

HPSG

Grammatikformalisen Sommer-Semester 2010

Antske Fokkens, Hans Uszkoreit und Judith Köhne

Department of Computational Linguistics
Saarland University

22 Juni 2010

Outline

- 1 Einführung
- 2 Lexikalische Beschränkungen
- 3 Modifikation
- 4 Sprachspezifische Schemata

Outline

- 1 Einführung
- 2 Lexikalische Beschränkungen
- 3 Modifikation
- 4 Sprachspezifische Schemata

Überblick

- Zeichen: die wichtigsten Attribute
- Phrasen: wie werden sie aufgebaut?
 - Schemata (HPSG “Regeln”)
- Prinzipien:
 - Head-Feature Principle (Kopfmerkmal-Prinzip)
 - Valence Principle (Valenzprinzip, auch Subcat-Prinzip genannt)
- **Syntaktische Beschränkungen**
 - **Schemata**
 - **Lexikalische Beschränkungen**
- Semantik
 - Semantic Principle (Semantik-Prinzip)
 - Wie wird die Semantik einer Phrase aufgebaut?

Vorlesung heute

- Kurze Wiederholung Vorlesung letzte Woche
- Überblick: weitere Beschränkungen der Grammatik
- Das Valenzprinzip: Valenz genauer erklärt
- Lexikalische Beschränkungen
- Modifikation
- (Schemata: sprachspezifische Beschränkungen)

Kurze Wiederholung (1)

Typen linguistischer Objekte:

- Umfassen alle linguistischen Elemente der Grammatik (lexikalische Elemente, Kategorien, Funktionen)
- Definiert über ihre Attribute, d.h. über ihre Merkmale und deren mögliche Werttypen
- Werttypen:
 - Atomar
 - Merkmalsstruktur
 - Eine Liste von Merkmalsstrukturen $\langle \dots \rangle$
 - Eine Menge von Merkmalsstrukturen $\{ \dots \}$

Basiert auf <http://www8.informatik.uni-erlangen.de/IMMD8/Lectures/GRAMMATIKTHEORIEN/>

Kurze Wiederholung (2)

- **Typendeklaration**
 - Logische Formeln für linguistische Repräsentationen
 - Darstellung als Attribut-Wert-Matrix
 - Nicht-spezifizierte Werte sind eingeschränkt durch Typen-spezifische Merkmaldeklarationen
- **Typendefinition bestimmt...**
 - ...welche Attribute ein Typ hat
 - ...welche Werttypen diese Attribute erlauben
 - (...und welchen bestimmten Wert ein Attribut annehmen muss)

Basiert auf <http://www8.informatik.uni-erlangen.de/IMMD8/Lectures/GRAMMATIKTHEORIEN/>

Kurze Wiederholung (3)

Von einem sprachwissenschaftlichen Gesichtspunkt besteht HPSG aus:

- Lexika (validieren Basis-Wörter einer bestimmten Sprache)
- Lexikalische Regeln (validieren abgeleitete Wörter)
- Immediate-Dominance Schemata (validieren Konstituenten Strukturen)
- “Linear Precedence” Regel (beschränkt Wortstellung)
- Grammatik-Prinzipien (äussern Generalisierungen über linguistische Objekte)

Based on: <http://www.coli.uni-saarland.de/courses/syntactic-theory-09/slides/hpsg2.pdf>

Kurze Wiederholung (4)

- Äußerungen können von Zeichen modelliert werden
- Die generative Idee: eine Äußerung besteht aus einem oder mehreren Teilen
- Diese Idee wird von Phrasen modelliert: eine Phrase ist ein Zeichen, das eine oder mehrere Töchter (d.h. Teile) hat

Kurze Wiederholung (5)

Wie können wir sicherstellen, dass nur wohlgeformte Phrasen gebaut werden?



