

Syntax natürlicher Sprachen

Vorlesung 6: Grammatische Merkmale

A. Wisiorek

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

23.11.2021

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 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

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 als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

2.1. Flexionskategorien

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - **Flexionskategorien**
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- **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)

- 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

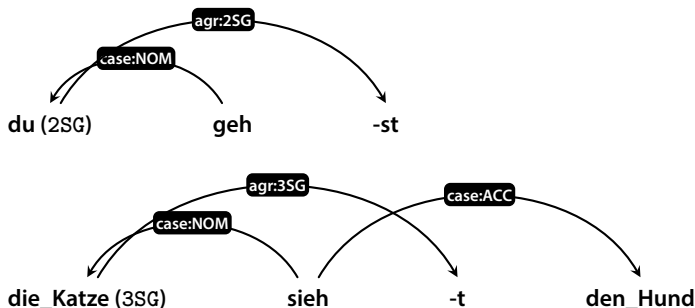
- auch Voice/Diathese: Kategorisierung der **Abbildung von semantischen Rollen auf die Grammatischen Relationen** (s.u.)

2.2. Kasus und Agreement als Marker Gramm. Relationen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

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	$\Leftarrow AGR \Rightarrow$	Person
Numerus	$\Leftarrow AGR \Rightarrow$	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

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 **Merkmalsstrukturen**
 - **Formale Repräsentation grammatischer Kategorien**
 - **Merkmalsstrukturen im NLTK**
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

3.1. Formale Repräsentation grammatischer Kategorien

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - **Formale Repräsentation grammatischer Kategorien**
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- 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:**

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

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

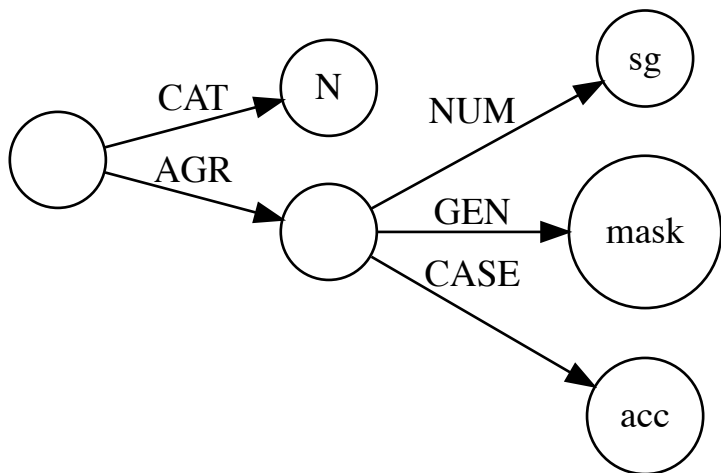
- Merkmale innerhalb einer Merkmalsstruktur können **Beschreibungen für die gleiche linguistische Einheit** sein (**koreferent** sein; s. Übung)
- durch **Forderung nach Koreferenz von Merkmalen** von durch PSG-Regeln festgelegte Konstituenten einer syntaktischen Kategorie (untereinander oder mit Merkmalen 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: $\langle N \text{ AGR CASE} \rangle = ACC$)

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



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

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

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

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

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

← Unterspezifikation Kasusmerkmal (unifiziert mit beliebigen Kasusspezifikationen)

- 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

3.2. Merkmalsstrukturen im NLTK

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - **Merkmalsstrukturen im NLTK**
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

```

1  http://www.nltk.org/howto/featstruct.html
2
3  fs1 = FeatStruct(number='singular', person=3)
4  print(fs1)
5  #[ number = 'singular' ]
6  #[ person = 3          ]
7
8  #nested feature structure:
9  fs2 = FeatStruct(type='NP', agr=fs1)
10 print(fs2)
11 #[ agr  = [ number = 'singular' ] ]
12 #[          [ person = 3          ] ]
13 #[          ]
14 #[ type = 'NP' ]

