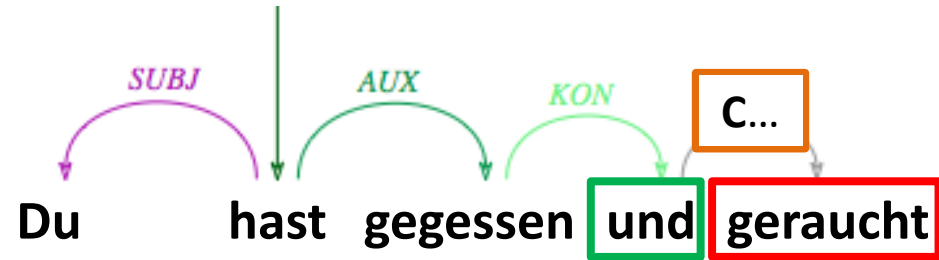
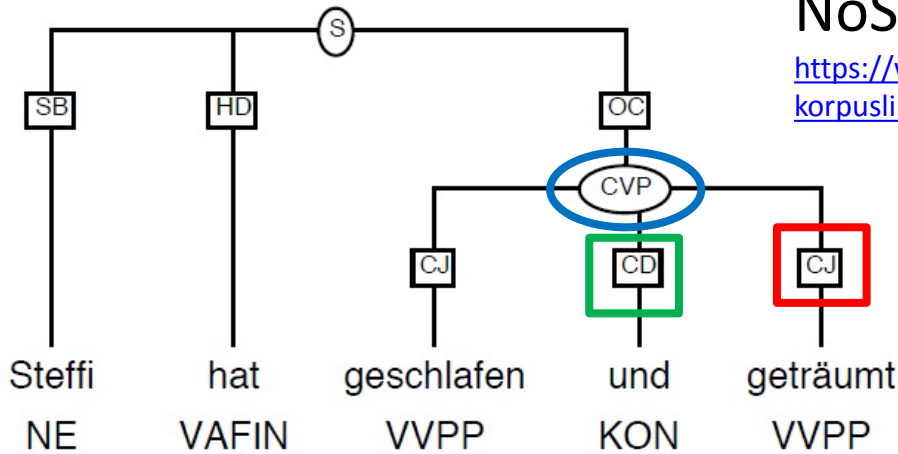
# Abhängigkeiten (Richtlinien)

TiGer 2003

S.117, Bsp. 2

## NoSta-D Richtlinien für Abhängigkeiten

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>



Abhängens	Regens	Label	Beispiel
(C...)-[CJ]	das innerhalb derselben (C...) unmittelbar links stehende (C...)-[C...], sofern ...	C...	S.117, Bsp. 2
Was	ist Tochter von	wie	

# Abhängigkeiten (Richtlinien)

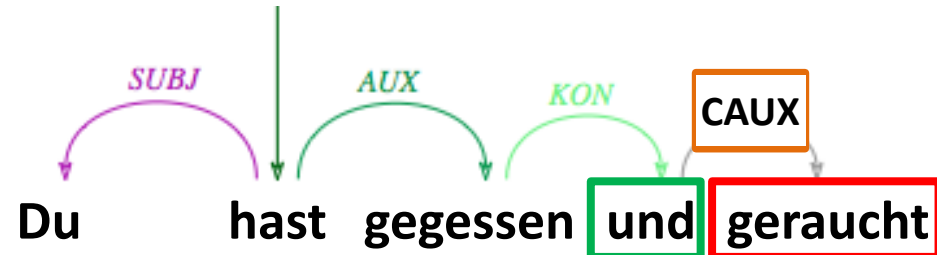
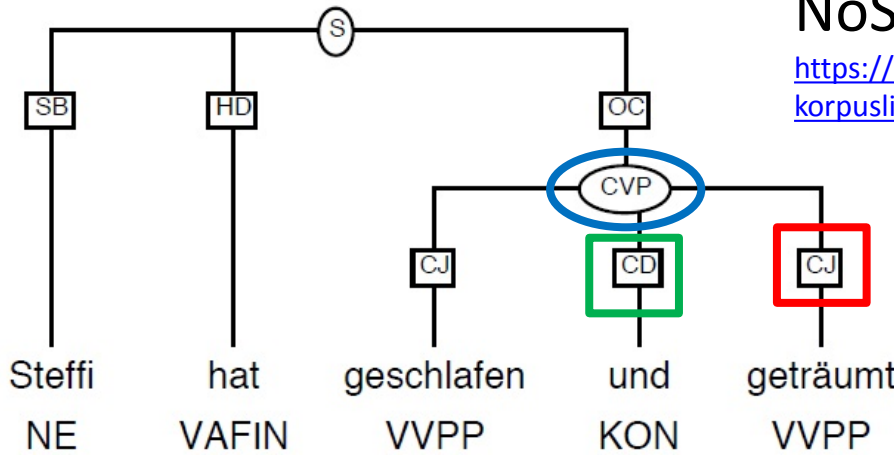
NoSta-D

