University of Bielefeld

Spezielle Themen der KI

Multimodalität:
Verarbeitung multimodaler Perzepte am Beispiel

(nach [Latoschik, 2001])

University of Bielefeld

Multimodale Systeme

Thematisch historische Einordnung

- **Bolt, 1980 (Put-That-There):** Statische Szene und statische Richtungserkennung. Getrieben durch Einzelworterkennung.
- **Böhm et al.,1992 und 1994 (Given):** Symbolische Gesten als Steuerungszeichen in der Virtuellen Realität.
- **Weimar & Ganapathy, 1992:** Kurvenmodifikation durch Sprache und Gestik.
- **Sparrell & Koons, 1994; Koons et al., 1993 (ICONIC):** Manipulation einer graphischen Szene. Dereferenzierung von Objekten und Richtungen. Einführung der Konzepte *gestlet* und *iconic mapping*.
- **Cavazza et al., 1995:** Dereferenzieren durch Kombination von Zeigen und Spracheingabe. Einführung des Begriffs des *extended pointing* im Gegensatz zum Mausclick. Sprachgetrieben (benutzen MIT-Gestenparser).
- **Lucente et al., 1998 (VisSpace):** Dynamische Projektionsanpassung vor einem Großbildschirm. Zeigen über Kopf-Hand-Differenzvektor (benutzen MIT-Pfinder).

University of Bielefeld

Gesten

Definition 1 (Geste) Eine Änderung der Körperhaltung, welche bedeutsame Signale aussendet oder genauer: Eine Geste sei eine dynamische Abfolge von äußerlich sichtbaren Konfigurationsänderungen des menschlichen Bewegungsapparates, welche einer Kommunikation dient.

Definition 2 (Postur) Eine Postur sei eine durch eine gesonderte Ruhephase ausgezeichnete Körperkonfiguration in der dynamischen Ausführung einer Geste.

Definition 3 (Mimik) Mimik sei die Untermenge der Gestik, welche hauptsächlich durch Tonusänderungen der Muskeln des Gesichtsbereichs im Gesichtsrelief zum Ausdruck kommt.

University of Bielefeld

Gestenklassifikation

Typ	Klassifikationsname	Charakteristika
I	1. Deiktisch	Referenzieren auf Objekt(e), Ort(e) und Richtung(en)im Raum.
II	1. Mimetisch 2. Ikonisch 3. Objektbezogen 4. Piktographisch	Die Extremitäten werden als Platzhalter benutzt, um das Verhalten eines beschriebenen Objektes oder Zustandes nachzubilden.
III	1. Physiographisch 2. Kinetographisch 3. Pantomimisch	Repräsentieren und verbildlichen das Zusammenspiel mit einem Objekt. Zeigen die Interaktion bei der Benutzung.
IV	1. Symbolisch 2. Moduseinstellend 3. Emblematisch	Haben eine eindeutige Semantik als alleinstehende Geste und verändern ggf. den Modus in welchem eine gleichzeitige verbale Äußerung interpretiert wird.
V	1. Ideographisch 2. Metaphorisch 3. Ikonisch	Veranschaulichen eine räumliche metaphorische Manifestation eines internen Zustands. Beziehen sich auf eine Interpretation.
VI	1. Beats 2. Gestikulation 3. Sprachmarkierend 4. Selbstregulierend	Geben einen Sprachrhythmus an. Betonung und gestische Expression fallen in den gleichen Takt.

Nach: (Efron, 1941), (Ekman & Friesen, 1979), (Nespoulous & Lecours, 1986), (Rimé & Schiaratura, 1991), (Wexelblat, 1994)

University of Bielefeld

Gestik

– Coverbale Gestik (Nespoulous & Lecour, 1986): expressive, paraverbale und illustrative Gesten:
  - Deiktisch: Das Zeigen auf über lexikalische Einheiten im sprachlichen Kanal geäußerte Referenzen.
  - Spatiographisch: Das Skizzieren der spatialen Konfiguration des Referenten einer lexikalischen Einheit.
  - Kinemimisch: Das Beschreiben einer durch eine lexikalische Einheit ausgedrückten Aktionen.
  - Pictomimisch: Das Beschreiben von Formeigenschaften des Referenten einer lexikalischen Einheit.