Prinzipien +
Syntaktische Eigenschaften der Wörter (und Morphologie) +
Linearpräzedenz Regel

Kurze Wiederholung (6)

Grundprinzipien in HPSG

- Immediate-Dominance-(ID) Prinzip
- Head-Feature-Prinzip (Kopf-Merkmal Prinzip)
- Valenz-Prinzip
- Semantik-Prinzip

Kurze Wiederholung (7)

ID Prinzip

Jede Phrase muss genau einem der ID-Schemata der Sprache entsprechen

- Welche Schemata zur Verfügung stehen ist sprachabhängig

Based on: <http://www.coli.uni-saarland.de/courses/syntactic-theory-09/slides/hpsg2.pdf>

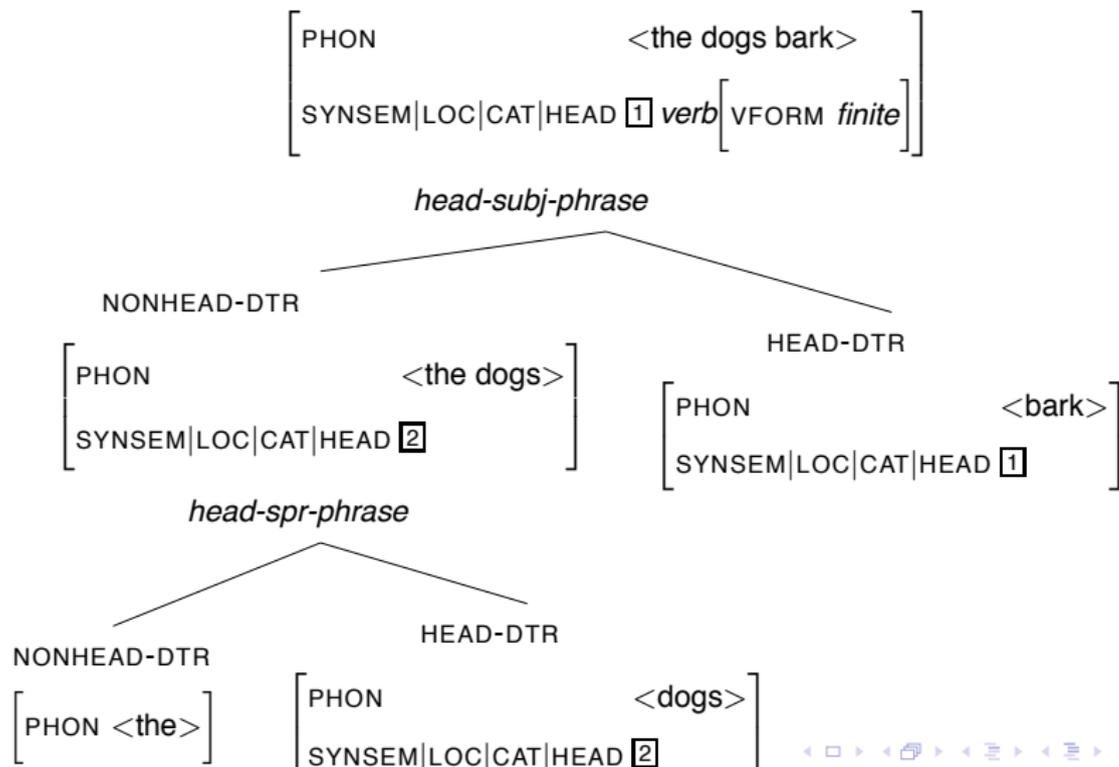
Kurze Wiederholung (8)

Kopf-Merkmal Prinzip (Head-Feature Principle)

Die Wert des HEAD-Attributs (Kopfattributs) einer *headed-phrase* ist **token-identisch** zum HEAD-Wert der Kopftöchter

$$\text{headed-phrase} \left[\begin{array}{l} \text{SYNSEM|LOC|CAT|HEAD} \quad \boxed{1} \\ \text{DTRS|HEAD-DTR|SYNSEM|LOC|CAT|HEAD} \quad \boxed{1} \end{array} \right]$$

