

The image features two thick black L-shaped brackets. One is positioned on the left side, with its vertical bar extending downwards and its horizontal bar extending to the right. The other is on the right side, with its vertical bar extending upwards and its horizontal bar extending to the left. These brackets frame the central text.

ATTRIBUTE, WERTE,
UNIFIKATION

MODELLE UND MERKMALSSTRUKTUREN



Attribut-Wert-Spezifikationen

- **Attribut-Wert-Spezifikation:** Formalismus, der die **sprachlichen Einheiten^a** mit **Merkmalen^b** versieht, die wiederum **Werte^c** erhalten
 - a) z.B. *die Wörter*
 - b) z.B. *>Kategorie<*
 - c) z.B. *>Verb<*

- Finden insbesondere in den modernen **Grammatiktheorien** Anwendung:
 - *Generalized Phrase Structure Grammar (GPSG, Gazdar et al. 1985)*
 - *Head- Driven Phrase Structure Grammar (HPSG, Pollard & Sag 1994)*
 - *Lexical Functional Grammar (LFG, Bresnan 2001)*

Linguistische Modelle

- Basis von Sprachverarbeitungssystemen
- Mathematisch formalisierte Konstrukte
- Nicht Teil der Grammatik: Sie bilden den Beschreibungsgegenstand, die Objektsprache
- Auf Instanzen der Objektsprache wird mit dem Terminus **>Merkmalsstruktur<** referiert
- Die Metasprache ist die Beschreibung der Modelle
- Das Ziel der Sprachverarbeitung ist eine (partielle) Beschreibung der Merkmalsstrukturen (innerhalb der Domäne des linguistischen Modells)

Formalismus (Beschreibungen der
Merkmalsstrukturen)

Merkmalsstrukturen
(Modell)

Linguistische Objekte
(Realität)

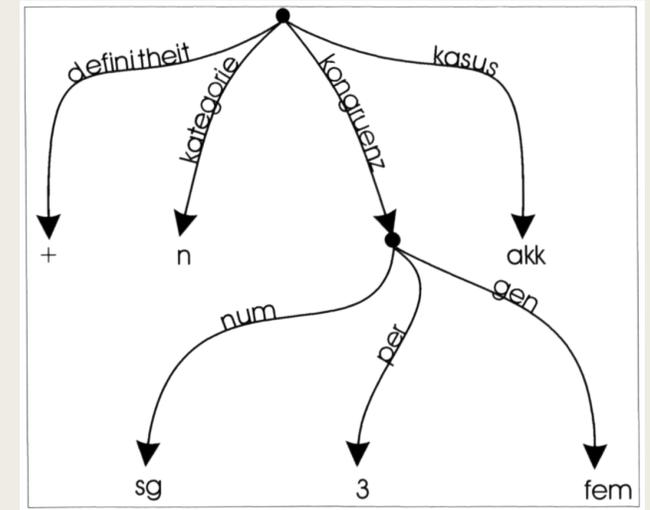
Ebenenmodell zur formalen
Beschreibung sprachlicher Einheiten

BESCHREIBUNGEN VON MERKMALSSTRUKTUREN

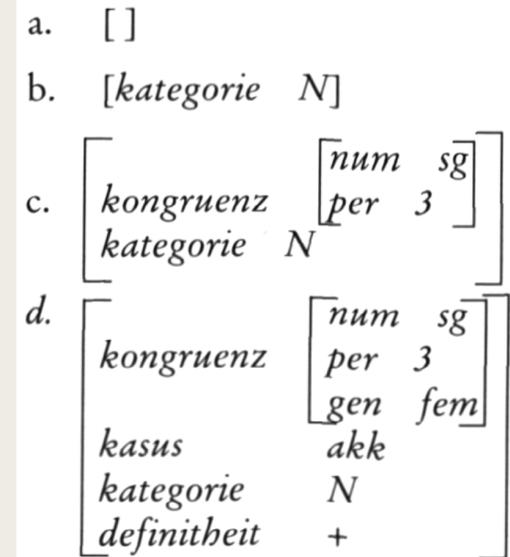


Beschreibung von Merkmalsstrukturen

- Merkmalsstrukturen können mittels unterschiedlicher Notationen repräsentiert werden
 - *Mathematisch definierte gerichtete Graphen*
 - *Diagramme von Attribut-Wert-Matrizen (AWMs)*
- Merkmalsstrukturen können wiederum als Beschreibungsinstrumentarium für Merkmalsstrukturen benutzt werden
 - *Um Ambiguitäten zu vermeiden, wird eine strikte Notationskonvention eingeführt* →