University of Bielefeld

Maschinelle Gestenerkennung

– Klassifikationsproblem: Qualitative Zuordnung bestimmter Eingabevektoren zu unterschiedlichen Klassen:
  - Einsatz von *HMM-basierten* Verfahren (s.a. Spracherkennung)
  - *Neuronale* Methoden
  - Explizite *Regelmodellierung* (z.B. über Templates)

➨ Existiert eine Gestengrammatik auf Basis diskreter Symbole einer kompositionellen Notation (z.B. HamNoSys)?

➨ Was geschieht mit quantitativen Daten (z.B. den Bewegungsattributen einer Trajektorie)?

University of Bielefeld

Gestenmerkmale

Definition 4 (Spatiotemporale Gestenexpression) Die spatiotemporale Gestenexpression sei die Gesamtheit der signifikanten Merkmale der die Geste ausführenden Extremitäten in Bezug auf: (1) die beschriebene spatiale Bahn, (2) die Beschreibungsdynamik durch die Trajektorie sowie (3) die daraus auf einer höheren Eben resultierenden Formen (Posturen) im Verlauf der Gestenäußerung.

Spatiotemporale Merkmale (zur Trennung gestischer Artikulation):

1. Aktion und Pause
2. Auslenkung aus Ruhestellung
3. Definite Posturform
4. Primitives Bewegungsprofil
5. Wiederholung
6. Interne Symmetrie
7. Externe Symmetrie
8. Externe Referenz

University of Bielefeld

Gestenverarbeitung

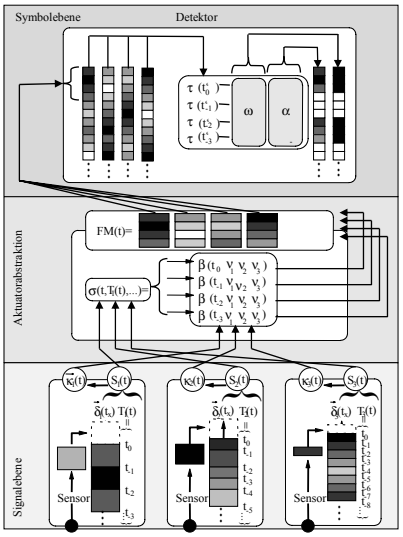
- **Gestenanalyse:** Feststellung der für das entsprechende Bewegungsmuster markanten Parameter (siehe Aktuatorvorverarbeitung)
- **Gestenerkennung:** Klassifikation der Gestenart
- ➨ *Ansatz:Mustersuche auf Basis vordefinierter Templates*

$$\begin{aligned} HOLDS?(Grasp,i) = \quad & HOLDS?((\texttt{AlignIndex} < 0.2), i_1) \\ & \wedge HOLDS?((\texttt{AlignMiddle} < 0.2), i_1) \\ & \wedge HOLDS?((\texttt{AlignPinkie} < 0.2), i_1) \\ & \wedge HOLDS?((\texttt{AlignRing} < 0.2), i_1) \end{aligned}$$

$\wedge HOLDS?((\texttt{AlignPinkie} < 0.2$
$\wedge HOLDS?((\texttt{AlignPinkie} < 0.2), i_1)$

University of Bielefeld

Gestenverarbeitung



Definition 8 (Detektor) Detektoren klassifizieren Bewegungsinformationen, indem sie Muster – logische Verknüpfungen von atomaren Testbedingungen - auf Bewegungsinformationen suchen. Als Resultat dieses Matchingvorgangs wird zu jedem Zeitpunkt t je ein Bewertungs- und ein Analysewert ermittelt. Ein Detektor ist ein 7-Tupel

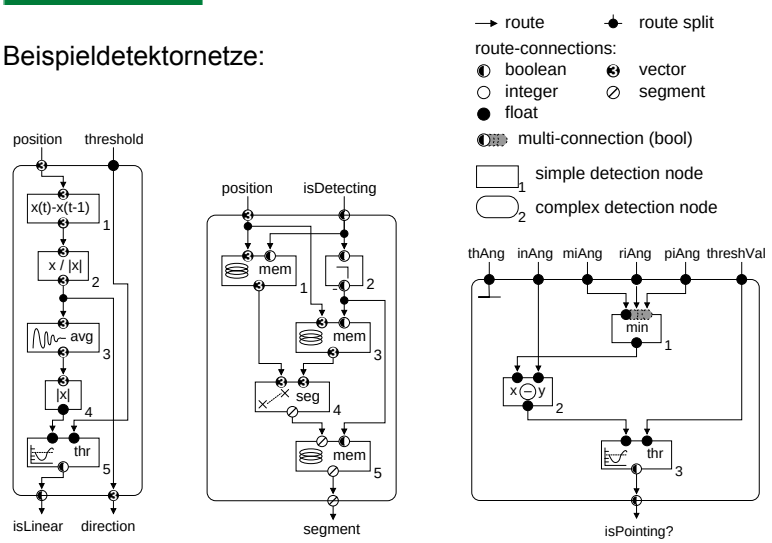
$$GD = (AI(t), \tau(t'), \omega(t'), T_{eval}(t), O(t), A(t))$$

