

# Syntax natürlicher Sprachen

## 8: Grammatische Merkmale und Merkmalstrukturen

A. Wisiorek

Centrum für Informations- und Sprachverarbeitung,  
Ludwig-Maximilians-Universität München

02.12.2025

# 1. Motivation für Feature-Modellierung

## 1 Motivation für Feature-Modellierung

## 2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

## 3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

## 4 Typhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

# Motivation für *Feature-based Grammars*

## Übergenerierung durch CFGs

- aufgrund Nichtberücksichtigung morphosyntaktischer Constraints

## 3 zentrale morphosyntaktische Constraints:

- ① Kasusrektion
- ② Kongruenz (Agreement)
- ③ Subkategorisierung (Art und Anzahl von Argumenten)

## Feature-based grammars (FCFGs)

- **Modellierung** von grammatischen Merkmalen über **Merkmalstrukturen**
- morphosyntaktische **Constraintregeln** als Anweisung auf Durchführung von **Unifikation** der Merkmalstrukturen  
→ *Unifikation: Vereinbarkeit von Merkmalstrukturen*

# Gründe für Übergenerierung von CFGs (auch X-Bar)

## Nichtberücksichtigung von Morphologie

- Rektion (Kasus):

*\*der Mann sieht des Kindes*

- Kongruenz (Agreement in Merkmalen):

*\*das Kinder*

## Nichtberücksichtigung von Subkategorisierung

- Art und/oder Anzahl von Komplementen

*\*der Hund geht die Katze*

# Lösungen für Übergenerierung

## Splitting atomarer Kategorien

- z.B. Numerus-Kongruenz NP:
- NP gesplittet in SgNP und PlNP:
  - SgN, PlN, SgDET, PlDET, SgNP, PlNP
  - SgNP → SgDET SgN, PlNP → PlDET PlN
- Problem: Regelvervielfachung

## Merkmale in Lexikon

- Merkmalstrukturen und Unifikationsconstraints erlauben Feststellung Merkmalskongruenz

## Auswahl durch probabilistisches Modell (PCFG)

- Übergenerierung erlauben
- ungrammatische Sätze als unwahrscheinliche ausschließen

# Einsatzbereiche von Unifikation

- Logik-Programmierung (z. B. Prolog)
- Pattern Matching (z. B. in funktionalen Programmiersprachen)
- Typinferenz (vor allem funktionale Programmiersprachen wie Haskell, Scala, etc. Eingeschränkt aber z. B. auch C#)
- Verarbeitung von **Merkmalstrukturen**
  - dienen der Beschreibung komplexer Objekte, z. B. grammatischer Merkmale
  - Unifikation zur Überprüfung der Vereinbarkeit von Merkmalstrukturen

## 2. Grammatische Merkmale

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typenhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

## 2.1. Morphosyntaktische Flexionskategorien des Deutschen

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- **Morphosyntaktische Flexionskategorien des Deutschen**

- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typenhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

# Nominale Flexionskategorien

## CASE (Kasus)

- nom
- gen
- dat
- akk

## GEN (Genus)

- fem
- mask
- neutr

## NUM (Numerus)

- sg
- pl

## AGR (Agreement)\*

- CASE
- GEN
- NUM

## PERSON

- 1
- 2
- 3

\* Nominales Agreement (N mit DET, ADJ) modelliert als komplexes Merkmal, d.h. mit anderen Merkmalen als Wertausprägungen

## AGR (Agreement)\*

- PERSON: 1/2/3
- NUM: sg/pl

\* Verbales Agreement (Verb mit Subjekt) modelliert als komplexes Merkmal, d.h. mit anderen Merkmalen als Wertausprägungen

## 2.2. Kodierung syntaktischer Funktionen

- 1 Motivation für Feature-Modellierung
- 2 Grammatische Merkmale
  - Morphosyntaktische Flexionskategorien des Deutschen
  - **Kodierung syntaktischer Funktionen**
  - Flexionskategorien
  - Kasus und Agreement
- 3 Merkmalstrukturen
  - Grundlagen
  - Constraintregeln und Unifikation
  - Unifikation und Subsumption
- 4 Typenhierarchie und getypte Merkmalstrukturen
  - Grundlagen
  - Subsumption und Unifikation getypter Merkmalsstrukturen
  - Bedingungen

# Typen syntaktischer Kodierung

→ Kodierung syntaktischer Funktionen, insbesondere zentraler Grammatische Relationen (Satzgliedfunktionen: Subjekt, Objekt usw.)

## A: Strukturelle Kodierung

- ① **Wortstellung:** SVO vs OVS etc.

## B: Morphologische Kodierung (*Morphosyntax / Flexion*)

- ② **Kasus-Markierung**
- ③ **Agreement-Markierung**  
→ d. h. über *grammatische Kategorien/Merkmale*

- Feature-Tagset:  
<http://universaldependencies.org/u/feat/index.html>

- **Sprachtypologie** = auf grammatische Struktur und die Varianz ihrer Kodierung bezogener Sprachvergleich

## Sprachbau-Typologie

- ① **Isolierender Sprachbau**: die syntaktischen Relationen werden primär durch Wortstellung kodiert (z. B. Vietnamesisch, Kantonesisch)
- ② **Analytischer Sprachbau**: Kodierung primär durch freie Morpheme = Funktionswörter (z. B. Deutsch)
- ③ **Synthetischer Sprachbau**: Kodierung primär durch gebundene Morpheme (z. B. Latein)

# Typisierung Sprachen mit morphologischer Kodierung

## Synthetischer vs. Analytischer Sprachbau

Differenzierung nach dem Typ der Morpheme (gebunden vs. frei)

## Agglutinierender vs. Flektierender Sprachbau

Subdifferenzierung synthetischer Sprachen nach dem **Fusionsgrad der Morpheme / Form-Funktions-Verhältnis**

## *dependent-marking* vs. *head-marking*

Differenzierung nach der Verwendung von **Kasus und Agreement**

## Akkusativ- vs. Ergativ- vs. Aktiv-System

Differenzierung nach der **Abbildung von semantischen Rollen auf Grammatische Relationen**

## *topic- vs. subject-prominent*

Differenzierung nach der **Abbildung von pragmatischen Rollen auf Grammatische Relationen**

## 2.3. Flexionskategorien

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- **Flexionskategorien**
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typ hierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

- **syntaktisch relevanter Teil der Morphologie (Morphosyntax)**
- **Kodierung syntaktischer Funktionen** zwischen den Wörtern im Satz durch **Formveränderung**  
→ *schließt insbesondere auch das konkatenative Hinzufügen von Morphemen oder Funktionswörtern ein*
- **substantielle Kodierung der syntaktischen Funktion**  
→ *durch Funktionsmarker, z. B. Akkusativ als Objektmarker*
- an Stelle von struktureller Kodierung über lineare Anordnung  
→ *Wortstellung, z. B. Subjekt vor Objekt*

# Flexionskategorien als Merkmal-Wert-Paare

- Flexionskategorie = Grammatisches Merkmal  
→ *Merkmal hat Merkmalsausprägungen = Werte*
- Beispiel: **Flexionskategorie Numerus** hat die **Werte Singular und Plural**:
  - NUM: sg, pl
- **Merkmalsausprägungen** werden durch **Morpheme** kodiert
  - *Morphem = kleinste bedeutungstragende Einheit der Sprache*
  - *nicht weiter segmentierbare substantielle Form-Funktions-Paare*
- Beispiel: Pluralmorphem im Englischen:
  - -s kodiert pl
  - Ø kodiert sg (Nullmorphem)