TiGer 2003

S.117, Bsp. 2

## NoSta-D Richtlinien für Abhängigkeiten

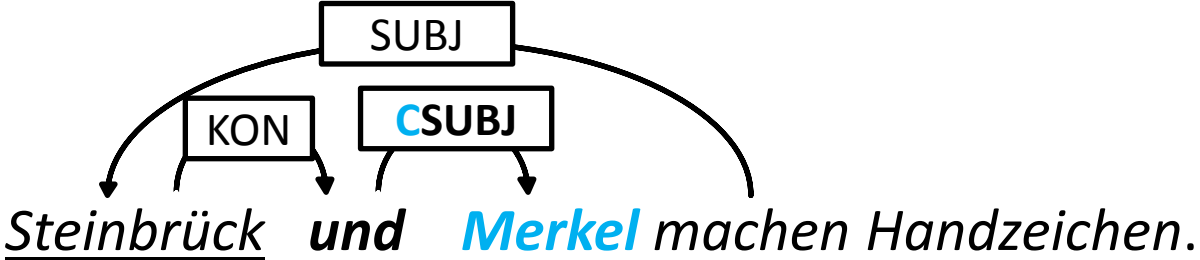
<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>

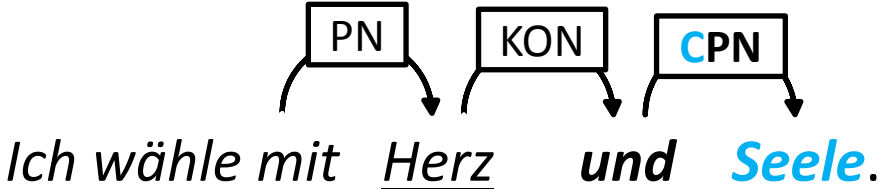


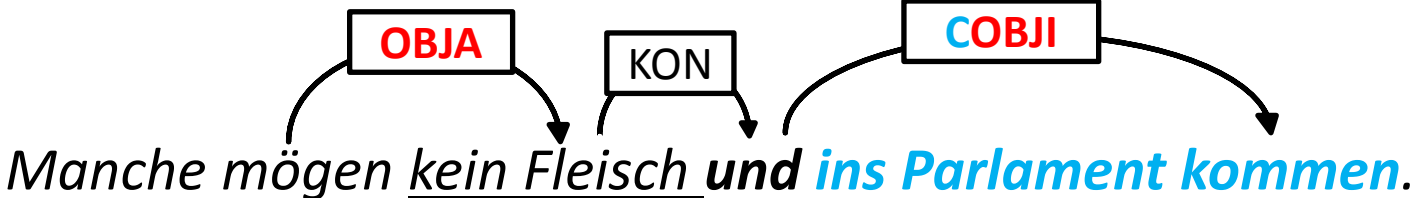
Dependens	Regens	Label	Beispiel
(C...)-[CJ]	das innerhalb derselben (C...) unmittelbar links stehende (C...)-[C...], sofern ... Das Restlabel ergibt sich aus der direkten Relation der Mutter des ersten koordinierten Elements zum Dependens.	C...	S.117, Bsp. 2
Was	ist Tochter von	wie	

## Kreuzklasse C

## Koordination

- CSUBJ**  


*Steinbrück und Merkel machen Handzeichen.*
- CPN**  


*Ich wähle mit Herz und Seele.*
- COBJI**  


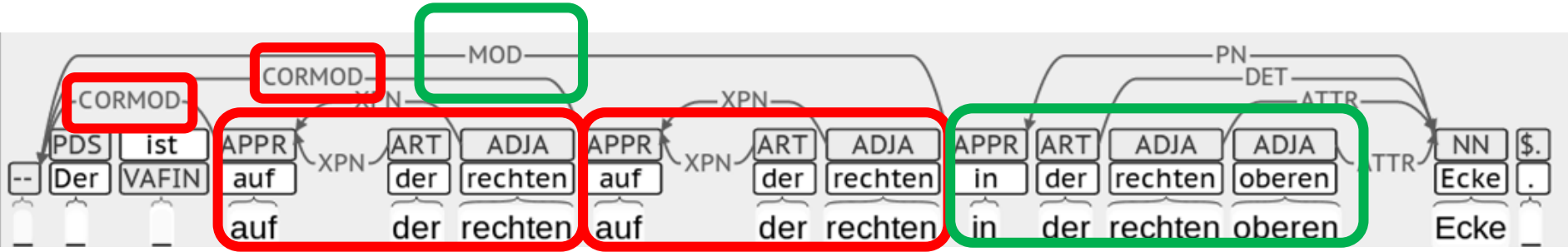
*Manche mögen kein Fleisch und ins Parlament kommen.*

# Abhängigkeiten (Guidelines)

## Kreuzklasse **COR...** (Korrektur)

Kreuzklasse über Label der Kante auf den höchsten Knoten einer Sequenz, die korrigiert wird.

- Kante auf höchsten Knoten = MOD → **CORMOD**
- [auf der rechten]<sub>**CORMOD**</sub> ... [in der rechten oberen Ecke]<sub>MOD</sub>



## C: Interessante Nichtstandard-Phänomene

# Named Entities (Chat)

## ■ Kreative Namensschreibung

#	Sprecher	Beitrag
25	system	<b>Lantonie</b> betritt den Raum.
26	<b>Lantonie</b>	:)
27	quaki	<b>lantonieeeeeee</b>
28	<b>Lantonie</b>	Hallo. :)
29	zora	<b>LANTOOO</b> :)))
34	marc30	<b>Lantöööö</b> :o)
35	TomcatMJ	hi <b>lanto</b>