- Singuläre Detektorauswertung an einem Aktuator mit drei Sensorquellen.
- Durch die gemeinsame Ein- und Ausgabestruktur (Attributsequenzen) ist eine hierarchische Detektorstaffelung möglich.

University of Bielefeld

Gestenverarbeitung

Beispieldetektornetze:



University of Bielefeld

Temporal Augmented Transition Network - tATN

- Verarbeitung multimodaler Eingaben basiert auf zeitlichen Informationen.
- Semantische „Zusammengehörigkeit“ wird durch Synchronität und Semantik determiniert.
- Synchronität zwischen Perzepten kann sehr unterschiedlich ausfallen (zur Modellierung siehe auch Allen‘sches Zeitkalkül).
- Gestenerkennung arbeitet im seltensten Fall „mutual exclusive“, d.h. zu einem gegebenen Zeitpunkt sind mehr als nur ein Erkennungsresultat gleichzeitig wahr.
- Herkömmliche ATN Modellierungen arbeiten
  1. sequentiell und auf linearem Inputstream und
  2. sehen nur einen aktiven Zweig vor.

➢ Um die für die Verarbeitung multimodaler Perzepte notwendigen Verfahren bereitzustellen, wurde das Konzept des temporal Augmented Transition Networks (tATN) eingeführt [Latoschik, 2001; Latoschik, 2002].
  - tATNs erhalten ein zusätzliches Zeitregister an jedem Knoten.
  - tATNs haben multiple aktive Zweige zu jedem Zeitpunkt.
  - tATNs haben ein deklaratives Beschreibungsformat:
    MIML - Multimodal Integration Markup Language

University of Bielefeld

tATN

- tATNs besitzen die folgenden Eigenschaften:
  - Verarbeitung zeitlich paralleler Eingaben.
  - Ausführung auch bei Teilparseergebnissen.
  - Vereinigung unterschiedlicher Granularitätsstufen der Perzepte.
  - Unterstützung von semantischen und zeitlichen Relationen.
  - Einfache Parametrisierung.
  - Anbindung an den Anwendungskontext.
  - Anbindung an Echtzeitsysteme.
- Im Unterschied zu ATNs gegebene Erweiterungen:
  - Ein zusätzliches Zeitregister an jedem Knoten.
  - Multiple aktive Zweige zu jedem Zeitpunkt.
  - Ein deklaratives Beschreibungsformat:
    MIML - Multimodal Integration Markup Language