Head-Feature Principle (Beispiel)



Schemata

- Die Definitionen der wichtigsten Schemata die wir bisher gesehen haben sind (leicht) vereinfacht
- Die Grammatik muss (mehr) beschränkt werden:
 - Die Schemata sind teilweise unterdefiniert (Kopf-Adjunkt-Schema)
 - Sie haben noch keine sprachspezifischen Beschränkungen
 - **Zusätzliche Beschränkungen kommen vom Lexikon**
- Die Semantik (Semantik Prinzip, semantische Kompositionalität) muss noch definiert werden

Valenz

- *valency* ist ein Typ, der (hier) drei Attribute hat:

SUBJ	<i>list-of-synsem</i>
COMPS	<i>list-of-synsem</i>
SPR	<i>list-of-synsem</i>

- Die subkategorisierten Argumente eines Wortes werden unter Valenz angegeben

Valenz: Beispiele

- die Valenz eines Wortes das ein Subjekt, 2 Komplemente

und keinen Spezifizierer hat:

SUBJ	<[]>
COMP	<[], []>
SPR	<>

- <> stellt die leere Liste da, <[]> ist eine Liste mit genau einem Element, <[], []> hat genau 2 Elemente.

Das Valenz und Schemata

Wie können wir sicher stellen, dass alle subkategorisierten Argumente gefunden werden, und keine überflüssigen Argumente auftauchen?

Das Valenz und Schemata

Wie können wir sicher stellen, dass alle subkategorisierten Argumente gefunden werden, und keine überflüssigen Argumente auftauchen?



Das Valenzprinzip

Die Schemata “registrieren” welches Argument gerade gefunden wurde (oder welche gefunden wurden), und welche Argumente von der Kopftochter immer noch gesucht werden

Das Valenzprinzip, teil 1

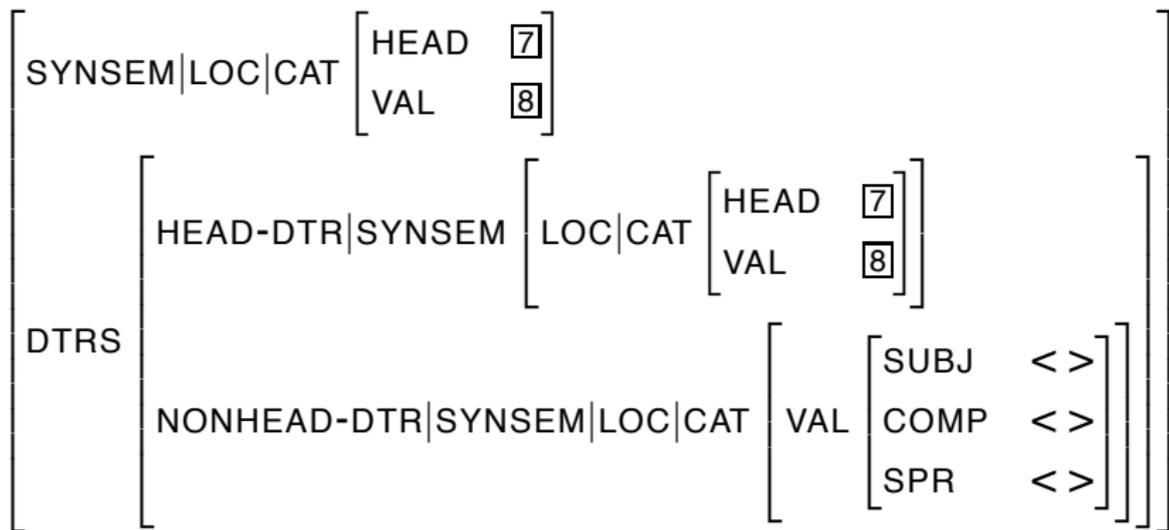
Valenzprinzip

Sofern nicht anders definiert, sind die Valenzwerte einer Phrase identisch zu den Valenzwerte ihrer Kopftochter