- **Affigierung:** Suffixe (Endungen), Präfixe, Infixe: *sag-t-e*
  - *konkatenative Morphologie*
  - *agglutinierend* bzw. *flektierend* (s.u.)
- **Funktionswörter ('freie Morpheme'):** *war gegangen*
  - *analytischer Sprachbau*
- **Ablaut** (Stammveränderung durch Vokalwechsel: *ich hänge* > *ich hing*)
- **Reduplikation:** lat. *pe-pend-i* 'ich hing'
- **Deutsch = gemischt analytisch-flektierend:** Verwendung von flektierten **Hilfswörtern** (Auxiliare, Funktionswörter)

# Unterscheidung nach Form-Funktionsverhältnis

- ① **1:1 = eine Form** (ein Morphem) **kodiert eine Funktion:**  
*ich sag-t-e*: say-PRT-1SG (t-Präteritum der schwachen Verben)  
→ **agglutinierend**
- ② **1:n = eine Form kodiert n Funktionen:**  
*ich sag-e*: say-1+SG  
→ **flektierend = Verschmelzung von Funktionen in einem Morphem**
- ③ **n:1 = Allomorphie: eine Funktion wird durch unterschiedliche Morpheme realisiert:**  
PL: *Kind-er; Tier-e; Essen-Ø*

- **(Un-)Markiertheit:** Form (Merkmalsausprägung), die die **default-Funktion** des Merkmals anzeigt, ist üblicherweise **substantiell minimal**, oft Fehlen einer substantiellen Form
  - Ansatz **Nullform** ( $\emptyset$ )
  - z. B. **Nominativ im Deutschen:**  
*Hund- $\emptyset$ :* dog-NOM  
*Hund-es:* dog-GEN

Deklination = nominale Flexion (Nomen, Adjektiv, Pronomen, Determinierer)

### *nominale Flexionskategorien des Deutschen:*

**Kasus**; Genus, Numerus, Person, Definitheit

#### **Kasus:** Nominativ / Akkusativ / Dativ / Genitiv

- in anderen Sprachen: **geringere Anzahl an Kasus** (Arabisch: 3; Berber: 2) oder **höhere** (Finnisch: 15) oder **kein morphologischer Kasus** (Kodierung durch Wortstellung oder Agreement)
  - ① Markierung **Grammatischer Relationen** im Satz (Subjekt, Objekt, Adverbial)
  - ② Markierung der **Modifikationsbeziehung innerhalb von NPs** (Attributfunktion, z. B. Genitiv-Attribut)

## Genus: Maskulin / Feminin / Neutrum

- **inhärente** Kategorisierung (nicht veränderbares Merkmal; semantisch nur noch zum Teil transparent)  
→ *in vielen Sprachen: Klassenmarker (chinesisch, Bantu-Sprachen): bezeichnen z. B. die Form von Dingen*

## Numerus: Singular / Plural

- Kategorisierung nach **Einheit/Vielheit**  
→ *zusätzlich häufig Dual = Zweiheit, z. B. im Arabischen*

## Person: 1 / 2 / 3

- Subkategorisierung beim Pronomen bzgl. der **Teilnehmer im Äußerungskontext**: Referenz auf Sprecher oder Adressat
- Substantive sind immer 3. Person

## Definitheit: Definit / Indefinit

- Kategorisierung bzgl. **Bekanntheit**

# Agreement in der Nominalphrase

- **Merkmalskongruenz zwischen Nomen** (als Kopf der Phrase) und den Dependenten **Determinativ und Adjektiv** in **Genus, Numerus und Kasus**
- Anzeige der **Dependenz nominaler Modifikatoren** durch **Kongruenz in Merkmalen mit dem nominalen Kopf**
- Im Deutschen trägt häufig **nur noch der Artikel bzw. das Adjektiv die Kasus-Merkmale**, da das Kasussystem im Deutschen stark abgebaut ist
- **Adjektiv-Kongruenz:** Merkmalskongruenz mit dem Nomen in Genus, Numerus und Kasus, aber **unterschiedlich je nach Vorhandensein des Artikels** (starke vs. schwache Formen)

# Konjugation = verbale Flexion

## *verbale Flexionskategorien des Deutschen:*

Person, Numerus (**Agreement**); Tempus, Modus, Genus verbi

### **Person+Numerus-Kongruenz:** 1sg/2sg/3sg/1pl/2pl/3pl

- Kongruenz/Agreement in Person und Numerus mit dem Subjekt

### **Tempus:** Präs. / Prät. / Perf. / Plsqperf. / FuturI/II

- Kategorisierung bzgl. des **Zeitpunkts des Geschehens relativ zum Moment der Aussage** (Vergangenheit, Gegenwart, Zukunft)

### **Modus:** Indikativ / Imperativ / Konjunktiv

- Kategorisierung bzgl. **Einstellung des Sprechers zur Aussage**

### **Genus verbi:** Aktiv / Passiv

- auch Voice/Diathese: Kategorisierung der **Abbildung von semantischen Rollen auf die Grammatischen Relationen**

## 2.4. Kasus und Agreement

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- **Kasus und Agreement**

3 Merkmalstrukturen

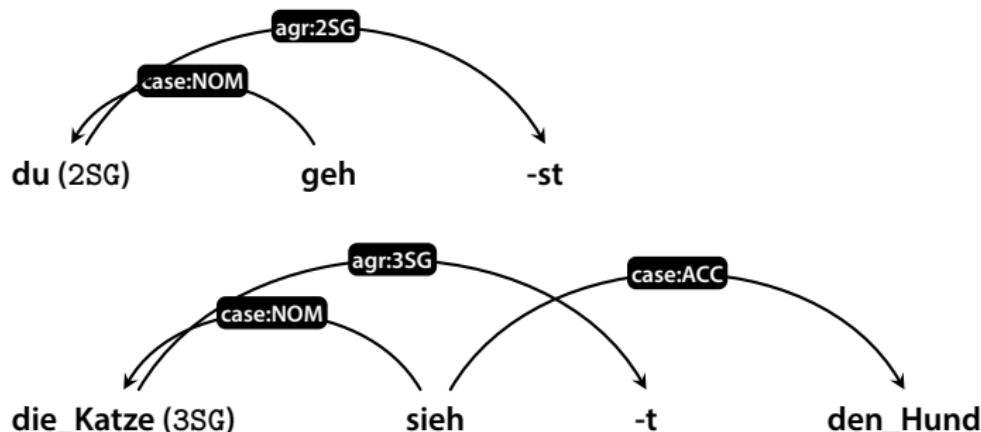
- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typ hierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

# Markierung zentraler syntaktischer Satzgliedfunktionen

- ① **Kasus:** *Markierung der syntaktischen Funktion eines verbalen Dependenten am Dependenten* (*dependent-marking*)
- ② **Agreement:** *Markierung der syntaktischen Funktion eines verbalen Dependenten am Kopf (Verb) über Kongruenz in Merkmalen mit dem Dependenten* (*head-marking*)



- **Markierung Grammatischer Relationen** durch **grammatisches Merkmal am Dependenten**
- **Varianz** der Werte des Kasusmerkmals **in Abhängigkeit von der zu kodierenden syntaktischen Funktion**, also vom syntaktischen Kontext (abhängiges Merkmal)
- **Typ1: Rektion:** Markierung Nomen entsprechend der Verbvalenz (Komplement) = **Kernkasus**
- **Typ2: Modifikation:** Markierung Nomen als Modifikator des Verbs (Adjunkt) = **obliquer Kasus**

- **Form von Kasus:** neben **morphologischem Kasus** (also mit Affix, meist Suffix, als Kasusmarker) auch durch **Adposition** (z. B. im Japanischen durch Postpositionen) oder durch **Kasusmarkierung am Artikel** (vgl. Deutsch)
- im Deutschen typischerweise:
  - **Nominativ** als **Subjektkasus**
  - **Akkusativ** als **Objektkasus** (selten auch Genitiv/Dativ/Präpos.)
  - **Dativ** als Kasus des **indirekten Objekts**
  - **Präpositionen und z.T. auch Genitiv und Akkusativ** als **Adverbialkasus**