# Named Entities (Chat)

## ■ Kreative Namensschreibung

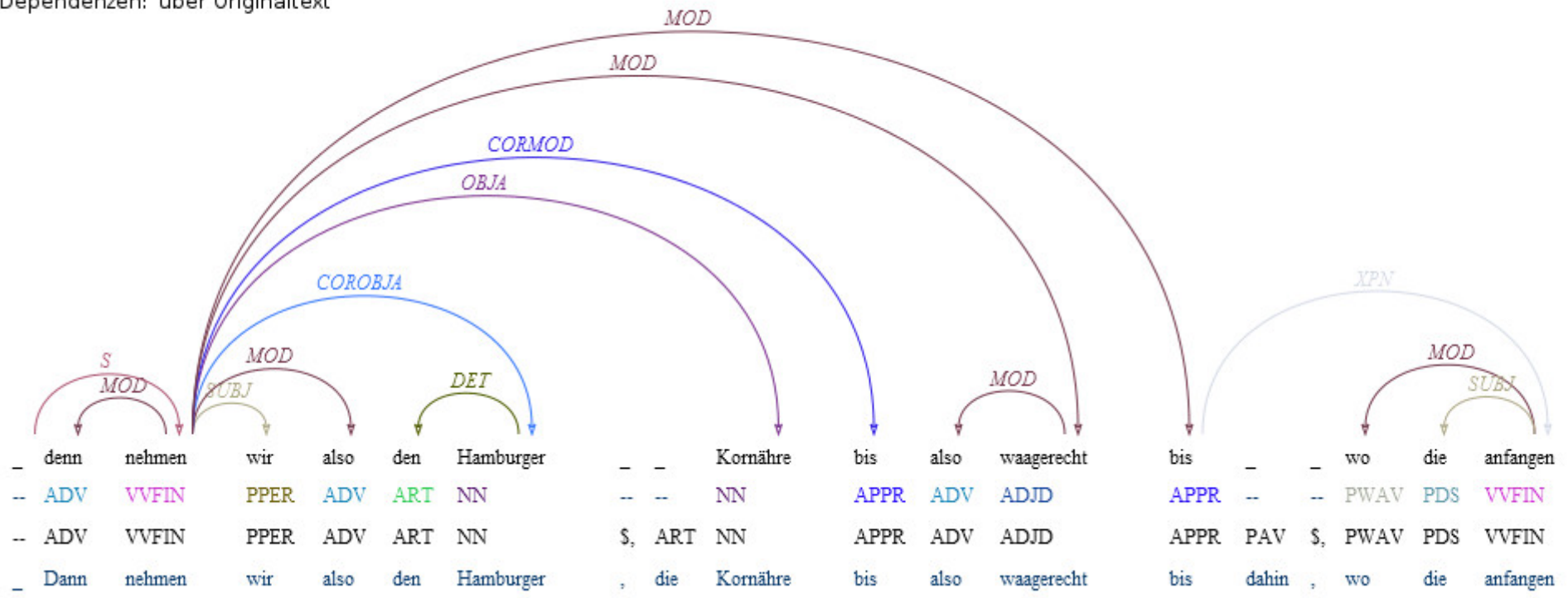
#	Sprecher	Beitrag
25	system	<b>Lantonie</b> betritt den Raum.
26	<b>Lantonie</b>	:)
27	quaki	<b>lantonieeeeeee</b>
28	<b>Lantonie</b>	Hallo. :)
29	zora	<b>LANTOOO</b> :)))
34	marc30	<b>Lantöööö</b> :o)
35	TomcatMJ	hi <b>lanto</b>

## ■ Kreative Namensgebung

#	Sprecher	Beitrag
434	system	<b>zurück</b> betritt den Raum.

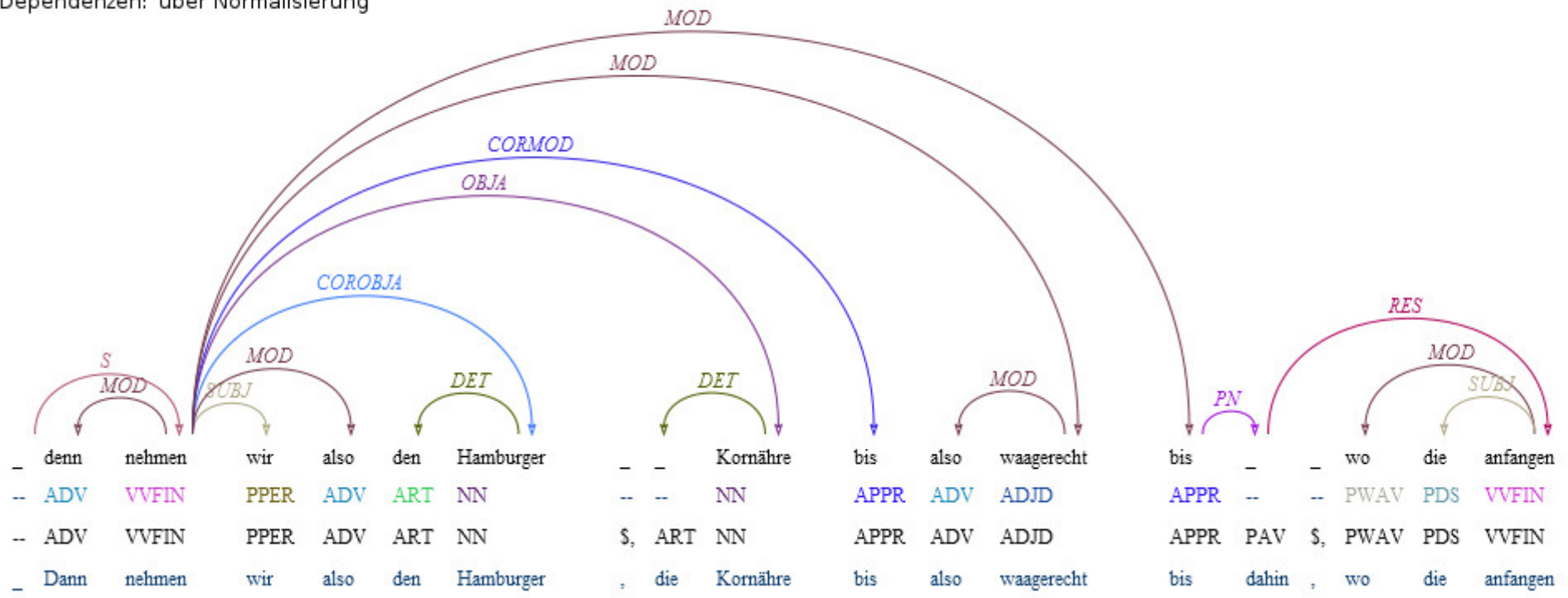
# Syntax: Korrekturen und Fragmente (auf ORIG)