Merkmalsstruktur, die weibliche Eigennamen im Akkusativ repräsentiert



Merkmalsstrukturen werden mittels AWMs beschrieben

<i>objekt</i>	<table border="1"> <tr> <td><i>kasus</i></td> <td><i>akk</i></td> </tr> <tr> <td><i>belebt</i></td> <td>-</td> </tr> </table>	<i>kasus</i>	<i>akk</i>	<i>belebt</i>	-
<i>kasus</i>	<i>akk</i>				
<i>belebt</i>	-				
<i>kongruenz</i>	<table border="1"> <tr> <td><i>person</i></td> <td>3</td> </tr> <tr> <td><i>numerusus</i></td> <td><i>sg</i></td> </tr> </table>	<i>person</i>	3	<i>numerusus</i>	<i>sg</i>
<i>person</i>	3				
<i>numerusus</i>	<i>sg</i>				

(unvollständige) Aussagen über die Wortform *isst*

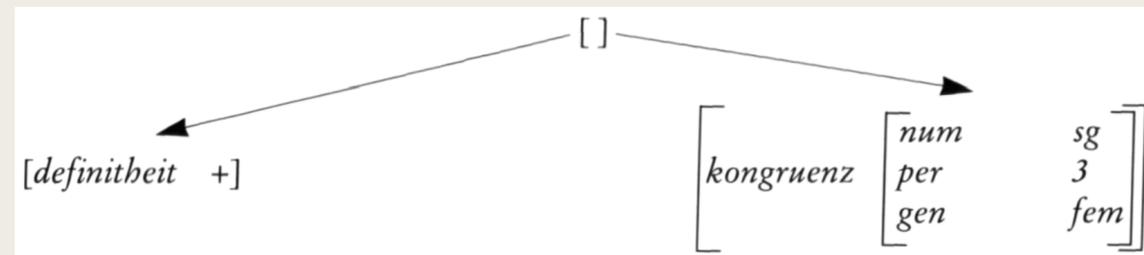
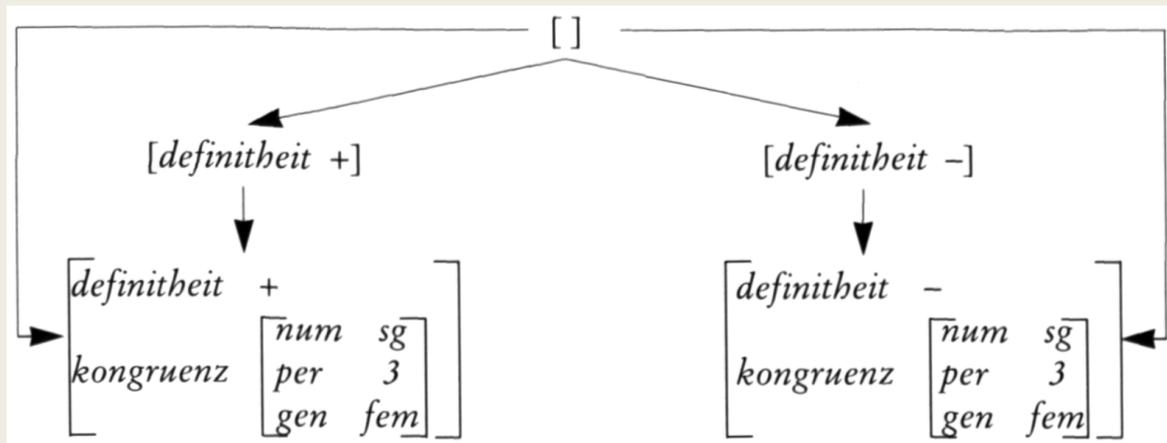
<i>subjekt</i>	<table border="1"> <tr> <td><i>kasus</i></td> <td><i>nom</i></td> </tr> <tr> <td><i>form</i></td> <td>1</td> </tr> <tr> <td><i>belebt</i></td> <td>+</td> </tr> </table>	<i>kasus</i>	<i>nom</i>	<i>form</i>	1	<i>belebt</i>	+
<i>kasus</i>	<i>nom</i>						
<i>form</i>	1						
<i>belebt</i>	+						
<i>form</i>	1						
	<table border="1"> <tr> <td><i>person</i></td> <td>3</td> </tr> <tr> <td><i>numerusus</i></td> <td><i>sg</i></td> </tr> </table>	<i>person</i>	3	<i>numerusus</i>	<i>sg</i>		
<i>person</i>	3						
<i>numerusus</i>	<i>sg</i>						