# Agreement / Merkmalskongruenz

- Markierung Grammatischer Relationen durch **Übereinstimmung des Kopfes** in grammatischen Merkmalen **mit Merkmalen des Dependenten**
- **Kovarianz morphologischer Eigenschaften des Verbs mit Eigenschaften der Subjekt-NP**
- im Deutschen: **Kongruenz des Verbs mit Subjekt** in den Merkmalen **Person und Numerus**

Subjekt-Merkmale		verbale Merkmale
Person	$\leftarrow AGR \rightarrow$	Person
Numerus	$\leftarrow AGR \rightarrow$	Numerus
Genus		Tempus
		Modus
Case	$\leftarrow$	

# Mono- vs. Polypersonales Agreement

- im Sprachvergleich: auch **Kodierung der syntaktischen Funktion weiterer Kernargumente** gegeben (*double-agreement* usw.)  
→ *entsprechend der GR-Hierarchie: Subjekt > Objekt > Ind. Objekt*

## Baskisch: Agreement mit Subjekt, Objekt und Indirektem Objekt

Oparitu d-i-a-t

give 3SG:P-have-2SG:IO-1SG:A\*

I have given it to you (as a present).

\*P = Patiensargument, A = Agensargument

- als *head-marking*-Strategie ermöglicht Agreement **Pro-Drop = pronominale Nicht-Besetzung von valenzgeforderten Stellen**
- verbale Agreement-Marker sind meist (bzw. sind Ergebnis der Grammatikalisierung von) **enklitische Personalpronomen**

### 3. Merkmalstrukturen

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typenhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

# 3.1. Grundlagen

- 1 Motivation für Feature-Modellierung
- 2 Grammatische Merkmale
  - Morphosyntaktische Flexionskategorien des Deutschen
  - Kodierung syntaktischer Funktionen
  - Flexionskategorien
  - Kasus und Agreement
- 3 Merkmalstrukturen
  - **Grundlagen**
  - Constraintregeln und Unifikation
  - Unifikation und Subsumption
- 4 Typenhierarchie und getypte Merkmalstrukturen
  - Grundlagen
  - Subsumption und Unifikation getypter Merkmalsstrukturen
  - Bedingungen

- auch: **Feature Structure** (FeatStruct) oder **Attribut-Wert-Matrix** (AVM)
- **formale Repräsentation komplexer Objekte**, die durch eine **Anzahl an Eigenschaften definiert** sind (Feature + Feature value):

$$\text{Merkmalstruktur} = \begin{bmatrix} \text{MERKMALE} & \text{WERT} \\ \text{MERKMALE} & \text{WERT} \end{bmatrix}$$

- **Repräsentation grammatischer Merkmale** als Merkmalstruktur:

$$N \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix}$$

$$N \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix} \quad \text{oder} \quad \begin{bmatrix} \text{CAT} & N \\ \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix}$$

- Variante 1: **Kategoriensymbol** + Merkmalstruktur als **Annotation** der Merkmale
- Variante 2: Repräsentation **gesamter Kategorie** als **Merkmalstruktur** (Kategorie als Merkmal CAT)

- Merkmalstrukturen werden in der Linguistik u. a. für Beschreibung phonetischer und semantischer Merkmale verwendet
- In der Syntaxanalyse zunächst für **Modellierung der Subkategorisierung** von Verben in Generativer Grammatik verwendet
- ab 1980: **Unifikationsgrammatiken** = Modelle, deren **syntaktische Kategorien Merkmalstrukturen** sind und die die Operation der **Merkmalsunifikation** für die Steuerung des Ableitungsprozesses verwenden (PATR-II, GPSG, LFG, HPSG)

- durch **Integration von Merkmalen in Kategoriensymbole**  
→ z. B. *IV*, *TV*; *N\_Sg*, *N\_Pl*
- **2 Probleme:**
  - solche erweiterten CFGs **vervielfachen** allerdings das **Regelsystem**
  - **strukturelle Ähnlichkeit** wird nur **suggeriert**
    - z. B. *N\_Sg* und *N\_Pl* als Subkategorien von *N*
    - die atomaren Nichtterminale sind aber **beliebige Variablen ohne Zusammenhang!**

- mit Merkmalstrukturen, d.h. aus **Merkmal-Wert-Paaren** zusammengesetzten komplexen Objekten, lassen sich **grammatikalische Zusammenhänge beschreibungsadäquater modellieren**:

$$\begin{bmatrix} \text{CAT} & N \\ \text{NUM} & SG \\ \text{CASE} & NOM \end{bmatrix}$$
$$\begin{bmatrix} \text{CAT} & N \\ \text{NUM} & PL \\ \text{CASE} & NOM \end{bmatrix}$$

# Unterspezifikation

- sowohl lexikalische Einheiten als auch lexikalische Kategorien können repräsentiert werden über ihre Merkmale:  
→ je weniger Merkmale (Informationen) desto allgemeinere Klasse von linguist. Objekten ist repräsentiert (**Unterspezifikation**):

Wortformen: *Hunden*  $\begin{bmatrix} \text{CAT} & N \\ \text{NUM} & PL \\ \text{GEN} & MASK \\ \text{CASE} & DAT \end{bmatrix}$ , *der*  $\begin{bmatrix} \text{CAT} & DET \\ \text{NUM} & SG \\ \text{GEN} & MASK \\ \text{CASE} & NOM \end{bmatrix}$

lexikalische Subkategorien (Maskulina):  $\begin{bmatrix} \text{CAT} & N \\ \text{GEN} & MASK \end{bmatrix}$

lexikalische Kategorien:  $\begin{bmatrix} \text{CAT} & N \end{bmatrix}$     $\begin{bmatrix} \text{CAT} & DET \end{bmatrix}$

- Merkmale innerhalb einer Merkmalstruktur können **Beschreibungen für die gleiche linguistische Einheit** sein (**koreferent** sein)
- durch **Forderung nach Koreferenz von Merkmalen** von durch PSG-Regeln festgelegte Konstituenten einer syntaktischen Kategorie (untereinander oder mit Merkmalen der Kategorie) können **Abhängigkeiten wie Kongruenz und Rektion modelliert werden** (=Beschränkungen/Constraintregeln)

$$\begin{bmatrix} \text{CAT} & \text{NP} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & \text{DET} \\ \text{AGR} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & \text{N} \\ \text{AGR} & \boxed{1} \end{bmatrix}$$

# Komplexe Werte

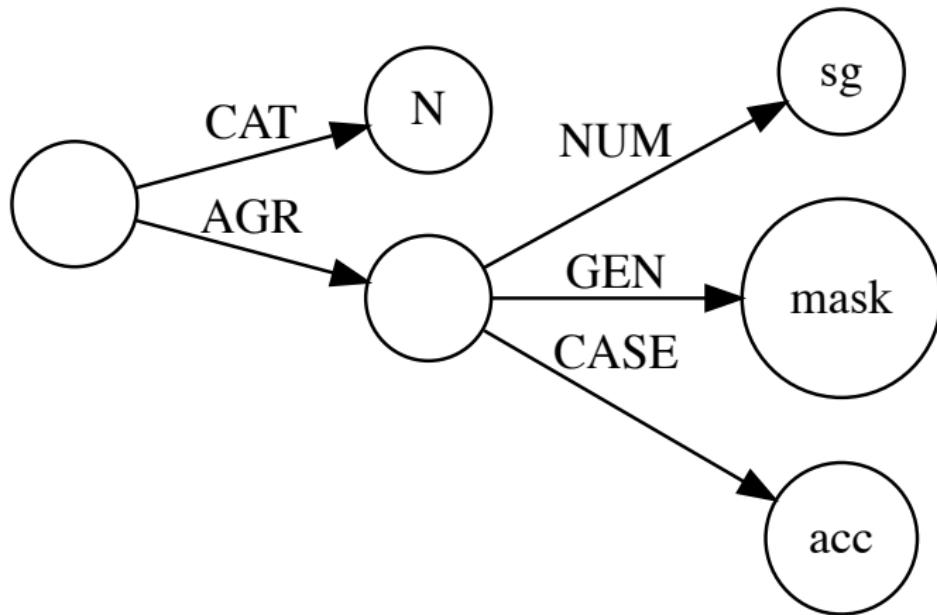
- neben atomaren Werten (SG, +) können auch **Merkmalsstrukturen als Werte in einer Merkmalsstruktur** vorkommen
- damit lassen sich **Kongruenzmerkmale** zusammenfassen:

CAT	N						
AGR	<table><tr><td>NUM</td><td>SG</td></tr><tr><td>GEN</td><td>MASK</td></tr><tr><td>CASE</td><td>ACC</td></tr></table>	NUM	SG	GEN	MASK	CASE	ACC
NUM	SG						
GEN	MASK						
CASE	ACC						

abkürzende Notation für Pfad in FeatStruct:  
(als Pfadgleichung: <N AGR CASE>=ACC)

CAT	N
AGR CASE	ACC

# Merkmalsgraph



# Lexikoneinträge mit komplexem AGR-Merkmal

Hund	CAT	N
	AGR	$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{bmatrix}$

Katze	CAT	N
	AGR	$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \end{bmatrix}$

der	CAT	DET
	AGR	$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{bmatrix}$
		$\begin{bmatrix} \text{CASE} & \text{NOM} \end{bmatrix}$

den	CAT	DET
	AGR	$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{bmatrix}$
		$\begin{bmatrix} \text{CASE} & \text{ACC} \end{bmatrix}$

die	CAT	DET
	AGR	$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \end{bmatrix}$

← Unterspezifikation Kasusmerkmal (unifiziert mit beliebigen Kasusspezifikationen)

## 3.2. Constraintregeln und Unifikation

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- **Constraintregeln und Unifikation**
- Unifikation und Subsumption

4 Typ hierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

- bloßer **Ersatz** von atomaren Kategoriensymbolen in PSG-Regeln durch **Merkmalstrukturen** schränkt **Übergenerierung nicht ein**:  
 $NP \rightarrow DET\ N$   
 $\begin{bmatrix} CAT & NP \end{bmatrix} \rightarrow \begin{bmatrix} CAT & DET \end{bmatrix} \begin{bmatrix} CAT & N \end{bmatrix}$
- **Zusatzregeln** notwendig, die **auf die Merkmale der Konstituenten Bezug nehmen** und **Abhängigkeiten** zwischen den durch unterspezifizierte Merkmalstrukturen repräsentierten linguistischen Objekten **ausdrücken** (Beschränkungen/Constraints)

- **nominales Agreement:** Beschränkung der durch die PSG-Regel repräsentierten Kombination von Determinativ und Nomen auf Übereinstimmung im AGR-Merkmal (Koreferenz)
- **Constraintregel als Pfadgleichung:**

$NP \rightarrow DET\ N$

$\langle DET\ AGR \rangle = \langle N\ AGR \rangle$

- **Alternative Darstellung mit Variable (NLTK: ?x):**

$$\begin{bmatrix} \text{CAT} & NP \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & DET \\ \text{AGR} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & N \\ \text{AGR} & \boxed{1} \end{bmatrix}$$

# Constraintregel als Unifikationsanweisung

- Anweisung auf Durchführung von **Unifikation zur Feststellung der Vereinbarkeit** dieser AGR-Teil-Merkmalstrukturen:

$\langle \text{der AGR} \rangle = \langle \text{Hund AGR} \rangle ?$

$$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix} \sqcup \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix} = \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{CASE} & \text{NOM} \end{bmatrix}$$

- Erkennung** (da unifizierbar,  $\langle \text{DET AGR} \rangle = \langle \text{N AGR} \rangle$ ):  
*der Hund, den Hund, die Katze*

# Constraintverletzung

$\langle \text{die AGR} \rangle = \langle \text{Hund AGR} \rangle?$

$$\begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \end{bmatrix} \sqcup \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{bmatrix} = \text{FAIL!}$$

- **Ablehnung** (da:  $\langle \text{DET AGR GEN} \rangle \neq \langle \text{N AGR GEN} \rangle$ ):

*die Hund* ( $\langle \text{DET AGR GEN} \rangle = \text{FEM}, \langle \text{N AGR GEN} \rangle = \text{MASK}$ )

*der Katze* ( $\langle \text{DET AGR GEN} \rangle = \text{MASK}, \langle \text{N AGR GEN} \rangle = \text{FEM}$ )

*den Katze* ( $\langle \text{DET AGR GEN} \rangle = \text{MASK}, \langle \text{N AGR GEN} \rangle = \text{FEM}$ )

### 3.3. Unifikation und Subsumption

- 1 Motivation für Feature-Modellierung
- 2 Grammatische Merkmale
  - Morphosyntaktische Flexionskategorien des Deutschen
  - Kodierung syntaktischer Funktionen
  - Flexionskategorien
  - Kasus und Agreement
- 3 Merkmalstrukturen
  - Grundlagen
  - Constraintregeln und Unifikation
  - **Unifikation und Subsumption**
- 4 Typenhierarchie und getypte Merkmalstrukturen
  - Grundlagen
  - Subsumption und Unifikation getypter Merkmalsstrukturen
  - Bedingungen

- **information combination operation** (ähnlich Mengenvereinigung)
- Zwei Merkmalstrukturen **unifizieren**, wenn sie **vereinbar** sind.
- **Ergebnis einer Unifikation:**
  - *Existiert*, wenn es (auch rekursiv) **keine widersprüchlichen Merkmal-Wert-Paare** gibt.
  - Enthält alle **Merkmal-Wert-Paare beider Merkmalstrukturen**.
    - d.h. die Informationen aus beiden Merkmalsstrukturen sind in ihrer Unifikation enthalten.

- Beispiel:  $\begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix} \sqcup \begin{bmatrix} A & 1 \\ C & 3 \end{bmatrix} = \begin{bmatrix} A & 1 \\ B & 2 \\ C & 3 \end{bmatrix}$

- NICHT definiert (widersprüchliche Werte):

$$\begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix} \sqcup \begin{bmatrix} A & 1 \\ B & 3 \end{bmatrix} ( = \text{NONE})$$

# Subsumption

- **information ordering relation** (ähnlich Teilmengenbeziehung)
- Zwei Merkmalstrukturen stehen in **Subsumption**, wenn eine Struktur **alle Merkmale der anderen enthält**.
- **Erfüllung der Subsumptionsrelation:**
  - **Ist erfüllt**, wenn die spezifischere Struktur **alle Merkmal-Wert-Paare der allgemeineren Struktur** enthält.
    - d.h.  $F_0$  **subsumiert**  $F_1$  ( $F_0 \sqsubseteq F_1$ ), wenn alle Informationen von  $F_0$  (der untergeordneten, allgemeineren Struktur) in  $F_1$  (der übergeordneten, spezifischeren enthalten) sind.
    - **d.h. existiert u.a. nicht**, wenn es **unterschiedliche Merkmal-Wert-Paare** gibt (insbesondere bei widersprüchlichen Merkmal-Wert-Paaren).
- Beispiel:  $\begin{bmatrix} A & 1 \end{bmatrix} \sqsubseteq \begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix}$  aber NICHT:  $\begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix} \sqsubseteq \begin{bmatrix} A & 1 \\ C & 3 \end{bmatrix}$