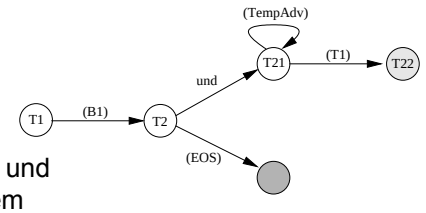
University of Bielefeld

tATN Parse-algorithmus

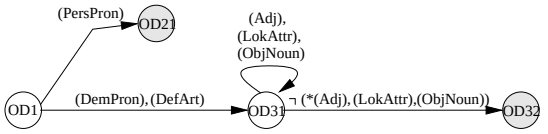
```
set s := startzustand;    ;;Initialisiere state auf Startzustand
for every newframe do     ;;Framesynchrone Ausführung
    lexqueue:update();    ;;Neue lexikalische Einträge in die Queue
    actuators:update();   ;;Aktuatoren (und damit Detektoren) updaten
    set EOP := false;     ;;EndOfParse initialisieren
    do until EOP          ;;Wiederhole solange es etwas zu parsen gibt
     set EOP := true;
     set newstate := false;    ;;Flag für vorzeitigen Abbruch
     ;;Zuerst alle constraints mit lexikalischem Eintrag testen
     for every arc in s:arcs and haslex(arc:constraints) do
       if arc:constraint do    ;;constraints erfüllt?
         set s := arc:tostate; ;;Traversiere den arc, advance state
         set arc:activate();   ;;Aktiviere Auswertefunktionen
         set newstate = true;
         set EOP = false;
         exit for            ;;Vorzeitiger Abbruch
       done
     done
     if newstate = false do    ;;Ist vorzeitiger Abbruch?
       ;;Jetzt alle übrigen constraints testen
       for every arc in s:arcs and haslex(arc:constraints) do
         if arc:constraints do
           set s := arc:tostate;
           set s:activate();
           exit for          ;;Vorzeitiger Abbruch
         done
       done
    done
    done              ;;Ende von do until EOP
done                  ;;Ende der Bearbeitung für diesen Frame
```


University of Bielefeld

tATN



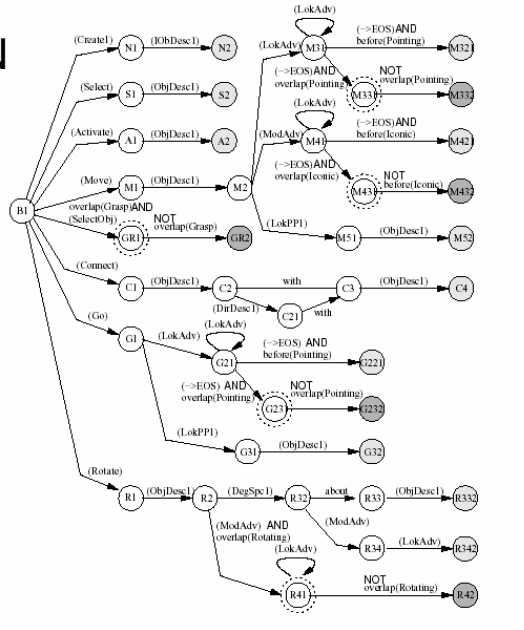
Beispiel aus dem SGIM- und Virtuelle Werkstatt System



University of Bielefeld

tATN

Beispiel aus dem SGIM und Virtuelle Werkstatt System



University of Bielefeld

tATN - MIML

```
<definition start="userInstruction">. . . </definition>

<extern apiCommand="apiRotateObjByHand-On"/>

<wordtype function="defArticle">
    <word>the, that, this</word>
</wordtype>

<gesturetype function="rotating" type="prBHistory">
    <history name="hist-rotating"/>
    <field name="axis"/>
    <field name="degree"/>
</gesturetype>
```

University of Bielefeld

tATN - MIML

```
<requirement name="rotateObject" function="userInstruction">
    <frame>
        <slot name="object" type="singleslot"/>
        <slot name="degree" type="singleslot" default="30"/>
        <slot name="axis" type="singleslot" default="0 0 1"/>
        <slot name="rotcenter" type="singleslot" default="@object" />
    </frame>
    <description>
        <temporalrelation type="sequential">
                <speech>
                        <function name="rotateAction"/>
                </speech>
                <requires>
                        <function name="objectDescription"/>
                        <fill-slot source="identifier" target="object"/>
                </requires>
```


University of Bielefeld

tATN - MIML

```
<select>
    <choice>
    ...
    </choice>
    <choice>
       <temporalrelation type="overlap">
          <speech>
             <function name="modalAdverb"/>
          </speech>
          <gesture>
             <function name="rotating"/>
             <exec-on-start>
                <apiCommand name="rotateObjectByHand-On"/>
             </exec-on-start>
             <exec-on-end>
                <apiCommand name="rotateObjectByHand-Off"/>
             </exec-on-end>
          </gesture>
       </temporalrelation>
    </choice>
</select>
...
```

University of Bielefeld

Probleme mit Annotation

- Annotation erfolgt für beliebige Arten von kontextfreien Grammatiken wie PSG oder Rekursive Übergangsnetzwerke (RTNs).
- Typische RTNs sind die erweiterten Übergangsnetzwerkgrammatiken (ATNs – Augmented Transition Networks, [Woods, 1970])

➢ Probleme:
  - Grammatik nicht mehr rein deklarativ.
  - Annotation durch „procedural attachements“ fördert die Verwendung dieser für die Grammatik selbst.
  - Registerbelegungen sollen unabhängig sein. In der Realität sind sie es häufig nicht (-> lexikalische Ambiguität, s. Beispiel vorher).