Aufbau von Merkmalsstrukturbeschreibungen

- Eine AWM besitzt endlich viele Merkmale
- Jedes Merkmal besitzt einen Wert
- Alle Merkmale besitzen Gültigkeit, d.h. es werden nur Merkmale beschrieben, die in der zu beschreibenden Merkmalsstruktur vorhanden sind
- Die Merkmale sind somit implizit mit dem logischen UND verknüpft
- Eine AWM, die keine Merkmale spezifiziert, heißt leere AWM
- Alle nicht-leeren AWMs spezifizieren Merkmale
- Die Werte der Merkmale können atomar oder komplex sein
- Die Reihenfolge, in der die Merkmale aufgeführt sind, ist irrelevant
- Verschiedene Attribute können auf denselben Wert zugreifen

Subsumtion

- **Subsumtion:** Eine AWM D subsumiert eine AWM D' ($D \sqsubseteq D'$) genau dann, wenn alle Informationen, die in D vorhanden sind auch in D' enthalten sind. Dies bedeutet, dass alle von D' beschriebenen Attribut-Wert-Strukturen auch von D beschrieben werden.
- Die leere AWM subsumiert jede Merkmalsstruktur



Subsumtionsbeziehung zwischen AWMs. Die hier nicht in einer Subsumtionsbeziehung stehenden Einheiten der Sprachbeschreibung sind zueinander inkompatibel.

(Fehlende) Subsumtionsbeziehung von zueinander kompatiblen AWMs

<i>kategorie</i>	<i>n</i>							
<i>definitheit</i>	+							
<i>kongruenz</i>	<table style="border-collapse: collapse; border: 1px solid black;"> <tr> <td style="padding: 5px;"><i>num</i></td> <td style="padding: 5px;"><i>sg</i></td> </tr> <tr> <td style="padding: 5px;"><i>per</i></td> <td style="padding: 5px;"><i>3</i></td> </tr> <tr> <td style="padding: 5px;"><i>gen</i></td> <td style="padding: 5px;"><i>fem</i></td> </tr> </table>	<i>num</i>	<i>sg</i>	<i>per</i>	<i>3</i>	<i>gen</i>	<i>fem</i>	
<i>num</i>	<i>sg</i>							
<i>per</i>	<i>3</i>							
<i>gen</i>	<i>fem</i>							

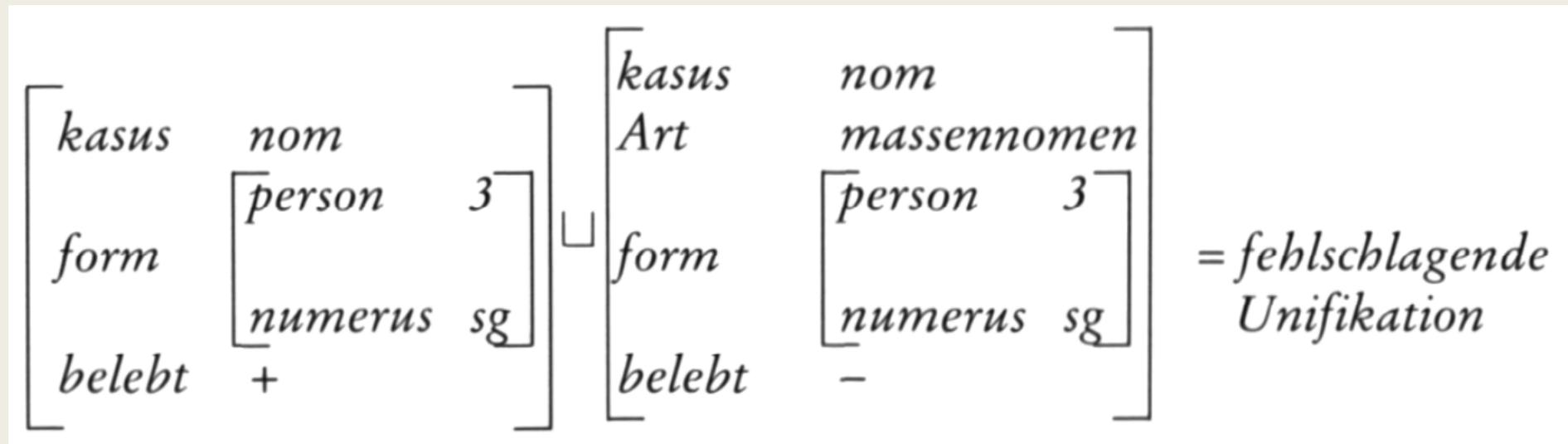
Allgemeinster Unifikator/most general unifier: Eine AWM, die genau die Informationen zweier kompatibler Merkmalsstrukturen enthält