# Subsumption und Unifikation

- das Ergebnis einer Unifikation ist die **kleinste obere Schranke** in der Subsumptionsbeziehung
    - d.h. die Unifikation zweier Merkmalstrukturen  $F_0$  und  $F_1$  (falls sie existiert) ist die kleinste Merkmalstruktur, die von  $F_0$  und  $F_1$  subsumiert wird
  - alle Informationen von  $F_0$  bzw.  $F_1$  sind in ihrer Unifikation enthalten
- Beispiel:  $\begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix} \sqsubseteq \begin{bmatrix} A & 1 \\ B & 2 \\ C & 3 \end{bmatrix}$  und  $\begin{bmatrix} A & 1 \\ C & 3 \end{bmatrix} \sqsubseteq \begin{bmatrix} A & 1 \\ B & 2 \\ C & 3 \end{bmatrix}$

# 4. Typhierarchie und getypte Merkmalstrukturen

1 Motivation für Feature-Modellierung

2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

4 Typhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- Bedingungen

# 4.1. Grundlagen

- 1 Motivation für Feature-Modellierung
- 2 Grammatische Merkmale
  - Morphosyntaktische Flexionskategorien des Deutschen
  - Kodierung syntaktischer Funktionen
  - Flexionskategorien
  - Kasus und Agreement
- 3 Merkmalstrukturen
  - Grundlagen
  - Constraintregeln und Unifikation
  - Unifikation und Subsumption
- 4 Typhierarchie und getypte Merkmalstrukturen
  - **Grundlagen**
  - Subsumption und Unifikation getypter Merkmalsstrukturen
  - Bedingungen

# Getypte Merkmalsstrukturen

- durch Definition einer Typhierarchie (bzgl. Subsumption) für die Werte eines Merkmals können auch nicht-identische Werte (Types) unifizieren

- Beispiel:  $\begin{bmatrix} \text{PERS} & 1 \end{bmatrix} \sqsubseteq \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & sg \end{bmatrix}$

$$\begin{bmatrix} \text{PERS} & 1 \end{bmatrix} \sqcup \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & sg \end{bmatrix} = \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & sg \end{bmatrix}$$

- Getypte Merkmalsstruktur:  $\begin{bmatrix} \text{AGR} & 1 \end{bmatrix} \sqsubseteq \begin{bmatrix} \text{AGR} & 1sg \end{bmatrix}$

$$\begin{bmatrix} \text{AGR} & 1 \end{bmatrix} \sqcup \begin{bmatrix} \text{AGR} & 1sg \end{bmatrix} = \begin{bmatrix} \text{AGR} & 1sg \end{bmatrix}$$

## Definition (Type, $\sqsubseteq$ )

Sei **Type** eine endliche Menge von Typen mit **Vererbungshierarchie**  $\sqsubseteq$ .

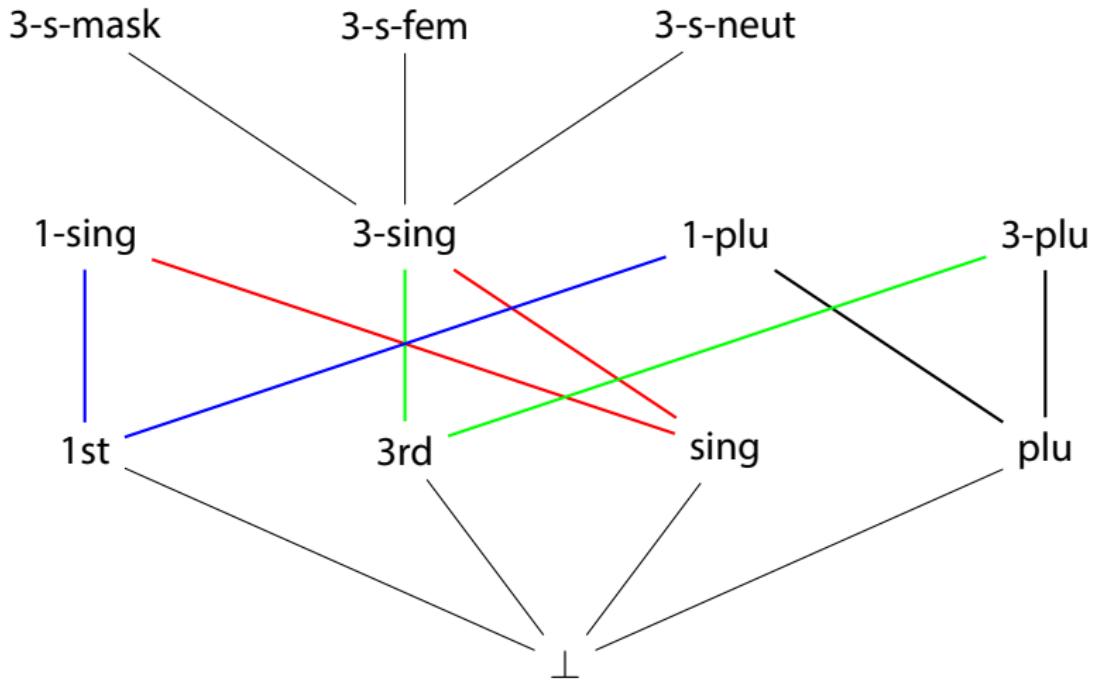
Wenn für  $A, B \in \text{Type}$  gilt, dass  $A \sqsubseteq B$ , dann

- erbt  $B$  Informationen von  $A$ .
- ist  $A$  Obertyp von  $B$ . (Alle  $A$ -Attribute sind auch  $B$ -Attribute.)
- $A$  subsumiert  $B$  ( $B$  wird von  $A$  subsumiert).
- ist  $A$  „allgemeiner oder gleich“  $B$ .
- ist  $B$  „spezieller oder gleich“  $A$ .

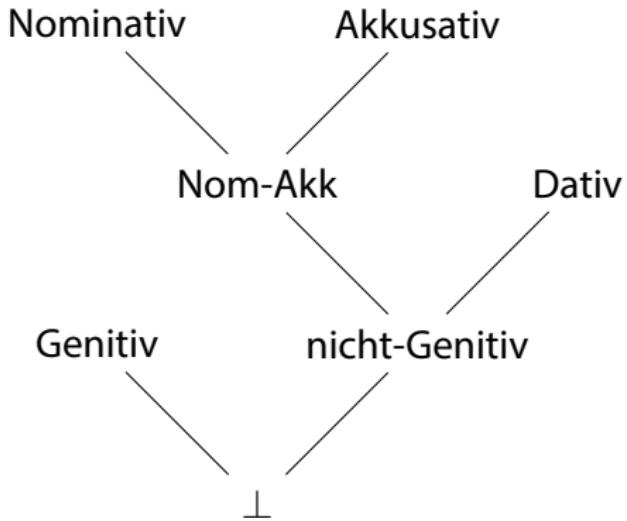
## Eigenschaften von Type, $\sqsubseteq$

- wohldefinierte Unifikationsoperation
- transitiv ( $\forall A, B, C \in \text{Type}. A \sqsubseteq B \wedge B \sqsubseteq C \Rightarrow A \sqsubseteq C$ )
- reflexiv ( $\forall A \in \text{Type}. A \sqsubseteq A$ )
- antisymmetrisch ( $\forall A, B \in \text{Type}. A \sqsubseteq B \wedge B \sqsubseteq A \Rightarrow A = B$ ) (keine Vererbungsschleifen)  
⇒ *partielle Ordnung* (d. h. nicht alle Elemente von Type müssen miteinander vergleichbar sein)
- Existenz eines eindeutigen allgemeinsten Typs  
( $\exists_1 A \in \text{Type}. \forall B \in \text{Type}. A \sqsubseteq B$ )  
⇒  $\perp$  definiert als kleinstes Element von Type bzgl.  $\sqsubseteq$

# Beispiel 1: Agreement-Typhierarchie (Person + Numerus)



## Beispiel 2: Kasus-Typhierarchie



Vgl. z. B. die Paradigmen:

der Hund, des **Hundes**, dem Hund, den Hund

**das** Buch, des **Buches**, dem Buch, **das** Buch

## Definition (Feat)

Sei **Feat** eine *endliche* Menge von Merkmalen (engl. *features*).

