

Syntax natürlicher Sprachen

Vorlesung 6: Grammatische Merkmale

A. Wisiorek

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

29.11.2022

Übgenerierung 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 grammatikalischen Merkmalen über **Merkmalsstrukturen**
- morphosyntaktische **Constraintregeln** als Anweisung auf Durchführung von **Unifikation** der Merkmalsstrukturen
→ *Unifikation: Vereinbarkeit von Merkmalsstrukturen*

1. Sprachliche Ausdrucksmittel syntaktischer Funktionen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatical Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation

Typen syntaktischer Kodierung

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

A: Strukturelle Kodierung

- 1 **Wortstellung** (s. Sitzung 8)

B: Morphologische Kodierung

- 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; s. Sitzung 8)
- 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**

repräsentative Ausdrucksmittel syntaktischer Funktionen

2. Grammatische Merkmale

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation

2.1. Flexionskategorien

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation

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

- Flexionskategorie = Grammatisches Merkmal
 - **Merkmal** hat Merkmalsausprägungen = **Werte**
 - z. B. **grammatisches Merkmal/Kategorie Numerus:**
hat die Werte: SG, PL
- Merkmalsausprägungen werden durch **Morpheme** kodiert
 - **Morphem** = **kleinste bedeutungstragende Einheit** der Sprache
 - nicht weiter segmentierbare substantielle **Form-Funktions-Paare**
 - z. B. Pluralmorphem Englisch: -s = PL

- **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)

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**

- **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, s. Übung)

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

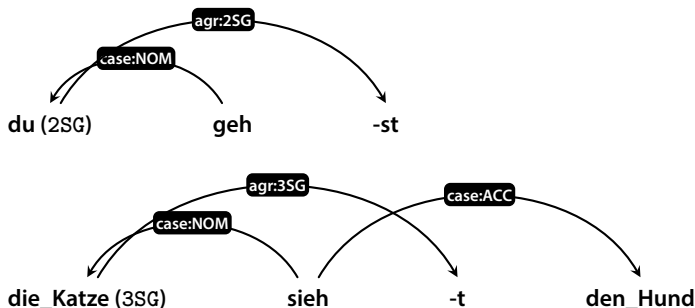
- auch Voice/Diathese: Kategorisierung der **Abbildung von semantischen Rollen auf die Grammatischen Relationen** (s.u.)

2.2. Kasus und Agreement

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - **Kasus und Agreement**
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation

Markierung zentraler syntaktischer Satzgliedfunktionen

- 1 **Kasus:** *Markierung der syntaktischen Funktion* eines verbalen Dependents *am Dependents* (*dependent-marking*)
- 2 **Agreement:** *Markierung der syntaktischen Funktion* eines verbalen Dependents *am Kopf* (Verb) über Kongruenz in Merkmalen mit dem Dependents (*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)
- **Typ2: Modifikation:** Markierung Nomen als Modifikator des Verbs (Adjunkt)

- **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** (auch: Genitiv/Dativ/Präpos.)
 - **Dativ** als Kasus des **indirekten Objekts**
 - **Präpositionen** und z.T. auch **Genitiv** und **Akkusativ** als **Adverbialkasus**

- Markierung Grammatischer Relationen durch **Übereinstimmung des Kopfes** in grammatischen Merkmalen **mit Merkmalen des Dependents**
- **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	$\Leftarrow \text{AGR} \Rightarrow$	Person
Numerus	$\Leftarrow \text{AGR} \Rightarrow$	Numerus
Genus		Tempus
		Modus
Case	\Leftarrow	

- 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).

- 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. Merkmalsstrukturen: Repräsentation grammatischer Kategorien

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation

3.1. Grundlagen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - **Grundlagen**
 - Constraintregeln und Unifikation

- auch: **Attribut-Wert-Matrix** (AVM)
- **formale Repräsentation komplexer Objekte**, die durch eine **Anzahl an Eigenschaften** definiert sind :

$$\text{Merkmalsstruktur} = \begin{bmatrix} \text{MERKMAL1} & \text{WERT1} \\ \text{MERKMAL2} & \text{WERT2} \end{bmatrix}$$

- **Repräsentation grammatischer Merkmale** als Merkmalsstruktur:

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

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

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

- Merkmalsstrukturen 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 Merkmalsstrukturen** sind und die die Operation der **Merkmalsunifikation** für die Steuerung des Ableitungsprozesses verwenden (PATR-II, GPSG, LFG, HPSG)

- **Modellierung der morphosyntaktischen Struktur (grammatischer Merkmale und ihrer Abhängigkeiten) einer Sprache, insbesondere von Rektions- und Kongruenzbeziehungen**
- **Nichtberücksichtigung in CFGs führt zu Übergenerierung**

- 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 **suggestiert**
→ z. B. *N_Sg* und *N_Pl* als Subkategorien von *N*
→ die atomaren Nichtterminale sind aber **beliebige Variablen ohne Zusammenhang!**

