

Syntax natürlicher Sprachen

8: Grammatische Merkmale und Merkmalstrukturen

A. Wisiorek

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

12.12.2023

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

- Subsumption und Unifikation
- Bedingungen

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 **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*

Splitting atomarer Kategorien

- z.B. Numerus-Kongruenz NP:
- NP gesplittet in SgNP und PlNP:
 - SgN, PlN, SgDET, PlDET, SgNP, PlNP
 - SgNP \rightarrow SgDET SgN, PlNP \rightarrow 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

- 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#)
- **Merkmalstrukturen** (zur Beschreibung komplexer Objekte, z. B. grammatischer Merkmale)

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 Typhierarchie und getypte Merkmalstrukturen

- Subsumption und Unifikation
- 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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - 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 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 mit Subjekt, modelliert als komplexes Merkmal, d.h. mit anderen Merkmalen als Wertausprägungen

2.2. Kodierung syntaktischer Funktionen

- 1 Motivation für Feature-Modellierung
- 2 Grammaticale 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
 - Subsumption und Unifikation
 - 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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - Bedingungen

- **syntaktisch relevanter Teil der Morphologie (Morphosyntax)**
- **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*
- an Stelle von struktureller Kodierung über lineare Anordnung
→ *Wortstellung, z. B. Subjekt vor Objekt*

- 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)

- 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**

- **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

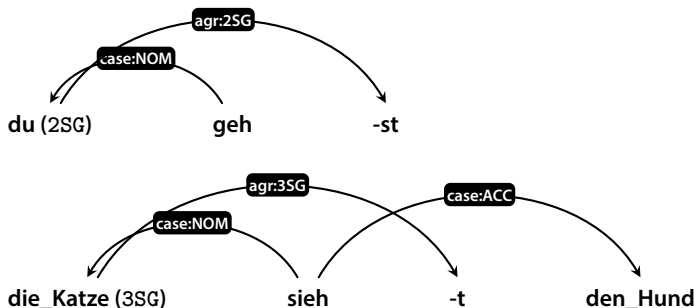
- auch Voice/Diathese: Kategorisierung der **Abbildung von semantischen Rollen auf die Grammatischen Relationen**

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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - Bedingungen

Markierung zentraler syntaktischer Satzgliedfunktionen

- 1 **Kasus:** *Markierung der syntaktischen Funktion* eines verbalen Dependents *am Dependenten* (*dependent-marking*)
- 2 **Agreement:** *Markierung der syntaktischen Funktion* eines verbalen Dependents *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**

- 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	← AGR →	Person
Numerus	← AGR →	Numerus
Genus		Tempus
		Modus
Case	←	

- 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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - 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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - Bedingungen

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

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

- **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 **suggestiert**
→ 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**:

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

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

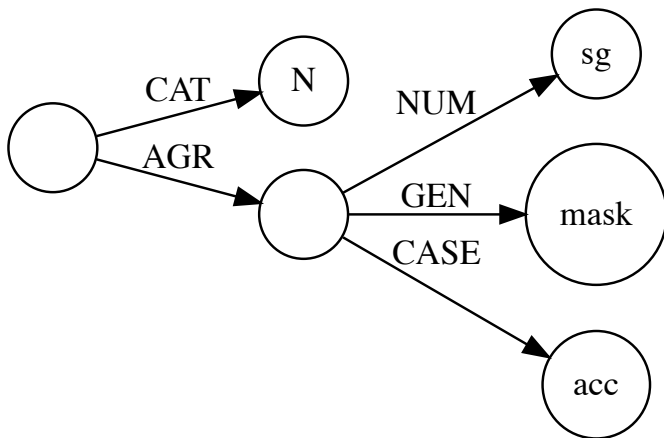
- Merkmale innerhalb einer Merkmalstruktur können **Beschreibungen für die gleiche linguistische Einheit** sein (**koreferent** sein)
- durch **Forderung nach Koreferenz von Merkmalen** von durch PSG-Regeln festgelegte Konstituenten einer syntaktischen Kategorie (untereinander oder mit Merkmalen der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