University of Bielefeld

Merkmalsstrukturen

Merkmalsstrukturen (MS):

- Verbreiteter Formalismus in der Computerlinguistik.
- Ausführlich untersucht und beschrieben, (zumindest im Kern) auf MS basierende Ansätze:
  - Categorical Unification Grammar [Uszkoreit, 1986]
  - Functional Unification Grammar (FUG) [Kay, 1979]
  - Generalized Phrase Structure Grammar GPSG [Gazda et al., 1985]
  - Head-driven Phrase Structure Grammar HPSG [Pollard und Sag, 1987]
  - Lexical Functional Grammar LFG [Kaplan und Bresnan, 1982]
  - …
- MS transzendieren die Grenzen zwischen linguistischen Beschreibungsebenen.
- MS bestehen aus implizit konjunktiv verknüpften Merkmalen und Merkmalswerten.
- Trennung zwischen der Sprache, in der Constraints über MS-Mengen spezifiziert werden, und den MS selbst [Johnson, 1990] ist sinnvoll.
- Unterspezifikation: Nicht alle Merkmale müssen explizit angegeben werden. Nichtauftreten bedeutet „keine Aussage über den entsprechenden Wert“.
- Zentrale Operation auf den MS ist die **Unifikation**, daher wird auch von **Unifikationsgrammatiken** gesprochen.

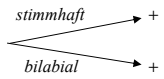
University of Bielefeld

Merkmalsstrukturen

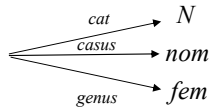
1. Binäre Merkmale:
   - bezeichnen Zutreffen eines Merkmals (+|-). Bsp. Für die Phoneme /b/ und /m/.
   - Früher auch für Semantikbeschreibung genutzt [Katz und Fodor, 1963].
2. Einfache Merkmale:
   - Wertbelegung eines Merkmals aus einer Menge von Werte.
   - Keine formale Möglichkeit der Wertrestriktion vorgesehen (aber Möglichkeit der Typisierung, s.u.)
3. Komplexe Merkmale:
   - Werte eines Merkmals sind selber Merkmalsstrukturen.
   - Einfache Merkmale äquivalent zu komplexen binären Merkmalen (mit endlicher Menge von Werten).
   - Zusammenfassung unter verschiedenen Aspekten zusammengehöriger Merkmale.

- Pfad: Reihe der Merkmale zu einem Wert: Bsp: *syn|agr|casus* ist *nom*.

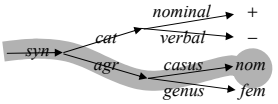
$$\begin{bmatrix} stimmhaft: & + \\ bilabial: & + \end{bmatrix}$$



$$\begin{bmatrix} cat: & N \\ casus: & nom \\ genus: & fem \end{bmatrix}$$



$$\begin{bmatrix} syn: & \begin{bmatrix} cat: & \begin{bmatrix} nominal: & + \\ verbal: & - \end{bmatrix} \\ agr: & \begin{bmatrix} casus: & nom \\ genus: & fem \end{bmatrix} \end{bmatrix} \end{bmatrix}$$



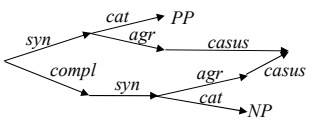
University of Bielefeld

Merkmalsstrukturen

Koreferenz und Pfadäquivalenz:

- Wertidentität für verschiedene Merkmale, etwa Casusübereinstimmung zwischen PP und NP:

$$\begin{bmatrix} syn: & \begin{bmatrix} cat: & PP \\ agr: & [casus: \langle 1 \rangle] \end{bmatrix} \\ compl: & \begin{bmatrix} syn: & \begin{bmatrix} cat: & NP \\ agr: & [casus \ \langle 1 \rangle] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$



- Da atomare Werte nicht unterspezifiziert sind –sie beschreiben genau ein Element- sind die folgenden MS schwach äquivalent. Die Koreferenz (bzw. Pfadäquivalenz) wäre in diesem Fall redundant*.

$$\begin{bmatrix} syn: & \begin{bmatrix} cat: & PP \\ agr: & [casus: \langle 1 \rangle dativ] \end{bmatrix} \\ compl: & \begin{bmatrix} syn: & \begin{bmatrix} cat: & NP \\ agr: & [casus \ \langle 1 \rangle] \end{bmatrix} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} syn: & \begin{bmatrix} cat: & PP \\ agr: & [casus: \langle 1 \rangle dativ] \end{bmatrix} \\ compl: & \begin{bmatrix} syn: & \begin{bmatrix} cat: & NP \\ agr: & [casus \ \langle 1 \rangle dativ] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

- Ist diese Äquivalenz auch für komplexe Werte gegeben?