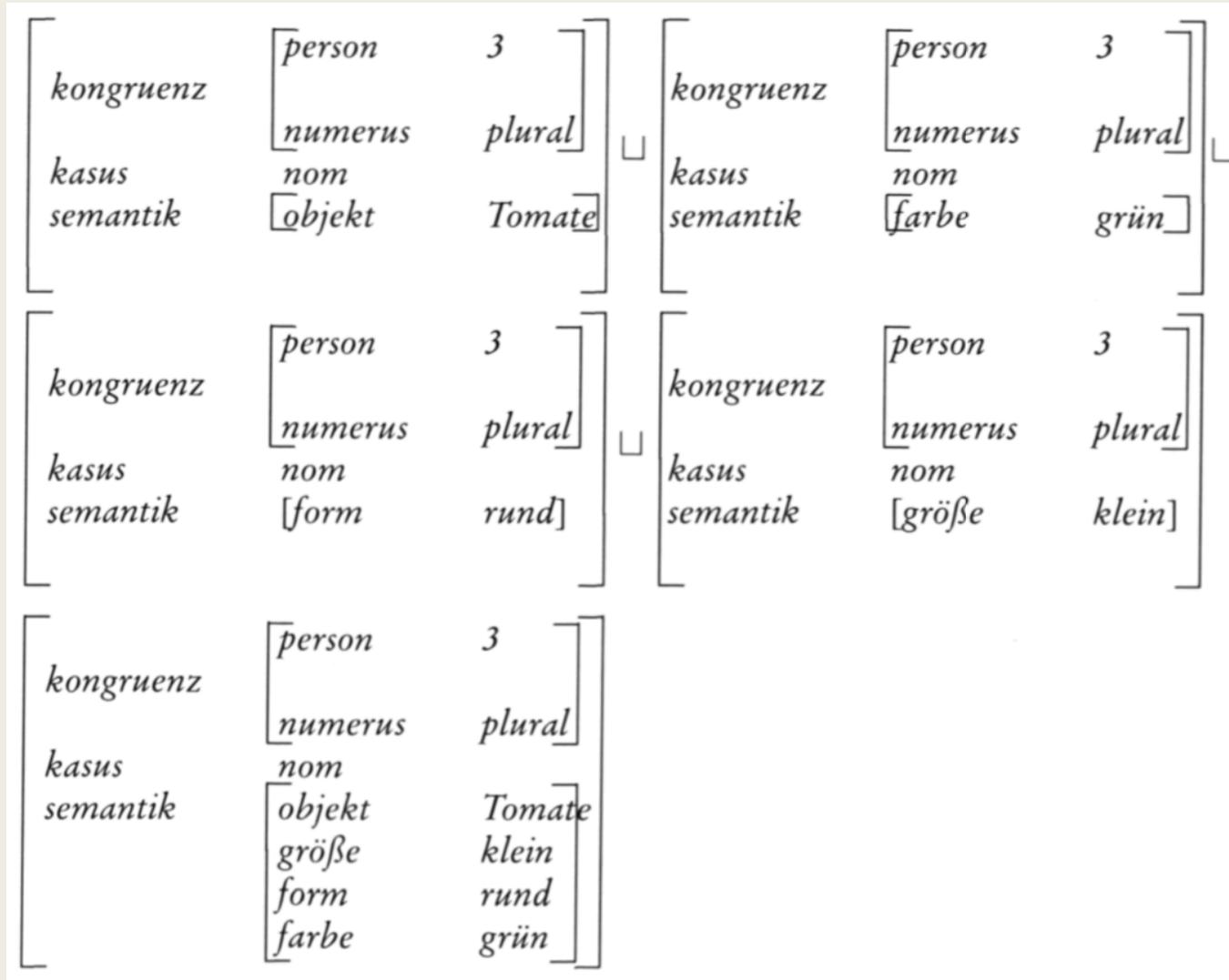
Unifikation

- Die in kompatiblen Merkmalsstrukturen enthaltenen unterschiedlichen Informationen können durch eine AWM beschrieben werden
- Das Ergebnis der Unifikation zweier AWMs D' und D'' ist die allgemeinste AWM D , die sowohl von D' als auch von D'' subsumiert wird. Wenn keine solche Struktur existiert, ist die Unifikation nicht definiert. Die Unifikation wird mittels des Symbols \sqcup dargestellt. Das Ergebnis von $D' \sqcup D''$ beschreibt alle Merkmalsstrukturen, die sowohl von D' als auch von D'' beschrieben werden.