Korpus: NoSta-D  
 Subkorpus: BeMaTaC\_2012-10-31-B  
 Segment: 58  
 Abhängigkeiten: über Originaltext



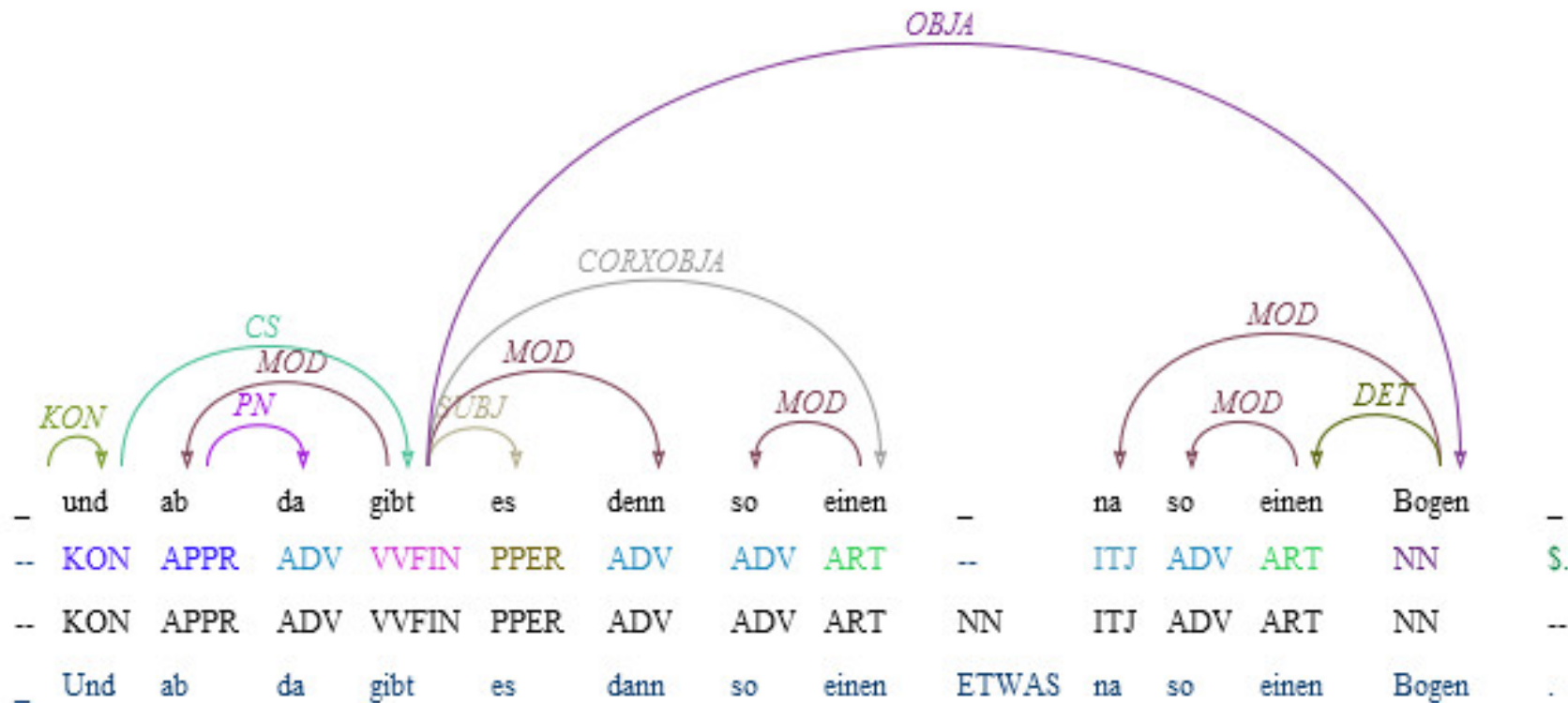
# Syntax: Korrekturen und Fragmente (auf NORM)

Korpus: NoSta-D  
 Subkorpus: BeMaTaC\_2012-10-31-B  
 Segment: 58  
 Abhängigkeiten: über Normalisierung