- neben atomaren Werten (SG, +) können auch **Merkmalstrukturen als Werte in einer Merkmalstruktur** 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: $\langle N \text{ AGR CASE} \rangle = ACC$)

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



Lexikoneinträge mit komplexem AGR-Merkmal

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 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
 - Subsumption und Unifikation
 - Bedingungen

- bloßer **Ersatz** von atomaren Kategoriensymbolen in PSG-Regeln **durch Merkmalstrukturen** 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 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):**

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

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

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

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

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

<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)

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 Typhierarchie und getypte Merkmalstrukturen
 - Subsumption und Unifikation
 - 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):

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

- **information ordering relation** (ähnlich Teilmengenbeziehung)
- Zwei Merkmalstrukturen stehen in **Subsumption**, wenn eine Struktur **alle Merkmale der anderen enthält**.
- **Ergebnis einer Subsumption:**
 - **Ist erfüllt**, wenn die übergeordnete (allgemeinere) Struktur **alle Merkmal-Wert-Paare der untergeordneten Struktur** enthält.
 - d.h. alle Informationen der übergeordneten (allgemeineren) Struktur sind in der untergeordneten (spezifischeren) enthalten
 - **Existiert 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: $\begin{bmatrix} A & 1 \\ B & 2 \end{bmatrix} \not\sqsubseteq \begin{bmatrix} A & 1 \\ C & 3 \end{bmatrix}$

- 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
 - Subsumption und Unifikation
 - 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 .

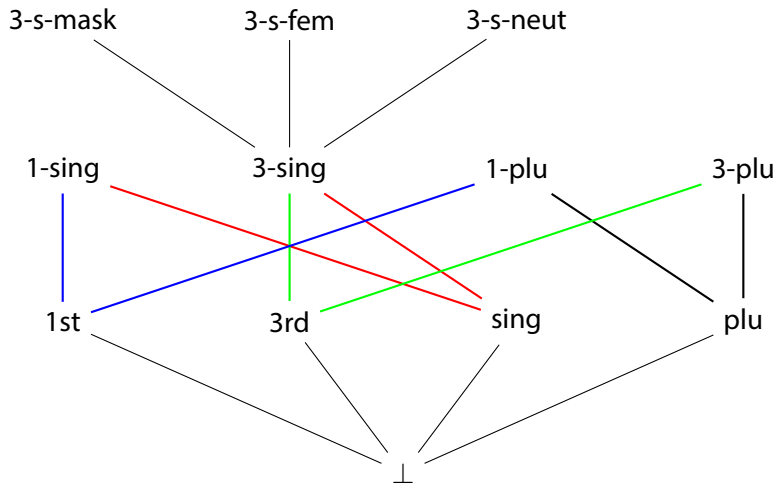
Wenn für $A, B \in \mathbf{Type}$ gilt, dass $A \sqsubseteq B$, dann

- *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 .

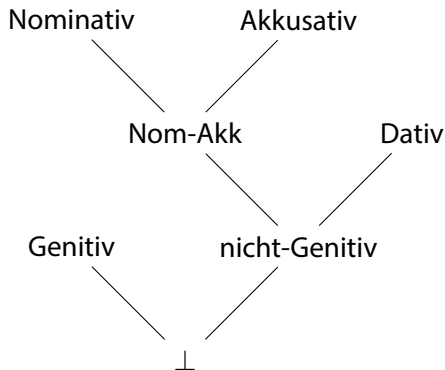
Eigenschaften von Type , \sqsubseteq

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

Beispiel: Typhierarchie



Noch ein Beispiel: 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
- $\theta: Q \rightarrow \mathbf{Type}$: totale Typisierungsfunktion
- $\delta: \mathbf{Feat} \times Q \rightarrow Q$: partielle Merkmal-Wert-Funktion

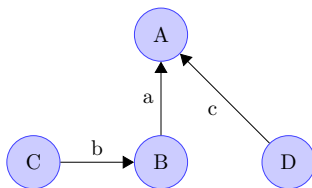
Sei \mathcal{F} die Menge aller Merkmalstrukturen.