```

```
1
2 #Variables are used to indicate that two features
   should be assigned the same value. For example, the
   following feature structure requires that the
   feature fs3['agr']['number'] be bound to the same
   value as the feature fs3['subj']['number'].
3 fs3 = FeatStruct(agr=FeatStruct(number=Variable('?n')),
   subj=FeatStruct(number=Variable('?n')))
4 print(fs3)
5 #[ agr = [ number = ?n ] ]
6 #[
7 #[ subj = [ number = ?n ] ]
```

```
1
2 #unification:
3 print(fs2.unify(fs3))
4 #[ agr  = [ number = 'singular' ] ]
5 #[      [ person = 3           ] ]
6 #[                                           ]
7 #[ subj = [ number = 'singular' ] ]
8 #[                                           ]
9 #[ type = 'NP'                             ]
10
11
12 #failed unification (inconsistent feature structures):
13 fs4 = FeatStruct(agr=FeatStruct(person=1))
14 print(fs4.unify(fs2))
15 #None
16 print(fs2.unify(fs4))
17 #None
```

4. **Varianz syntaktischer Kodierung*

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- **funktionale Kategorien** wie die **semantische** oder die **pragmatische Rolle von Argumenten** werden in funktionalen Ansätzen zur **Erklärung der Akzeptabilität syntaktischer Strukturen** verwendet
- Integration funktionaler Kategorien **in die Beschreibung** von syntaktischen Einheiten **durch Merkmalsstrukturen**
- anders als im Deutschen und den umliegenden Sprachen gibt es **Sprachsysteme**, die **primär die semantischen oder pragmatischen Rollen** der Argumente morphologisch markieren
- ebenso nehmen bestimmte Sprachen eine **andere Abbildung von semantischen Rollen auf die Grammatischen Relationen** vor (z. B. Ergativität)

4.1. *Semantische Rolle

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- auch: **thematische Rolle**
- **Rolle von Argumenten** des Verbs **im durch den Satz ausgedrücktem Geschehen**
- unterschiedliches Rolleninventar ja nach Theorie
- Makrorollen: **Actor** (Agens, Experiencer usw.) und **Undergoer** (Patiens, Theme usw.)
- **semantische Hierarchie** (nach Simon Dik):

Agens>Patiens>Recipient>Benefaktiv>Instrument>Locative>Time

→ *Anordnung bzgl. Besetzung der syntaktischen Position: je weiter rechts desto unwahrscheinlicher Realisierung als Subjekt/Objekt usw.*

→ **Beispiel:**

Gestern (TIME) hat Paul (AG) Petra (BEN) im Wohnzimmer (LOC) den Computer (PAT) mit einem Schraubenzieher (INSTR) repariert, den er ihr (REC) geschenkt hatte.

Auswahl semantischer Rollen

- **AGENS:** Person oder Sache, die eine Handlung ausführt.
→ *DER KAPITÄN ändert den Kurs.*
- **PATIENS:** Person oder Sache, die von einer Handlung als Objekt betroffen ist
→ *Der Kapitän schlägt DEN MAAT.*
- **CAUSE:** Person oder Objekt, das ein Ereignis verursacht
→ *DER STURM zerriss die Segel wie Papier.*
- **BENEFAKTIV:** Nutznießer oder Geschädigter einer Handlung
→ *Der Kapitän trug DER MEERJUNGFRAU die Handtasche.*
- **EXPERIENCER (s. Übung):** Person, die psychisch oder physisch von einem Ereignis betroffen ist
→ *DER MAAT fürchtete sich vor dem Klabautermann.*
- **SOURCE (s. Übung):** Ausgangspunkt eines gerichteten Ereignisses
→ *Die Santa Maria segelte von SANSIBAR über Madeira nach Casablanca.*
- **GOAL:** Ziel eines gerichteten Ereignisses
→ *Die Santa Maria segelte von Sansibar über Madeira nach CASABLANCA.*

4.2. *Relationale Typologie

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - ***Relationale Typologie**
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- im Sprachvergleich: **Differenz in der Abbildung von semantischen Rollen auf die Grammatischen Relationen** Subjekt und Objekt
- systematische **Differenz in der Kodierung der beiden Argumente A=Agens und P=Patiens eines transitiven Satzes** im Vergleich mit der Kodierung des Hauptarguments des intransitiven Satzes (S = intransitives Subjekt)