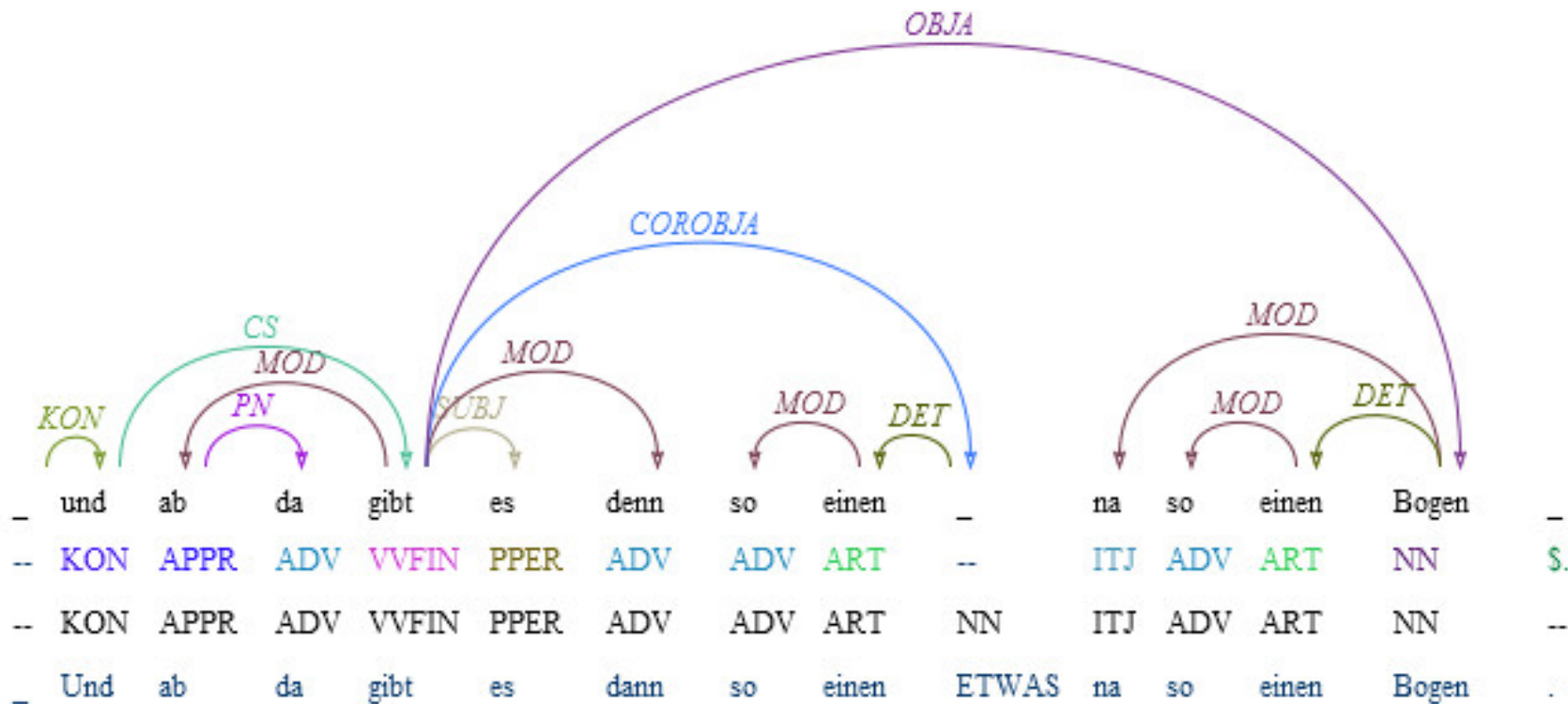
# Syntax: Korrekturen und Fragmente (auf ORIG)

Korpus: NoSta-D  
 Subkorpus: BeMaTaC\_2012-10-31-B  
 Segment: 60  
 Abhängigkeiten: über Originaltext  
 Bemerkung: aktuell gültige Annotation nach NoSta-D-AS