Beispiel für eine inkompatible AWM

- Wenn wir versuchen, die Subjektspezifizierung des Verbs *isst* mit einer Beschreibung eines unbelebten Subjekts, z.B. Wasser, zu unifizieren, ergibt sich:





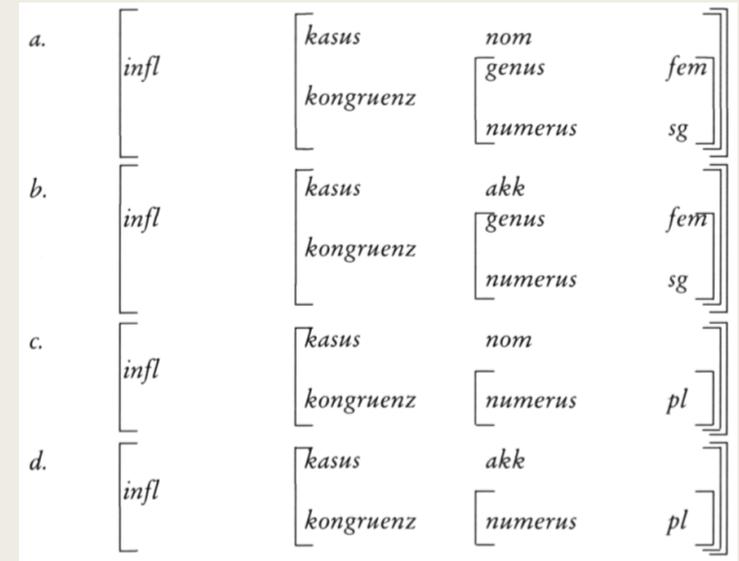
Beispiel für eine kompatible AWM

← Auf der Seite der Syntax lassen sich beliebig viele Adjektive vor das Substantiv stellen. Jedes dieser Adjektive trägt auch zur Bedeutung bei.

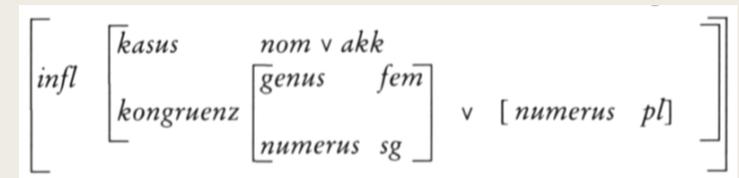
Disjunktion

- Karttunen (1984) schlägt Erweiterungen der Merkmalsstrukturbeschreibungen vor
- Eine AWM ist eine Konjunktion der Beschreibungen
- Es ist nicht möglich auszudrücken, dass eines von mehreren zur Auswahl stehenden Attribut-Wert-Paaren gelten soll
→ *Disjunktion*
- Kasper und Rounds (1986): Der Berechnungsaufwand der Unifikation disjunkter AWMs besitzt eine exponentielle Komplexität
- Schnellere Verarbeitung mit dem Rechner:
 - *Merkmalsstrukturbeschreibungen mit disjunktiv spezifizierten Werten in disjunktionsfreie AWMs umwandeln*
 - *Erweiterung des Formalismus nicht zulassen*

Um eine Merkmalsstruktur des deutschen Artikels die zu beschreiben, können die folgenden vier AWMs formuliert werden:

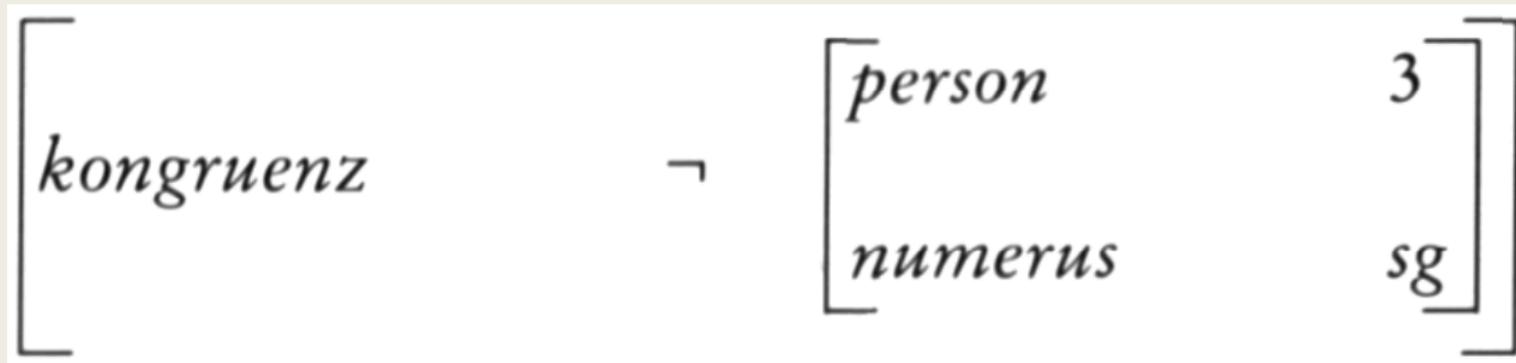


Durch Zulassung disjunkter Werte kann die in den vier AWMs enthaltene Information in einer AWM ausgedrückt werden:



Negation

- Die unflektierte Form von Verben im Englischen ist mit allen Subjekten verträglich, die nicht in der 3. Person Singular stehen (keine -s-Flexion)
→ *Negierter Wert des Kongruenzmerkmals:*



- Unter gewissen Voraussetzungen kann man eine Merkmalsstruktur mit Negation in eine ohne Negation umwandeln, die dafür Disjunktionen enthält

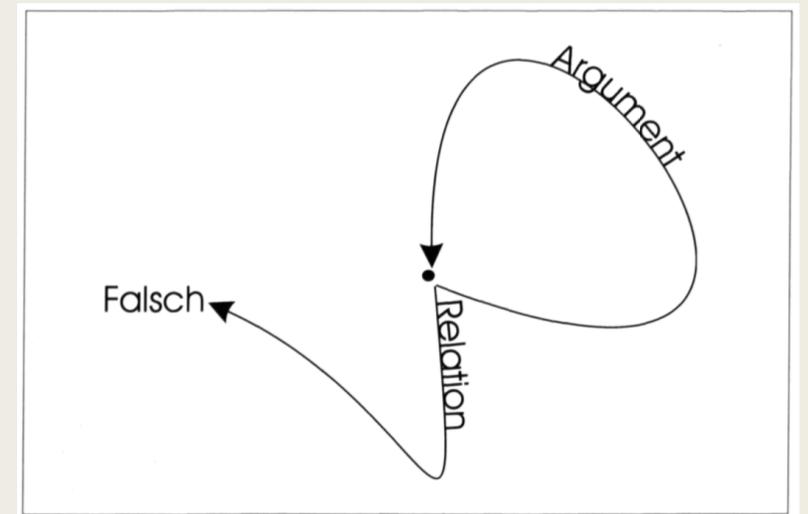
Zyklizität

- Merkmalsstrukturen, in denen Pfade von einem Merkmal auf dasselbe Merkmal zeigen, sind **zyklisch**
- Karttunen (1984): zyklische Strukturen sind nicht notwendig, da es nicht gezeigt werden kann, dass sie notwendig sind
- Carpenter (1992): in der natürlichen Sprache gibt es zyklische Strukturen:

a. Dieser Satz ist falsch.

b.

1	Relation falsch
	Argument 1



Derartige Strukturen sind sehr selten und werden deshalb von manchen Formalismen verboten.