$$\left[\begin{array}{cc} \text{CAT} & N \\ \text{AGR} & \left[\begin{array}{cc} \text{CASE} & \text{Nom} \\ \text{NUM} & \text{Pl} \end{array} \right] \end{array} \right]$$

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:

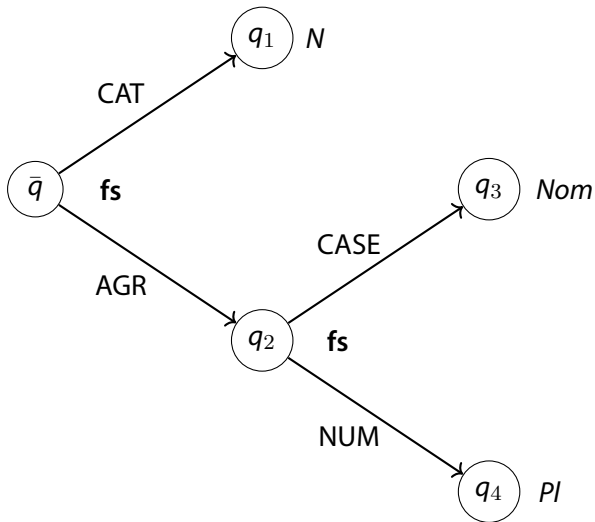
- $V := Q$
- $E := \{(q_1, q_2) \mid \exists f. \delta(f, q_1) = q_2\}$
- $L_V := \mathbf{Type}; l_V := \theta$
- $L_E := \mathbf{Feat}; l_E(q_1, q_2) := \{f \mid \delta(f, q_1) = q_2\}$

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 $\mathbf{Var} = \mathbb{N}$ benutzt.
- Es gibt aber auch andere Möglichkeiten;
z. B. im NLTK: ASCII-Identifer ($?x, ?y, \dots$)

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	SBJ	1
	OBJ	2
SEM	AGT	1
	PAT	2

ORTH	<i>folgt</i>	
SYN	SBJ	1 <i>Hund</i>
	OBJ	2 <i>Katze</i>
SEM	AGT	1
	PAT	2

4.1. Subsumption 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 Typhierarchie und getypte Merkmalstrukturen
 - **Subsumption und Unifikation**
 - 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 $f \in \mathbf{Feat}$,
für die $\delta(f, q)$ definiert ist

Beispiel

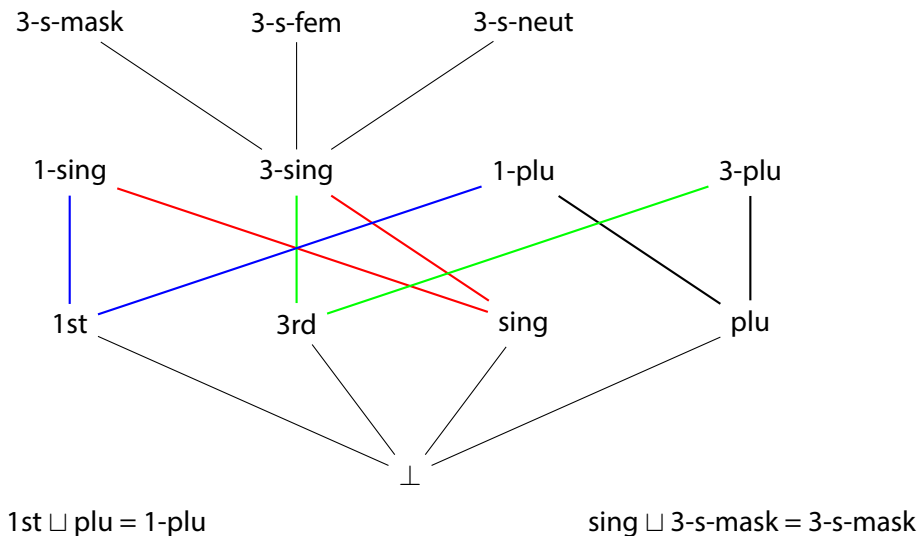
$$\bar{q} \begin{bmatrix} q_1 \text{ CAT} & N \end{bmatrix} \sqsubseteq \bar{q}' \begin{bmatrix} q'_1 \text{ CAT} & N \\ q'_2 \text{ GEN} & \text{mask} \end{bmatrix}$$