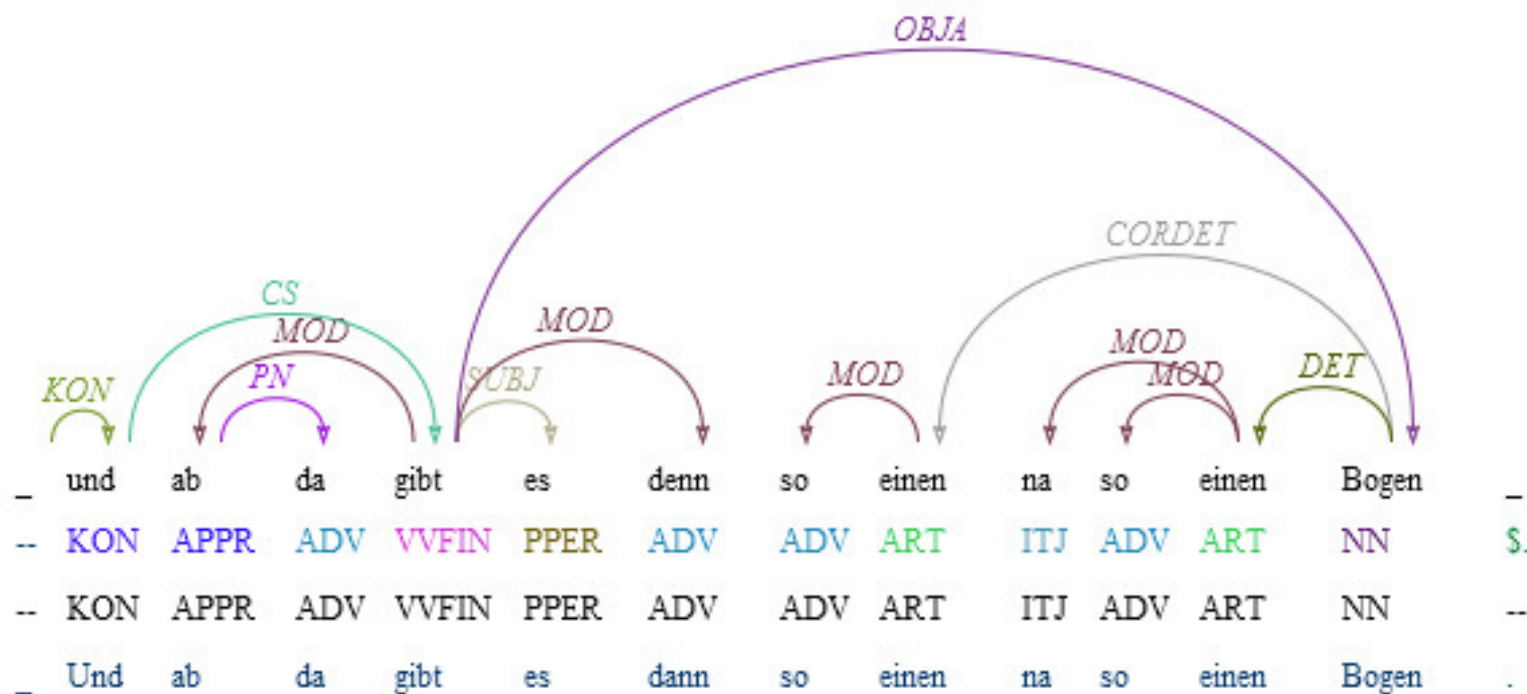
# Syntax: Korrekturen und Fragmente (auf NORM)

Korpus: NoSta-D  
 Subkorpus: BeMaTaC\_2012-10-31-B  
 Segment: 60  
 Abhängigkeiten: über Normalisierung  
 Bemerkung: aktuell gültige Annotation nach NoSta-D-AS



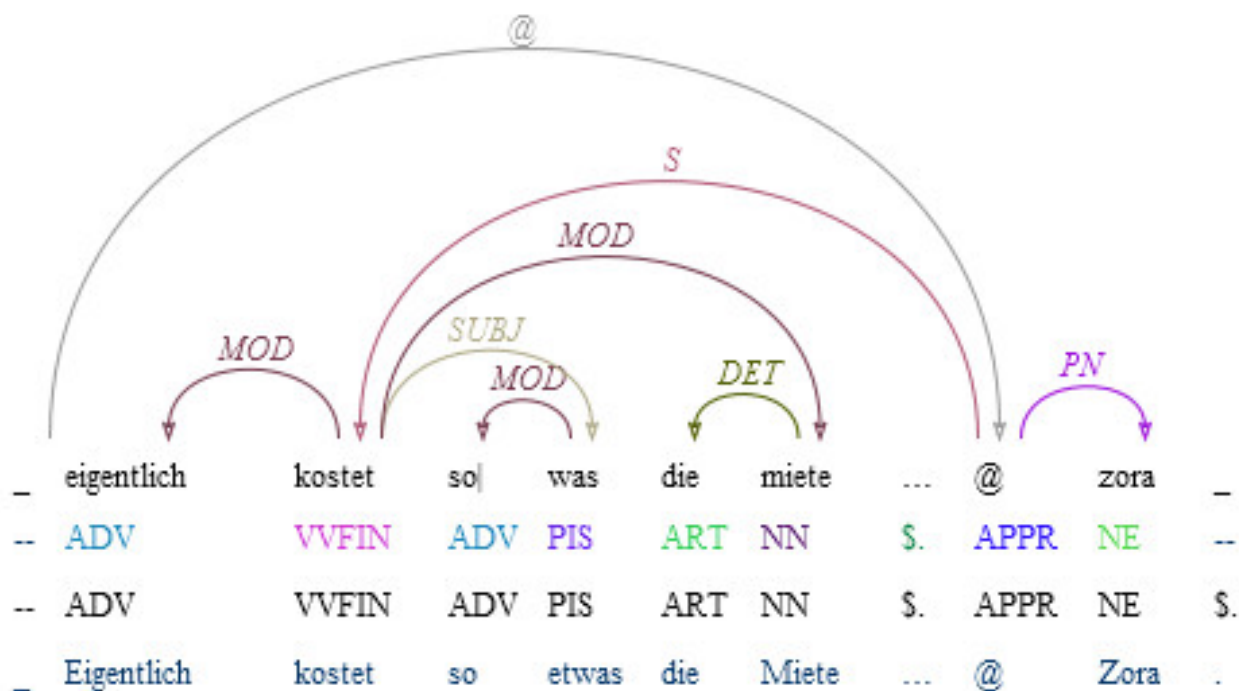
# Syntax: Korrekturen und Fragmente

Korpus: NoSta-D  
 Subkorpus: BeMaTaC\_2012-10-31-B  
 Segment: 60  
 Abhängigkeiten: über Originaltext (= über Normalisierung)  
 Bemerkung: im aktuellen NoSta-D-AS verworfene Annotation