➢ Nein! Komplexe Werte beschreiben Mengen von Elementen.

* Diese Ansicht wird nicht von allen Autoren geteilt (s. [Shieber, 1986])

University of Bielefeld

Merkmalsstrukturen

Disjunktive Merkmale:

- Modellierung der Ambiguität linguistischer Strukturen entweder durch Unterspezifikation (wenn etwa die modellierte Theorie alle Belegungen für dieses atomare Merkmal zulässt:

$$sie := \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ agr: & \begin{bmatrix} pers: & 3 \\ num: & sg \end{bmatrix} \end{bmatrix} \end{bmatrix} \text{ oder } \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ agr: & \begin{bmatrix} pers: & 3 \\ num: & pl \end{bmatrix} \end{bmatrix} \end{bmatrix} \xrightarrow{\text{Untersp.}} \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ agr: & [pers: & 3] \end{bmatrix} \end{bmatrix}$$

oder durch Disjunktion (wenn nur eine eingeschränkte Menge als Belegung für atomares Merkmal zulässig ist):

$$sie := \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ agr: & \begin{bmatrix} pers: & 3 \\ casus: & nom \end{bmatrix} \end{bmatrix} \end{bmatrix} \text{ oder } \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ agr: & \begin{bmatrix} pers: & 3 \\ casus: & akk \end{bmatrix} \end{bmatrix} \end{bmatrix} \xrightarrow{\text{atomare Disjunktion}} \begin{bmatrix} syn: & \begin{bmatrix} cat: & PersPron \\ casus: & \{nom, akk\} \end{bmatrix} \end{bmatrix}$$

oder durch Disjunktion (wenn nur eine eingeschränkte Menge als Belegung für komplexes Merkmal zulässig ist):

$$\begin{bmatrix} syn: & \begin{bmatrix} cat: & \begin{bmatrix} nominal: & + \\ verbal: & - \end{bmatrix} \\ agr: & \left\{ \begin{bmatrix} genus: & fem \\ num: & sg \end{bmatrix}; \begin{bmatrix} casus: & \{nom; gen; akk\} \\ genus: & fem \\ num: & pl \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

University of Bielefeld

Merkmalsstrukturen

- Jede Merkmalsstruktur beschreibt eine potentiell unendliche Menge von Elementen anhand festgelegter Bedingungen für diese Elemente.
- Daher ergibt sich eine partielle Ordnung über den Merkmalsstrukturen.
- Eine MS ist spezieller, wenn sie für eine Merkmal einen Wert enthält, der für die andere unterspezifiziert ist.

Subsumption: Partielle Ordnungsrelation welche allgemeinere MS vor speziellere MS ordnet:

$$[cat: [nominal: +]] \sqsubseteq \begin{bmatrix} cat: & \begin{bmatrix} nominal: & + \\ verbal: & - \end{bmatrix} \\ agr: & [casus: nom] \end{bmatrix} \sqsubseteq \begin{bmatrix} cat: & \begin{bmatrix} nominal: & + \\ verbal: & - \end{bmatrix} \\ agr: & \begin{bmatrix} casus: & nom \\ genus: & fem \end{bmatrix} \end{bmatrix}$$

Erzeugung eines Verbandes durch folgende Elemente:

- Topelement $\top$ bezeichnet die allgemeinste Merkmalsstruktur, die keine Information enthält.
- Bottomelement $\bot$ bezeichnet die inkonsistente Merkmalsstruktur.
- ➢ Alle MS werden von $\top$ subsumiert, keine von $\bot$.

Unifikation: Die auf der Ordnungsrelation operierende konjunktive Verknüpfung der Merkmale zweier MS -> Schnittmenge der durch die beiden Eingabestrukturen denotierten Elementmengen.