- Im **Ergativsystem** wird das **Patiens-Argument P** des transitiven Satzes **kodiert wie das Argument S** des intransitiven Satzes: $S = P$
 → Kasus: **ABSOLUTIV**: prototypisch nullmarkiert, Zitierform
 → **verbales Agreement** (analog intransitivem Subjekt) mit **P**
- das **Agens-Argument A** ist **sondermarkiert**: $S \neq A$
 → Kasus: **ERGATIV**
- In **Akkusativsystemen** gilt: $S = A$, $S \neq P$
 → **Sondermarkierung Patiens-Argument durch Akkusativ**
 → **verbales Agreement** (analog intransitivem Subjekt) mit **A**

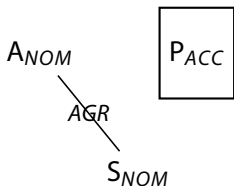


Abbildung: Akkusativisches Muster (Box = Sondermarkierung)

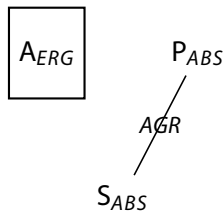


Abbildung: Ergativisches Muster (Box = Sondermarkierung)

Dyirbal (intransitiver Satz, dependent-marking):

ɲuma banaga-nyu
father.ABS:S return-NONFUT
Father returned.

Dyirbal (transitiver Satz, dependent-marking):

yabu ɲuma-ɲgu bura-n
mother.ABS:P father-ERG:A see-NONFUT
Father saw mother.

Akkusativsystem	CASE	AGR
S = Subjekt intrans.	NOM	+ / A
A = Agens trans.	NOM	+ / A
P = Patiens trans.	ACC	- / B
	$S = A \neq P$	$S = A \neq P$
Ergativsystem	CASE	AGR
S = Subjekt intrans.	ABS	+ / A
A = Agens trans.	ERG	- / B
P = Patiens trans.	ABS	+ / A
	$S = P \neq A$	$S = P \neq A$

- 'Aktiv'- oder 'Split-S'-System
- **Differenzierung** beim intransitiven Verb **nach semantischer Rolle** (S_A vs. S_P)
- vgl. Deutsch: *mich friert* (mich = inaktiv; Kodierung wie Patiens: S_P)
- Aktivsprache: **systematische Kodierung der semantischen Rolle**

Aktivsystem	CASE	AGR
S_a = Subjekt intrans. (Agens)	ACT	A
S_p = Subjekt intrans. (Patiens)	INACT	B
A = Agens trans.	ACT	A
P = Patiens trans.	INACT	B
	$S_A = A; S_P = P$	$S_A = A; S_P = P$

Georgisch (S_A = intransitiv mit Agensargument):

k'ac-ma išira

man-ACT scream:AOR.3.SG

Der Mann schrie.

Georgisch (S_P = intransitiv mit Patiensargument):

k'ac-i mok'vda

man-INACT die:AOR.3.SG

Der Mann starb.

Georgisch (A, P = transitiv):

k'ac-ma k'al-i mok'la.

man-ACT woman-INACT kill.3.SG

Der Mann tötete die Frau.

Guaraní	Agens:	Patiens:
transitiv:	<i>a-gwerú aína</i> A.1.SG-bring jetzt 'ich bringe (sie) jetzt'	<i>še-rerahá</i> P.1.SG-carry.off '(es) trägt mich dahin'
intransitiv:	<i>a-xá</i> A.1.SG-go 'ich gehe'	<i>šé-rasí</i> P.1.SG-sick 'ich bin krank'

- Kodierung der Agens- bzw. Patiens-Funktion des **empathischen Aktanten**

4.3. *Diathesen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - ****Diathesen***
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

Diathese

- syntaktische Operation der **Manipulation der Abbildung semantischer Rollen auf Grammatische Relationen**
- funktional-kognitive Interpretation als **Umstellung** bzgl. *figure-ground*-Schema
→ *Vordergrund/Hintergrund in Bühnenmodell*

z. B. Passivierung

- **Promotion** von Argument mit semantischer Rolle **Patiens** (Objekt) in **Subjektposition**
- **Demotion** von **agensartigem** Argument (Subjekt) in **adverbiale Position**