$$h(\bar{q}) = \bar{q}' \quad h(q_1) = q'_1$$

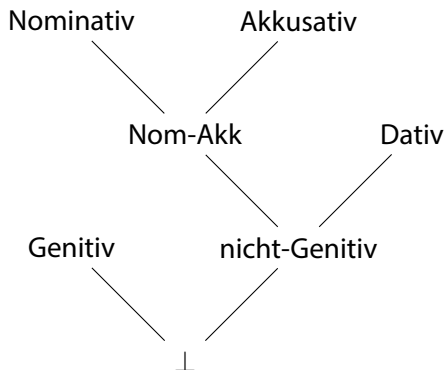
Unifikation (\sqcup) für Typen

- Das Ergebnis der Unifikation zweier Typen $A, B \in \mathbf{Type}$ ist ihre kleinste obere Schranke in \mathbf{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: Typunifikation



Noch ein Beispiel: 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 Unifiziere 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

$$q_1 \left[\begin{array}{cc} q_2 \text{ CAT} & N \\ q_3 \text{ AGR} & \left[\begin{array}{cc} q_4 \text{ NUM} & Sg \\ q_5 \text{ CAS} & \textit{nicht-Gen} \end{array} \right] \end{array} \right]$$

$$q_6 \left[\begin{array}{cc} q_7 \text{ ORTH} & \textit{Hund} \\ q_8 \text{ AGR} & \left[\begin{array}{cc} q_9 \text{ NUM} & Sg \\ q_{10} \text{ CAS} & \textit{Nom} \end{array} \right] \end{array} \right]$$

1 Identifikation korrespondierender Knoten

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

Beispiel: (Formale) Unifikation

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

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).

4.2. 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
 - Subsumption und Unifikation
 - **Bedingungen**

Pfade

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

Beispiele

- AGR-NUM
- SYN-SBJ-AGR-NUM
- ORTH
- ε (der leere Pfad)

Definition (Beschreibung Desc)

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

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

Beispiel

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

Erfülltheit

Die Erfülltheitsrelation \models^α 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 $x \in \mathbf{Var}$, $F \models^\alpha x \iff \alpha(x) = F$
- $F \models^\alpha d \wedge e \iff F \models^\alpha d \text{ und } F \models^\alpha e$

Sei F eine Merkmalstruktur.

$$F = \left[\begin{array}{ll} \text{CAT} & \boxed{1}N \\ \text{POS} & \boxed{2} \\ \text{AGR} & \left[\begin{array}{ll} \text{NUM} & Sg \\ \text{CAS} & \textit{Nominativ} \end{array} \right] \end{array} \right]$$

$$\alpha(\boxed{1}) = \alpha(\boxed{2})$$

Welche Beschreibungen aus Desc erfüllt F ?

- $F \models^{\alpha} N$?
- $F \models^{\alpha} \text{CAT} : N$?
- $F \models^{\alpha} \text{AGR-CAS} : \textit{nicht-Genitiv}$?

Nein!

Ja!

Ja!

Denn: *nicht-Genitiv* \sqsubseteq *Nominativ*

- $F \models^{\alpha} \text{POS} : N$?

Ja!

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 $MGSat(1) = [\underline{1}]$
- $MGSat(d \wedge e) = MGSat(d) \sqcup MGSat(e)$

Grammatikregel mit Constraint

$NP[CAS=?y] \rightarrow DET[GEN=?x, CAS=?y] \ N[GEN=?x]$

Bedingungen als Beschreibungen

- $type : NP \wedge CAS : \boxed{2}$
- $type : DET \wedge GEN : \boxed{1} \wedge CAS : \boxed{2}$
- $type : N \wedge GEN : \boxed{1}$

Bedingungen als Merkmalstrukturen

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