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

CAT	<i>N</i>
NUM	<i>SG</i>
CASE	<i>NOM</i>

CAT	<i>N</i>
NUM	<i>PL</i>
CASE	<i>NOM</i>

- **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: $\begin{bmatrix} \text{CAT} & N \end{bmatrix}$ $\begin{bmatrix} \text{CAT} & DET \end{bmatrix}$

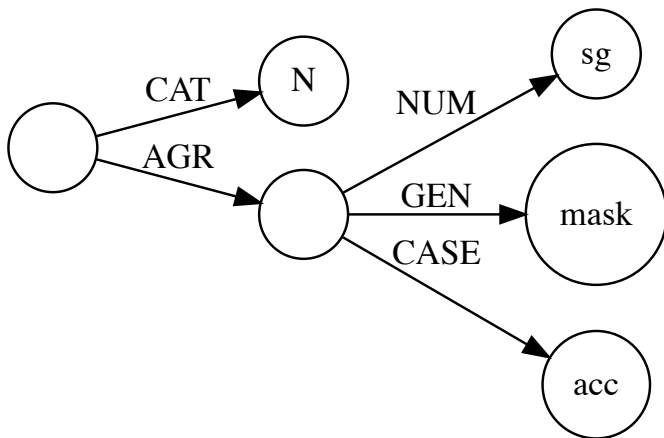
- Merkmale innerhalb einer Merkmalsstruktur können **Beschreibungen für die gleiche linguistische Einheit** sein (koreferent sein; s. Übung)
- durch **Forderung nach Koreferenz von Merkmalen** von durch PSG-Regeln festgelegte Konstituenten einer syntaktischen Kategorie (untereinander oder mit Merkmalen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

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

$$\left[\begin{array}{cc} \text{CAT} & N \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & MASK \\ \text{CASE} & ACC \end{array} \right] \end{array} \right]$$

abkürzende Notation für Pfad in AVM:
(als Pfadgleichung: $\langle N \text{ AGR CASE} \rangle = ACC$)

$$\left[\begin{array}{cc} \text{CAT} & N \\ \text{AGR|CASE} & ACC \end{array} \right]$$



Hund $\left[\begin{array}{cc} \text{CAT} & N \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & MASK \end{array} \right] \end{array} \right]$

der $\left[\begin{array}{cc} \text{CAT} & DET \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & MASK \\ \text{CASE} & NOM \end{array} \right] \end{array} \right]$

die $\left[\begin{array}{cc} \text{CAT} & DET \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & FEM \end{array} \right] \end{array} \right]$

Katze $\left[\begin{array}{cc} \text{CAT} & N \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & FEM \end{array} \right] \end{array} \right]$

den $\left[\begin{array}{cc} \text{CAT} & DET \\ \text{AGR} & \left[\begin{array}{cc} \text{NUM} & SG \\ \text{GEN} & MASK \\ \text{CASE} & ACC \end{array} \right] \end{array} \right]$

← Unterspezifikation Kasusmerkmal (unifiziert mit beliebigen Kasusspezifikationen)

3.2. Constraintregeln und Unifikation

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - **Constraintregeln und Unifikation**

- bloßer **Ersatz** von atomaren Kategoriensymbolen in PSG-Regeln **durch Merkmalsstrukturen** schränkt **Übergenerierung** nicht ein:

$NP \rightarrow DET\ N$

$$\left[\text{CAT} \quad NP \right] \rightarrow \left[\text{CAT} \quad DET \right] \left[\text{CAT} \quad N \right]$$

- **Zusatzregeln** notwendig, die auf die **Merkmale der Konstituenten** **Bezug nehmen** und **Abhängigkeiten** zwischen den durch unterspezifizierte Merkmalsstrukturen 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):**

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

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

<der AGR> = <Hund AGR>?

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

- **Erkennung** (da unifizierbar, <DET AGR> = <N AGR>):
der Hund, den Hund, die Katze

<die AGR>= <Hund AGR>?

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

- **Ablehnung** (da: <DET AGR GEN> \neq <N AGR GEN>):

die Hund (<DET AGR GEN> = FEM, <N AGR GEN> = MASK)

der Katze (<DET AGR GEN> = MASK, <N AGR GEN> = FEM)

den Katze (<DET AGR GEN> = MASK, <N AGR GEN> = FEM)

- Zwei Merkmalsstrukturen **unifizieren**, wenn sie **vereinbar** sind.
- **Ergebnis einer Unifikation:**
 - **existiert** nur, wenn es (auch rekursiv) **keine widersprüchlichen Merkmal-Wert-Paare** gibt
 - **enthält** alle **Merkmal-Wert-Paare** beider Merkmalstrukturen

Rückblick auf heutige Themen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement
- 3 Merkmalsstrukturen: Repräsentation grammatischer Kategorien
 - Grundlagen
 - Constraintregeln und Unifikation