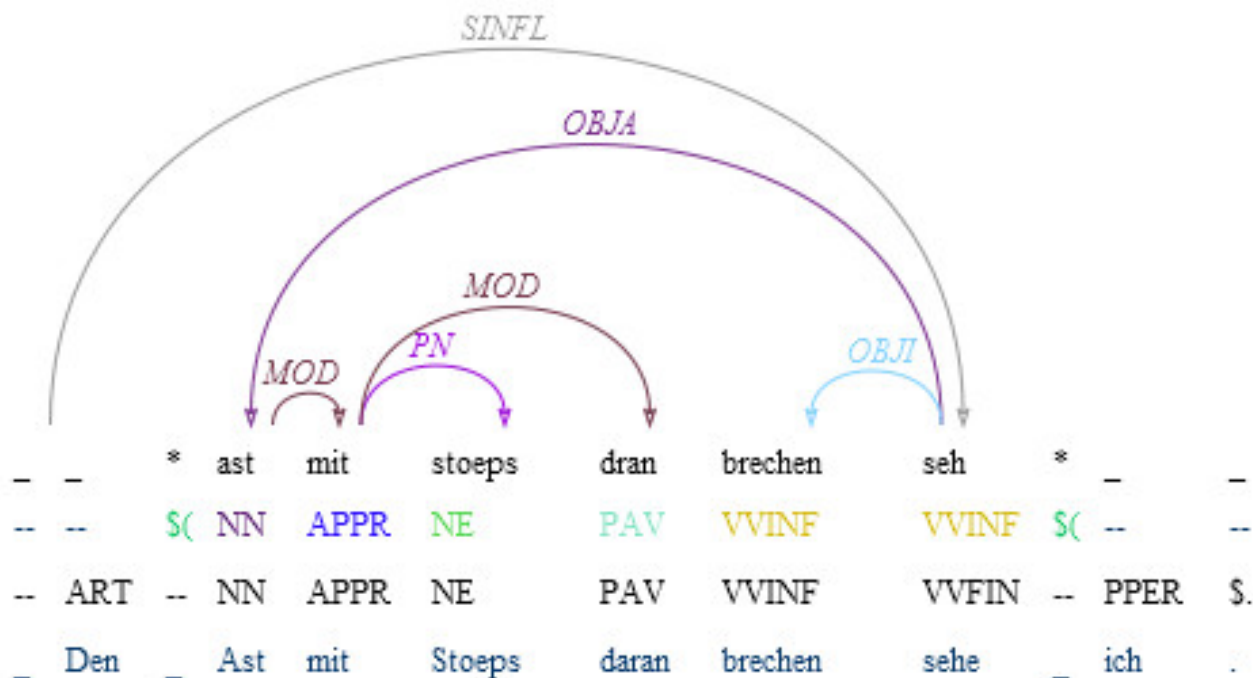
# Syntax: @-Adressierung

Korpus: NoSta-D  
 Subkorpus: unicum\_21-02-2003\_1  
 Segment: 221  
 Abhängigkeiten: über Originaltext (= über Normalisierung)



# Syntax: Inflektive (auf ORIG)

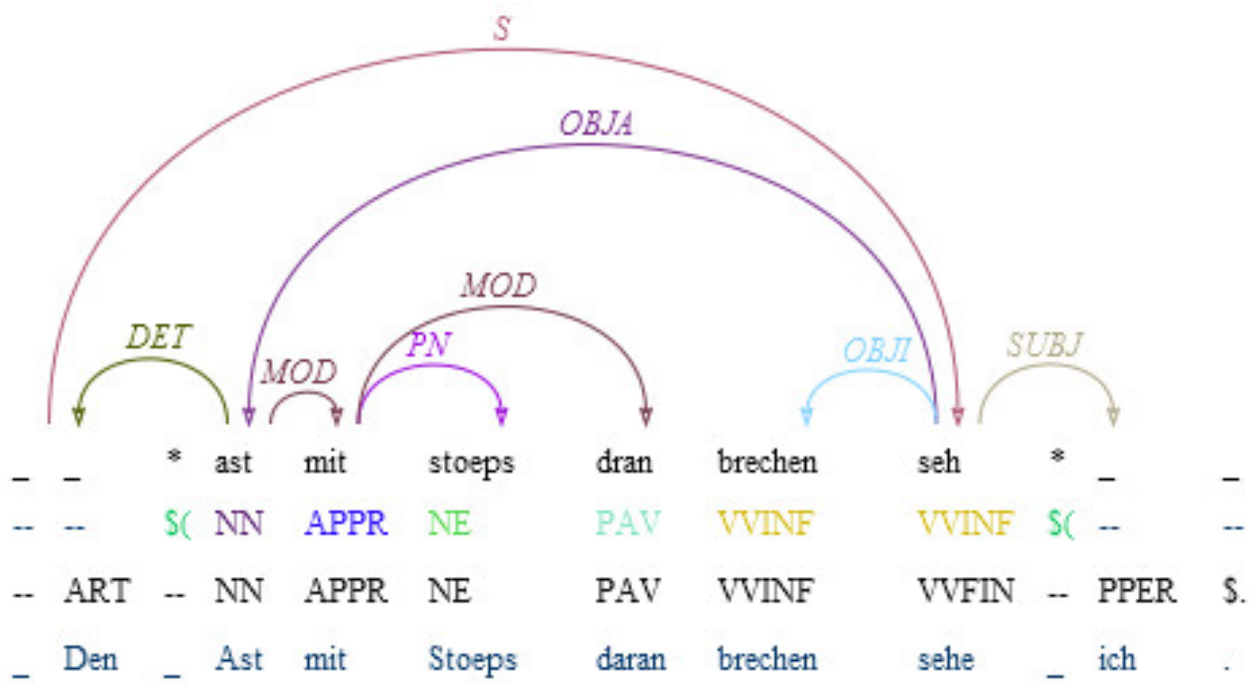
Korpus: NoSta-D  
 Subkorpus: unicum\_21-02-2003\_1  
 Segment-Nr.: 3  
 Abhängigkeiten: über Originaltext