(Ohne weitere Anforderungen an Struktur oder Eigenschaften)

## Beispiel

**Feat** = {GEN, CASE, NUM, AGR, PER, MOOD, CAT, TENSE}

## Definition

Eine Merkmalstruktur über **Type** und **Feat** ist definiert als Tupel  $F = (Q, \bar{q}, \theta, \delta)$  mit:

- $Q$ : endliche Menge von Knoten (Einträge)
- $\bar{q} \in Q$ : Wurzelknoten
- $\theta : Q \rightarrow \text{Type}$ : totale Typisierungsfunktion
- $\delta : \text{Feat} \times Q \rightarrow Q$ : partielle Merkmal-Wert-Funktion

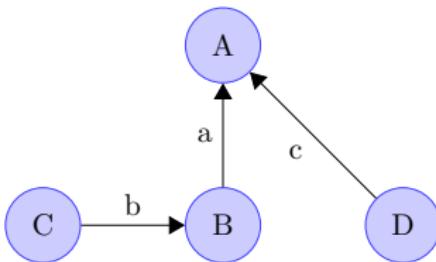
Sei  $\mathcal{F}$  die Menge aller Merkmalstrukturen.

CAT	$N$
AGR	$\begin{bmatrix} \text{CASE} & \textit{Nom} \\ \text{NUM} & \textit{Pl} \end{bmatrix}$

## Beschrifteter Graph

Ein *beschrifteter Graph* ist definiert als Tupel  $G = (V, E, I_V, I_E, L_V, L_E)$  mit

- $V$ : Menge der Knoten (engl. *vertices*)
- $E \subseteq V \times V$ : Menge der Kanten (engl. *edges*)
- $I_V : V \rightarrow L_V$ : Beschriftungsfunktion für Knoten (engl. *label*)
- $I_E : E \rightarrow L_E$ : Beschriftungsfunktion für Kanten
- $L_X$ : Menge von Beschriftungen für  $X$



## Visualisierung

Der Graph zu einer Merkmalstruktur  $F = (Q, \bar{q}, \theta, \delta)$  ist gegeben durch:

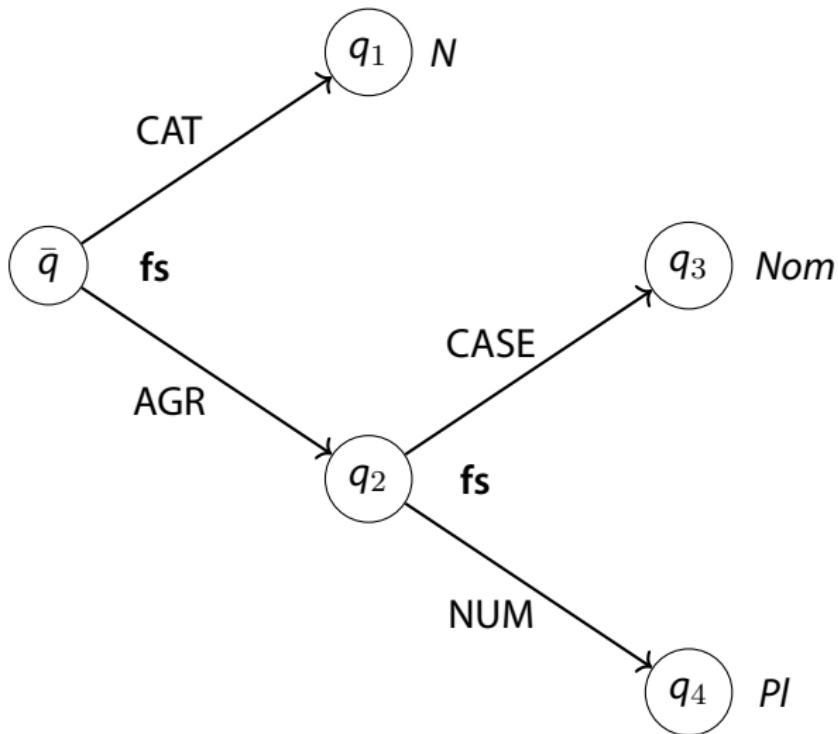
- $V := Q$
- $E := \{(q_1, q_2) \mid \exists f. \delta(f, q_1) = q_2\}$
- $L_V := \text{Type}; l_V := \theta$
- $L_E := \text{Feat}; l_E(q_1, q_2) := \{f \mid \delta(f, q_1) = q_2\}$

## Anmerkung

Zur Vereinfachung werden einelementige Mengen ohne Mengenklammern geschrieben.

Also  $a$  statt  $\{a\}$ .

## Beispiel: Graphdarstellung



## Variablen

- **Var** sei eine abzählbar unendliche Menge von Variablen.
- Häufig wird **Var** =  $\mathbb{N}$  benutzt.
- Es gibt aber auch andere Möglichkeiten;  
z. B. im NLTK: ASCII-Identifier (?x, ?y, ...)

## Definition (Zuweisungsfunktion, Valuation)

Eine Zuweisung  $\alpha : \mathbf{Var} \rightarrow \mathcal{F}$  ist eine totale Funktion, die alle Variablen an Merkmalstrukturen (Knoten, Einträge) bindet.

## Reentrance (dt. *Wiedereintritt*)

Durch das Aufstellen von Bedingungen (s. später) können Variablen an verschiedene Teile von Merkmalstrukturen gebunden werden. *Diese müssen gleich sein.*

## Beispiel

ORTH	<i>folgt</i>				
SYN	<table><tr><td>SBJ</td><td>[1]</td></tr><tr><td>OBJ</td><td>[2]</td></tr></table>	SBJ	[1]	OBJ	[2]
SBJ	[1]				
OBJ	[2]				
SEM	<table><tr><td>AGT</td><td>[1]</td></tr><tr><td>PAT</td><td>[2]</td></tr></table>	AGT	[1]	PAT	[2]
AGT	[1]				
PAT	[2]				

ORTH	<i>folgt</i>				
SYN	<table><tr><td>SBJ</td><td>[1]<i>Hund</i></td></tr><tr><td>OBJ</td><td>[2]<i>Katze</i></td></tr></table>	SBJ	[1] <i>Hund</i>	OBJ	[2] <i>Katze</i>
SBJ	[1] <i>Hund</i>				
OBJ	[2] <i>Katze</i>				
SEM	<table><tr><td>AGT</td><td>[1]</td></tr><tr><td>PAT</td><td>[2]</td></tr></table>	AGT	[1]	PAT	[2]
AGT	[1]				
PAT	[2]				

## 4.2. Subsumption und Unifikation getypter Merkmalsstrukturen

### 1 Motivation für Feature-Modellierung

### 2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

### 3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

### 4 Typhierarchie und getypte Merkmalstrukturen

- Grundlagen
- **Subsumption und Unifikation getypter Merkmalsstrukturen**
- Bedingungen

## Erweiterung auf Merkmalstrukturen

$F = (Q, \bar{q}, \theta, \delta)$  subsumiert  $F' = (Q', \bar{q}', \theta', \delta')$ , genau dann wenn es eine totale Funktion  $h : Q \rightarrow Q'$  gibt, sodass:

- $h(\bar{q}) = \bar{q}'$
- $\theta(q) \sqsubseteq \theta'(h(q))$  für alle  $q \in Q$
- $h(\delta(f, q)) = \delta'(f, h(q))$  für alle  $q \in Q$  und  $f \in \text{Feat}$ ,  
für die  $\delta(f, q)$  definiert ist

## Beispiel

$$\bar{q} \begin{bmatrix} q_1 \text{ CAT} & N \end{bmatrix} \sqsubseteq \bar{q}' \begin{bmatrix} q'_1 \text{ CAT} & N \\ q'_2 \text{ GEN} & mask \end{bmatrix} \quad h(\bar{q}) = \bar{q}' \quad h(q_1) = q'_1$$

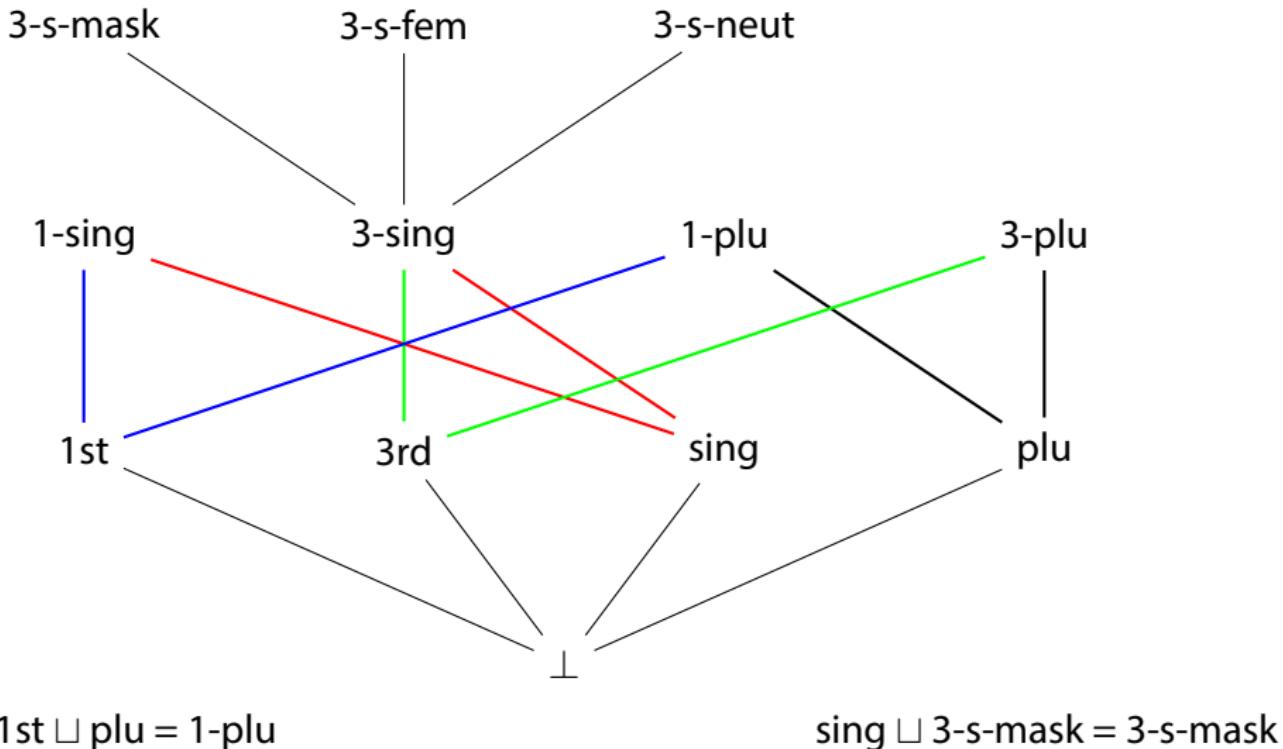
## Unifikation ( $\sqcup$ ) für Typen

- Das Ergebnis der Unifikation zweier Typen  $A, B \in \mathbf{Type}$  ist ihre kleinste obere Schranke in **Type** bzgl.  $\sqsubseteq$ .
- Diese kann auch undefiniert sein (Typen unifizieren nicht).
- $A \sqcup B = C \iff A \sqsubseteq C \text{ und } B \sqsubseteq C \text{ und }$

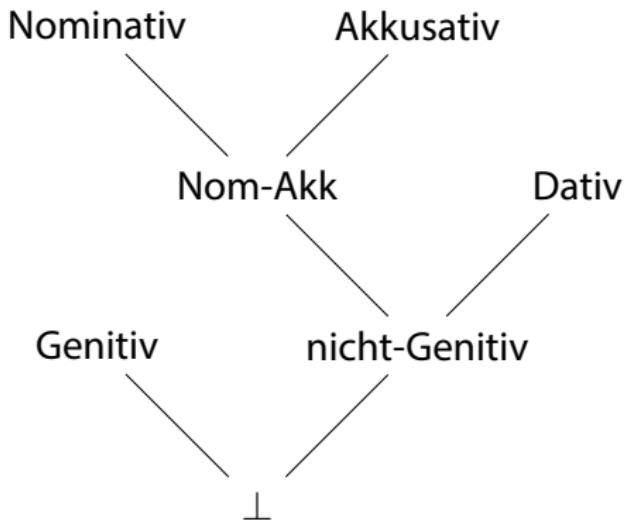
$$\forall D \in \mathbf{Type}. A \sqsubseteq D \wedge B \sqsubseteq D \implies C \sqsubseteq D$$

(Vgl. Mengenvereinigung und Untermengenbeziehung)

# Beispiel 1: Typunifikation



## Beispiel 2: Typunifikation



nicht-Genitiv  $\sqcup$  Nominativ = Nominativ

Nom-Akk  $\sqcup$  Dativ = *undefiniert*

## Unifikation ( $\sqcup$ ) für Merkmalstrukturen

- Idee: Unifikation ebenfalls kleinste obere Schranke bzgl. der Subsumptionsrelation  $\sqsubseteq$  auf Merkmalstrukturen
- Algorithmus in zwei Schritten:
  - ① Identifiziere korrespondierende (äquivalente) Knoten
  - ② Unifizierte deren Typen

### Formale Definition: Identifikation (Schritt 1)

Für Merkmalstrukturen  $F = (Q, \bar{q}, \theta, \delta), F' = (Q', \bar{q}', \theta', \delta')$  mit  $Q \cap Q' = \emptyset$  sei die Äquivalenzrelation  $\equiv$  wie folgt definiert:

- $\bar{q} \equiv \bar{q}'$
- $\delta(f, q) \equiv \delta'(f, q')$  wenn beide Seiten definiert und  $q \equiv q'$

## Formale Definition: Typunifikation (Schritt 2)

Die Unifikation von  $F$  und  $F'$  ist dann wie folgt definiert:

$$F \sqcup F' = ((Q \cup Q') / \equiv, [\bar{q}] \equiv, \theta^{\equiv}, \delta^{\equiv})$$

mit

$$\theta^{\equiv}([q] \equiv) = \bigsqcup \{(\theta \cup \theta')(q') \mid q' \equiv q\}$$

und

$$\delta^{\equiv}(f, [q] \equiv) = \begin{cases} [(\delta \cup \delta')(f, q)] \equiv & \text{falls } (\delta \cup \delta')(f, q) \text{ definiert} \\ \text{undefiniert} & \text{sonst} \end{cases}$$

## Notation (für $\equiv$ Äquivalenzrelation über $X$ )

- $[x] \equiv = \{y \in X \mid y \equiv x\}$
- $X / \equiv = \{[x] \equiv \mid x \in X\}$