BESCHREIBUNGEN
GETYPTER
MERKMALSSTRUKTUREN



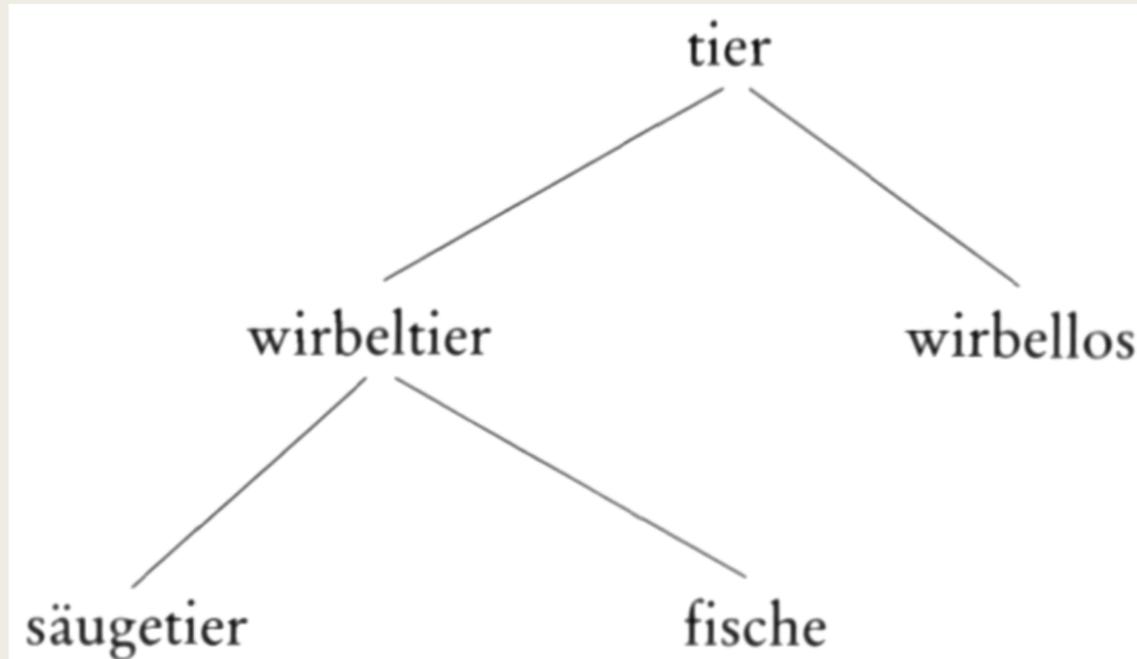
Getypte Attribut-Wert-Strukturen

- Getypte Attribut-Wert-Strukturen werden durch getypte AWMs beschrieben
- Getypte AWMs unterscheiden sich von einfachen AWMs durch die Angabe eines Typs, der einen Typen in einem Typensystem denotiert
- Der Typ einer Attribut-Wert-Struktur bestimmt, welche Merkmale sie besitzt
→ *Typen stellen Angemessenheitsbedingungen auf*
- Wenn eine Attribut-Wert-Struktur ausschließlich Merkmale besitzt, die durch ihren Typ legitimiert sind, ist diese Attribut-Wert-Struktur richtig getypt (engl. well-typed)
- Wenn alle durch ihren Typ erlaubten Merkmale einer Attribut-Wert-Struktur spezifiziert wurden, wird davon gesprochen, dass die Attribut-Wert-Struktur vollständig richtig getypt ist (engl. totally well-typed)

Das Typensystem

- Das Typensystem besteht aus einer endlichen Menge von Typen → **Type**
- Die Typen besitzen einen unterschiedlichen Grad der Allgemeinheit → Der Informationsgehalt der Typen ist unterschiedlich
- Manche Typen beschreiben Entitäten, die auch von anderen Typen beschrieben werden
 - *zwei (oder mehr) Typen besitzen denselben Informationsgehalt: alphabetische Varianten desselben Typs*
 - *einer von zwei Typen besitzt einen höheren Informationsgehalt als der andere: der mit dem höheren Informationsgehalt ist spezifischer als der Typ mit dem geringeren Informationsgehalt*
- Das Typensystem besitzt eine Ordnung, die sich nach der Spezifik der enthaltenen Typen richtet
- Die Ordnung bewirkt, dass allgemeinere Typen ihren Informationsgehalt an spezifischere Typen vererben

Das Typensystem



Klassifikation von Tieren in einer Hierarchie

Der Typ *tier* ist Supertyp seiner Subtypen *wirbeltier*, *wirbellos*, *säugetier* und *fisch*.

Die Typenmenge besteht aus:

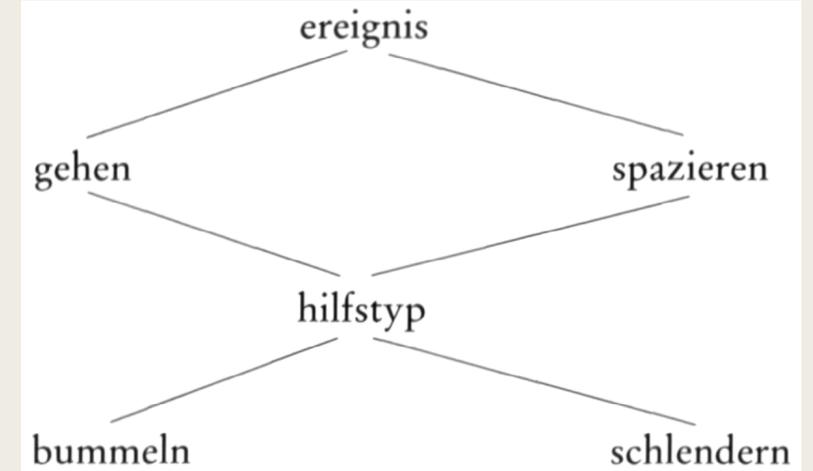
$$T = \left\{ \begin{array}{l} \text{tier} \\ \text{wirbeltier} \\ \text{wirbellos} \\ \text{säugetier} \\ \text{fisch} \end{array} \right\}$$

Es gelten die folgenden Subsumtionsbeziehungen:

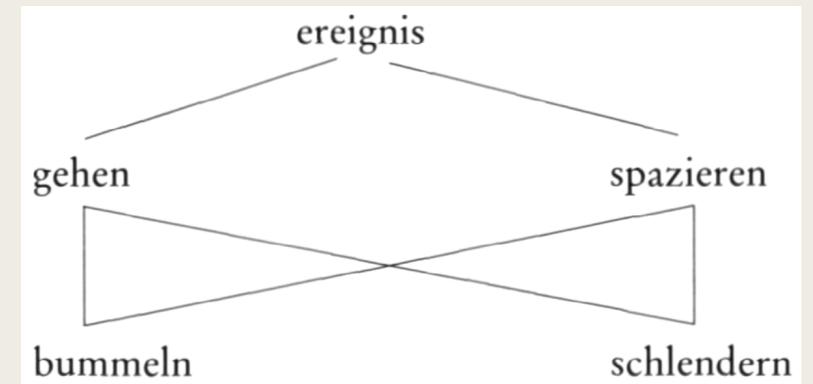
- $tier \sqsubseteq wirbeltier$,
- $tier \sqsubseteq wirbellos$,
- $tier \sqsubseteq säugetier$,
- $tier \sqsubseteq fisch$,
- $wirbeltier \sqsubseteq säugetier$ und
- $wirbeltier \sqsubseteq fisch$

Das Typensystem

- Die in den neueren Grammatiktheorien verwendeten Typensysteme müssen den folgenden Anforderungen genügen:
 1. *Das Typensystem besteht aus einer Menge von Typen*
 2. *Die Ordnung des Typensystems wird durch die Subsumtionsrelation ausgedrückt*
 3. *Zwei Typen des Typensystems, die gemeinsame Subtypen besitzen, besitzen genau einen gemeinsamen allgemeinsten Subtyp*



Akzeptable Typenhierarchie



Nicht akzeptierte Typenhierarchie

Getypte Attribut-Wert-Matrizen

- Getypte Attribut-Wert-Matrizen: Attribut-Wert-Matrizen, die getypte Attribut-Wert-Strukturen beschreiben können
- Eine getypte Attribut-Wert-Matrix besitzt endlich viele Merkmale und einen Typ
- Die Merkmale können atomar oder komplex sein
- Eine Attribut-Wert-Matrix, die keine Merkmale beschreibt und den allgemeinsten Typ denotiert, ist das Analogon zur leeren Attribut-Wert-Matrix $\sigma[]$
 - *Die Typenbezeichnung σ ist relativ zu dem definierten Typensystem zu sehen*

Operationen an getypten Attribut-Wert-Matrizen

- **Subsumtion:**

Eine AWM D_σ subsumiert eine AWM D'_τ ($D_\sigma \sqsubseteq D'_\tau$) genau dann, wenn alle Informationen, die in D_σ vorhanden sind, auch in D'_τ enthalten sind und ($\sigma \sqsubseteq \tau$) darüber hinaus gilt. Dies bedeutet, dass alle von D'_τ beschriebenen Attribut-Wert-Strukturen auch von D_σ beschrieben werden.

- **Unifikation:**

Das Ergebnis der Unifikation zweier getypter Attribut-Wert-Matrizen D'_σ und D''_τ ist die allgemeinste getypte Attribut-Wert-Matrix D'_γ , die von D'_σ und D''_τ subsumiert wird. Wenn keine solche Struktur existiert, ist die Unifikation nicht definiert.

Weitere Punkte

- Mit den vorgestellten AWMs werden alle Eigenschaften linguistischer Objekte gleichermaßen beschrieben
- Linguistische Generalisierungen innerhalb der Teiltheorien (für Phonologie, Morphologie, Syntax und Semantik) und auch innerhalb der gesamten Theorie lassen sich über die Modellierung im Typsystem gut erfassen
- Dinge, die für eine bestimmte Klasse von Objekten charakteristisch sind, werden in einer Typdefinition zusammengefasst
- Die Idee von der Universalgrammatik lässt sich mit Typen adäquat umsetzen:
 - *es gibt allgemeine Typen, die in den Typsystemen aller Sprachen benutzt werden, und*
 - *es gibt von diesen Typen abgeleitete Subtypen, die nur für bestimmte Sprachfamilien oder Einzelsprachen relevant sind*
- Andere Vorteile des Typsystems aus computerlinguistischer Sicht:
 - *das Wissen kann über die Grammatik kompakt und redundanzfrei repräsentieren werden:*
 - *Änderungen an der Spezifikation eines Typs werden durch die Vererbung automatisch auf alle Untertypen übertragen*
 - *Alle Objekte, die durch AWMs mit entsprechenden Typen beschrieben werden, haben automatisch die geänderte Spezifikation*

Quelle

- Witt, Andreas & Stefan Müller. 2009. Grundlagen für den Computereinsatz in der Linguistik: Attribute, Werte, Unifikation. In Horst M. Müller (ed.), 425 – 442. 2., überarbeitete und aktualisierte Auflage. Paderborn: Schöningh.