# Syntax: Inflektive (auf NORM)

Korpus: NoSta-D  
 Subkorpus: unicum\_21-02-2003\_1  
 Segment-Nr.: 3  
 Abhängigkeiten: über Normalisierung



## ▪ Fehlerhafte Verwendung von Pronomen

Wenn **man** sich mit dieser Frage im Rahmen der Ethik beschäftigt, wird **er** fast auf jeden Fall sagen dass Kriminalität sich nicht auszahlt.

Falko\_fk002\_2006\_08

## ▪ Nicht-realisierte Referenten

516 Erdbeer\$ bochum ist ne stadt im pott

519 quaki Bochum is dunkelgrün???

524 Erdbeere\$ ne ∅ grau

528 Erdbeere\$ ∅ im grauen pott

534 Erdbeere\$ ∅ hell schwarz?

221006\_unicum\_21-02-2003\_1

- **NoSta-D → Deutsche Pilot-Ressource für Training und Entwicklung für Nichtstandard-Varietät verfügbar**
- **Ergänzungen zu gängigen Richtlinien (NER, Abhängigkeiten und Korreferenzen) → gute Abdeckung von NoSta-D**
- **Normalisierung explizit in die Korpora integrieren → Nachvollziehbarkeit strittiger Annotationen erhöhen**
- **Parallele Annotation von Normalisierung und abgeleitet daraus der Originaldaten → Vereinbarkeit komplementärer Untersuchungsansätze**

- **Aufbau größerer Ressourcen und weiterer Varietäten für Training**
- **Untersuchung zusätzlicher Annotationstypen**
- **Grundsätzliche Kritik an der linguistischen Adäquatheit dependenzgrammatischer Modelle auch für den deutschen Standard**
  - u. a. Koordination finiter Verben
- **Generell stärkerer Fokus auf den Einsatz von Mehrebenenarchitekturen in Annotations-, Such- und Analysetools**
  - u. a. gleichzeitige Darstellung und Verarbeitung konkurrierender Normalisierungsebenen

# Danke!

Projektseite:

<https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/clarin-d>

Kontakt:

[burkhard.dietterle.1@hu-berlin.de](mailto:burkhard.dietterle.1@hu-berlin.de)  
[marc.reznicek@hu-berlin.de](mailto:marc.reznicek@hu-berlin.de)

- Albert, Stefanie; Anderssen, Jan; Bader, Regine; Becker, Stefanie; Bracht, Tobias; Brants, Thorsten et al. (2003):** TIGER Annotationsschema.
- Dipper, Stefanie; Lüdeling, Anke; Reznicek, Marc (erscheint):** *NoSta-D. A Corpus of German Non-Standard Varieties*. In: Zampieri, Marcos; Diwersy, Sascha (Hgg.): *Non-Standard Data Sources in Corpus-Based Research* (ZMS-Studien - Schriften des Zentrums Sprachenvielfalt und Mehrsprachigkeit der Universität zu Köln 5), Hamburg: Shaker.
- Foth, Kilian A. (2006):** Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Technischer Report. Universität Hamburg. Hamburg.
- Kaupat, David; Warzecha, Saskia; Stede, Manfred (2013):** Koreferenz. Chapter 5. Erweiterung des PoCoS-Kernschemas.
- Hirschmann, Hagen; Doolittle, Seanna ; Lüdeling, Anke (2007):** Syntactic annotation of non-canonical linguistic structures. In: *Proceedings of Corpus Linguistics 2007*, Birmingham, <http://linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiterinnen/anke/pdf/HirschmannDoolittleLuedelingCL2007.pdf>
- Naumann, Karin (2007):** Manual for the Annotation of In-Document Referential Relations. Seminar für Sprachwissenschaft, Abt. Computerlinguistik Universität Tübingen, [http://www.sfb441.uni-tuebingen.de/a1/Publikationen/tuebadz\\_relations\\_man.pdf](http://www.sfb441.uni-tuebingen.de/a1/Publikationen/tuebadz_relations_man.pdf)
- Schiller, Anne; Teufel, Simone; Stöckert, Christine; Thielen, Christine (1999):** Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical Report. University of Stuttgart; University of Tübingen, <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.
- Telljohann, Heike; Hinrichs, Erhard W.; Kübler, Sandra; Zinsmeister, Heike; Beck (2012):** Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Seminar für Sprachwissenschaft. Universität Tübingen, <http://www.sfs.uni-tuebingen.de/resources/tuebadz-stylebook-1201.pdf>.