# Beispiel: (Formale) Unifikation

$q_1 \left[ \begin{array}{ll} q_2 \text{ CAT } N \\ q_3 \text{ AGR } \left[ \begin{array}{ll} q_4 \text{ NUM } Sg \\ q_5 \text{ CAS } \text{ nicht-Gen} \end{array} \right] \end{array} \right]$	$q_6 \left[ \begin{array}{ll} q_7 \text{ ORTH } \text{ Hund} \\ q_8 \text{ AGR } \left[ \begin{array}{ll} q_9 \text{ NUM } Sg \\ q_{10} \text{ CAS } \text{ Nom} \end{array} \right] \end{array} \right]$
--	---

## ① Identifikation korrespondierender Knoten

- $q_1 \equiv q_6$  (Initialisierung)
- Nach 1 Schritt mit  $\delta$ :
  - $q_3 \equiv q_8$
- Nach 2 Schritten mit  $\delta$ :
  - $q_4 \equiv q_9$
  - $q_5 \equiv q_{10}$

# Beispiel: (Formale) Unifikation

$$q_1 \begin{bmatrix} q_2 \text{ CAT} & N \\ q_3 \text{ AGR} & \begin{bmatrix} q_4 \text{ NUM} & Sg \\ q_5 \text{ CAS} & \text{nicht-Gen} \end{bmatrix} \end{bmatrix} \quad q_6 \begin{bmatrix} q_7 \text{ ORTH} & \text{Hund} \\ q_8 \text{ AGR} & \begin{bmatrix} q_9 \text{ NUM} & Sg \\ q_{10} \text{ CAS} & \text{Nom} \end{bmatrix} \end{bmatrix}$$

## ② Typunifikation

- $Q_U = \{\{q_1, q_6\}, \{q_2\}, \{q_7\}, \{q_3, q_8\}, \{q_4, q_9\}, \{q_5, q_{10}\}\}$
- $\bar{q}_U = \{q_1, q_6\}$
- $\theta^{\equiv}(\{q_2\}) = N, \theta^{\equiv}(\{q_7\}) = \text{Hund}, \theta^{\equiv}(\{q_3, q_8\}) = \mathbf{fs},$   
 $\theta^{\equiv}(\{q_4, q_9\}) = Sg, \theta^{\equiv}(\{q_5, q_{10}\}) = \text{Nom}, \theta^{\equiv}(\{q_1, q_6\}) = \mathbf{fs}$
- $\delta(\text{CAT}, \{q_1, q_6\}) = \{q_2\}, \delta(\text{ORTH}, \{q_1, q_6\}) = \{q_7\}, \dots$

## Lemma

Wenn  $F \sqcup F'$  definiert ist, dann ist  $F \sqcup F' \in \mathcal{F}$  eine Merkmalstruktur.

## Theorem

$F \sqcup F'$  ist die *kleinste obere Schranke* von  $F$  und  $F'$  in  $(\mathcal{F}, \sqsubseteq)$ , falls  $F$  und  $F'$  eine obere Schranke haben.

Für Beweise siehe (**Carpenter:Log-TyFeat**).

## 4.3. Bedingungen

### 1 Motivation für Feature-Modellierung

### 2 Grammatische Merkmale

- Morphosyntaktische Flexionskategorien des Deutschen
- Kodierung syntaktischer Funktionen
- Flexionskategorien
- Kasus und Agreement

### 3 Merkmalstrukturen

- Grundlagen
- Constraintregeln und Unifikation
- Unifikation und Subsumption

### 4 Typhierarchie und getypte Merkmalstrukturen

- Grundlagen
- Subsumption und Unifikation getypter Merkmalsstrukturen
- **Bedingungen**

## Pfade

- Sequenzen von Merkmalen werden *Pfade* genannt.
- **Path** =  $\text{Feat}^*$  sei die Menge aller Pfade.
- Für  $p \in \text{Path}$ ,  $F \in \mathcal{F}$  sei  $F@p$  der Knoten in  $F$ , den man am Ende von Pfad  $p$  erhält.

## Beispiele

- AGR–NUM
- SYN–SBJ–AGR–NUM
- ORTH
- $\varepsilon$  (der leere Pfad)

## Definition (Beschreibung Desc)

Die Menge der Beschreibungen über **Type** und **Feat** sei die kleinste Menge, die folgende Bedingungen erfüllt:

- $A \in \text{Desc}$ , für alle  $A \in \text{Type}$
- $p : d \in \text{Desc}$ , für  $p \in \text{Path}$ ,  $d \in \text{Desc}$
- $x \in \text{Desc}$ , für alle  $x \in \text{Var}$
- $d \wedge e \in \text{Desc}$ , für  $d, e \in \text{Desc}$

## Beispiel

- AGR–NUM :  $Sg$
- SYN–SBJ :  $\boxed{1} \wedge \text{SEM–AGT} : \boxed{1}$

## Erfülltheit

Die Erfülltheitsrelation  $\models^\alpha$  zwischen Merkmalstrukturen und Beschreibungen ist gegeben durch:

- Für  $A \in \mathbf{Type}, F \models^\alpha A \iff A \sqsubseteq \theta(\bar{q})$
- $F \models^\alpha p : d \iff F @ p \models^\alpha d$
- Für  $x \in \mathbf{Var}, F \models^\alpha x \iff \alpha(x) = F$
- $F \models^\alpha d \wedge e \iff F \models^\alpha d \text{ und } F \models^\alpha e$

# Erfülltheit: Beispiel

Sei  $F$  eine Merkmalstruktur.

$$F = \begin{bmatrix} \text{CAT} & \boxed{1}N \\ \text{POS} & \boxed{2} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & Sg \\ \text{CAS} & \text{Nominativ} \end{bmatrix} \end{bmatrix} \quad \alpha(\boxed{1}) = \alpha(\boxed{2})$$

Welche Beschreibungen aus Desc erfüllt  $F$ ?

- $F \models^\alpha N ?$  Nein!
  - $F \models^\alpha \text{CAT}: N ?$  Ja!
  - $F \models^\alpha \text{AGR-CAS}: \text{nicht-Genitiv} ?$  Ja!
  - $F \models^\alpha \text{POS}: N ?$  Ja!
- Denn: *nicht-Genitiv*  $\sqsubseteq$  *Nominativ*

# Beschreibungen als Merkmalstrukturen

## *MGSat* (allgemeinster Erfüller)

Zu jeder *konsistenten* (widerspruchsfreien) Beschreibung  $d \in \mathbf{Desc}$  gibt es eine Merkmalstruktur  $MGSat(d) \in \mathcal{F}$  mit der Eigenschaft

$$\forall F \in \mathcal{F}. F \models d \iff MGSat(d) \sqsubseteq F$$

## Konstruktion

- Für  $A \in \mathbf{Type}$ :  $MGSat(A) = [A]$
- $MGSat(f_1 f_2 \dots f_n : d) = \left[ f_1 \quad \left[ f_2 \quad \dots \left[ f_n \quad MGSat(d) \right] \right] \right]$
- Wenn  $\mathbf{Var} = \mathbb{N}$ , dann  $MGSat(1) = [1]$
- $MGSat(d \wedge e) = MGSat(d) \sqcup MGSat(e)$

# Bedingungsprüfung per Unifikation: Beispiel

## Grammatikregel mit Constraint

NP [CAS=?y] → DET [GEN=?x, CAS=?y] N [GEN=?x]

## Bedingungen als Beschreibungen

- type : NP ∧ CAS : 2
- type : DET ∧ GEN : 1 ∧ CAS : 2
- type : N ∧ GEN : 1

## Bedingungen als Merkmalstrukturen

$$\begin{bmatrix} \text{type} & \text{NP} \\ \text{CAS} & \boxed{2} \end{bmatrix} \rightarrow \begin{bmatrix} \text{type} & \text{DET} \\ \text{GEN} & \boxed{1} \\ \text{CAS} & \boxed{2} \end{bmatrix} \begin{bmatrix} \text{type} & \text{N} \\ \text{GEN} & \boxed{1} \end{bmatrix}$$