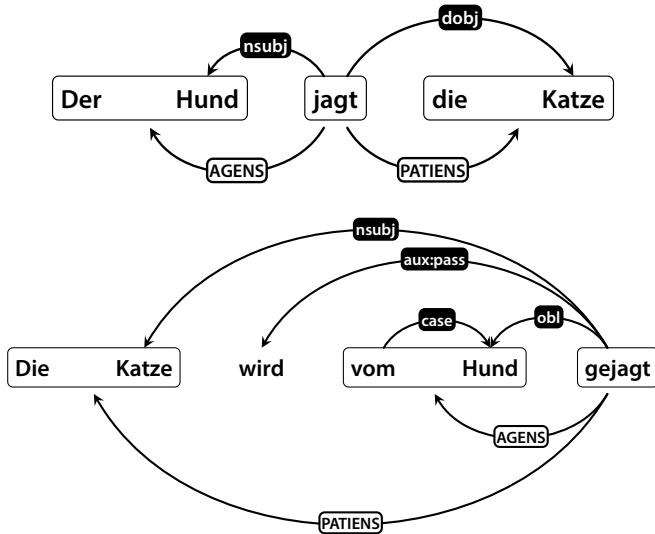


Abbildung: Aktiv- und Passivsatz Deutsch

- **Aktiv-Passiv-Unterscheidung** als **Flexionskategorie** des Verbs
- **Aktiv** ist die **unmarkierte Diathese**: Abbildung Agens auf Subjekt im transitiven Satz
- die **Passiv-Operation** bildet dagegen das **Patiens-Argument auf das Subjekt** ab: **Beförderung zum Subjekt (Promotion)**
- das **Agens-Argument** wird in die **Adverbialfunktion** herabgestuft (**Demotion**; rutscht auf der Hierarchie Grammatischer Relationen nach unten)
- **Valenzreduktion**: statt 2 (Subjekt+Objekt) nur noch 1 Kernargument (Subjekt)
- **Demotion wichtiger als Promotion**, da im Deutschen auch **intransitive Sätze passivierbar** sind (kein Objekt, das promoviert werden kann):
es wurde getanzt

- **Rezipientenpassiv:** Promotion Recipient-Argument von indirekter-Objekt-Position in Subjektposition
du (IO > S) bekommst etwas (O) geschenkt
- Valenzerhöhende Diathese **Kausativ:** Angabe einer verursachenden Instanz für Sachverhalt; im Deutschen periphrastische Konstruktion mit *lassen*:
er (S) kochte Tee (O) > er (CAUSEE) ließ ihn (S) Tee (O) kochen
- Valenzerhöhende Diathese **Applikativ:** Beförderung niederrangigen Arguments in Objektposition:
Heu (O) auf den Wagen (ADV) laden > den Wagen (ADV>O) mit Heu (O>ADV) beladen
- Valenzerhaltende Diathesen **Reflexiv/Medium:** Diathese zwischen Aktiv und Passiv, Patiens und Agens haben gleiche Bezug oder sind gleich prominent:
Er wäscht das Auto > er wäscht sich; Das Buch liest sich leicht.

4.4. *Pragmatische Rolle

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - ***Pragmatische Rolle**
 - **Topikalisierung und Fokussierung*

- **Funktion linguistischer Einheiten in Abhängigkeit vom Äußerungskontext**
- viele theoretische Ansätze und sich überschneidende Begrifflichkeiten
- **Topik = Satzgegenstand** (worüber etwas ausgesagt wird)
- **Fokus = Informationsschwerpunkt**, die neue Information über Satzgegenstand
- **je nach Kontext** kann ein Satz mit gleicher semantisch-logischer Struktur **unterschiedliche Äußerungsbedeutung** haben (eine andere Topik-Fokus-Struktur)
- **Kenntlichmachung der Topik-Fokus-Struktur eines Satzes** über verschiedene **syntaktische Operationen** wie Linksversetzung oder Cleftsätze

Topic-prominente Sprache markiert in Flexionsmorphologie primär die pragmatische Rolle:

Japanisch:

haha wa ko-no hon o kat-te kure-ta

Mutter TOP dies-ADJ Buch AKK kauf-GER geb-PRT

Mutter hat mir dieses Buch gekauft.

Japanisch:

ko-no hon wa haha ga kat-te kure-ta

dies-ADJ Buch TOP Mutter NOM kauf-GER geb-PRT

Dieses Buch hat Mutter mir gekauft.

4.5. *Topikalisierung und Fokussierung

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*

- **Wort- und Satzgliedstellung**
- Tendenz, vorne zu stehen = **Linksversetzung** (*left dislocation*, 'Herausstellung'): *Was Max (TOP) betrifft, so hat er seinen Schlüssel vergessen.*
- aber auch **Rechtsversetzung** möglich (nachgestellt): *Der ist ganz schön lang, dieser Zug (TOP)!*
- **Topikmarker** (wie im Japanischen, s. o.)
- häufig **pronominal** oder **durch Nullform** realisiert

- **Wort- und Satzgliedstellung**
- **Spaltsatz** (Cleft): *Es ist Max (FOC), der seinen Schlüssel vergessen hat!*
- **Fokusmarker**
- emphatischer **Satzakzent**, nach Akzent abfallende **Intonation**
- meist **lexikalisch realisiert**, da unbekannte Information
→ *Aussage neuer Information (Fokus) über bekannten Gegenstand (Topik)*

Was gestern betrifft, was ist da passiert?

→ *Gestern (TOP,ADV) [hat der Hund die Katze gejagt] (FOC,SATZ).*

Was den Hund betrifft, was hat er gestern gemacht?

→ *Er (TOP, SUBJ) [hat gestern die Katze gejagt] (FOC,SATZ).*

Was den Hund betrifft, wen hat er gestern gejagt?

→ *[Die Katze] (FOC,OBJ) hat er (TOP,SUBJ) gestern gejagt.*

Was die Katze betrifft, wer hat sie gestern gejagt?

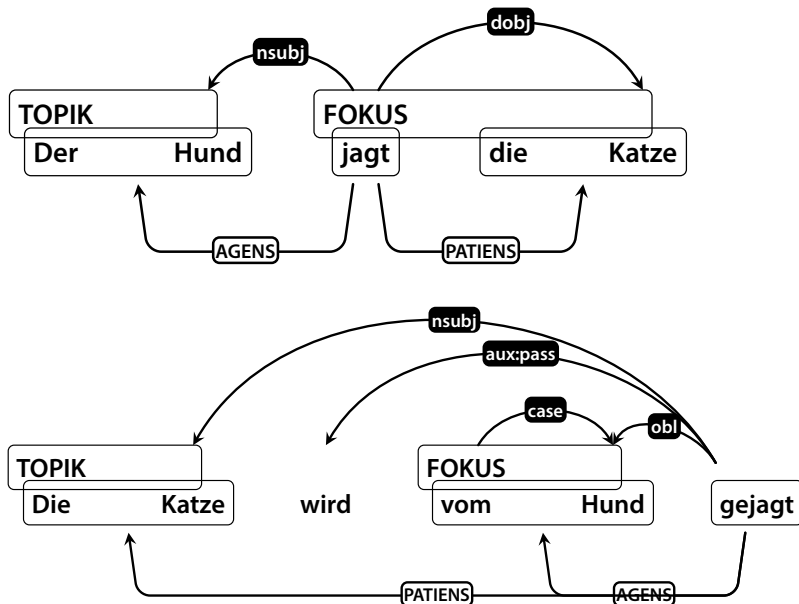
→ *[Der Hund] (FOC,SUBJ) hat sie (TOP,OBJ) gestern gejagt.*

- **Passivierung** als Mittel, das **Topik in die Subjektposition** zu bringen (bevorzugte Topikposition im Deutschen):

Was die Katze betrifft, von wem wurde sie gestern gejagt?

→ *Sie (TOP, SUBJ) wurde gestern vom Hund (FOC,ADV) gejagt.*

Passivierung als Mittel der Topikalisierung/Fokussierung



Rückblick auf heutige Themen

- 1 Sprachliche Ausdrucksmittel syntaktischer Funktionen
- 2 Grammatische Merkmale
 - Flexionskategorien
 - Kasus und Agreement als Marker Gramm. Relationen
- 3 Merkmalsstrukturen
 - Formale Repräsentation grammatischer Kategorien
 - Merkmalsstrukturen im NLTK
- 4 **Varianz syntaktischer Kodierung*
 - **Semantische Rolle*
 - **Relationale Typologie*
 - **Diathesen*
 - **Pragmatische Rolle*
 - **Topikalisierung und Fokussierung*