- Die Argumente, die eine Kopftochter braucht (und nicht eine der NON-HEAD-Töchter der Phrase sind), werden von der Mutter immer noch gebraucht:
- Eine Kopf-Adjunkt Phrase kombiniert keine Argumente mit der Kopftochter → identische Valenzwerte zwischen Kopftochter und Mutter
 - Eine Kopf-Subjekt Phrase kombiniert keine Komplemente oder Spezifizierer → identische COMPS und SPR Werte
 - Eine Kopf-Komplement Phrase kombiniert keine Subjekte oder Spezifizierer → identische SUBJ und SPR Werte
 - Eine Kopf-Spezifizierer Phrase kombiniert keine Komplemente oder Subjekte → identische COMPS und SUBJ Werte

Kopf-Adjunkt-Phrase

- *Head-Adjunct-Schema* mit HFP und Valenzprinzip:



Das Valenzprinzip (Teil 2)

- Es muss auch registriert werden, welche Argumente gerade gefunden wurden: eine Phrase die Kopftochter und Subjekt kombiniert hat, braucht kein Subjekt mehr
 - Das Argument, das gerade gefunden wurde, muss von der betroffene Valenzliste entfernt werden
 - Wenn eine Kopftochter z.B. zwei Komplemente sucht, und die Head-Comp-Phrase sie mit einer der Komplemente kombiniert, muss das andere Komplement auf die COMPS Liste der Mutter stehen
- Liste der Mutter = Liste der Kopftochter – *synsem* der Nichtkopf-Tochter
- = Liste der Kopftochter = Liste der Mutter + *synsem* der Nichtkopf-Tochter

Das Valenzprinzip (Teil 2)

Für ein Valenzattribut F gilt: in einer *head-f-phrase* ist der Wert vom F der Kopftochter die Verkettung von dem Wert von F der Phrase und die *synsem* der NON-HEAD-DTRS

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT|VAL} \left[F \quad \boxed{1} \right] \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR|SYNSEM|LOC|CAT|VAL} \left[F \quad \boxed{1} \oplus < \boxed{2} > \right] \\ \text{NON-HEAD-DTR|SYNSEM} \quad \boxed{2} \end{array} \right] \end{array} \right]$$

Outline

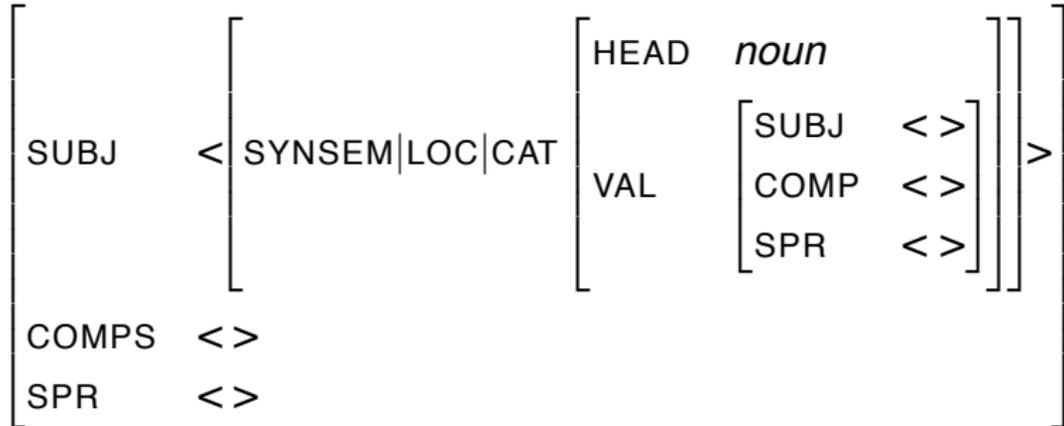
- 1 Einführung
- 2 Lexikalische Beschränkungen**
- 3 Modifikation
- 4 Sprachspezifische Schemata

Der Valenzwert von lexikalischen Einträgen

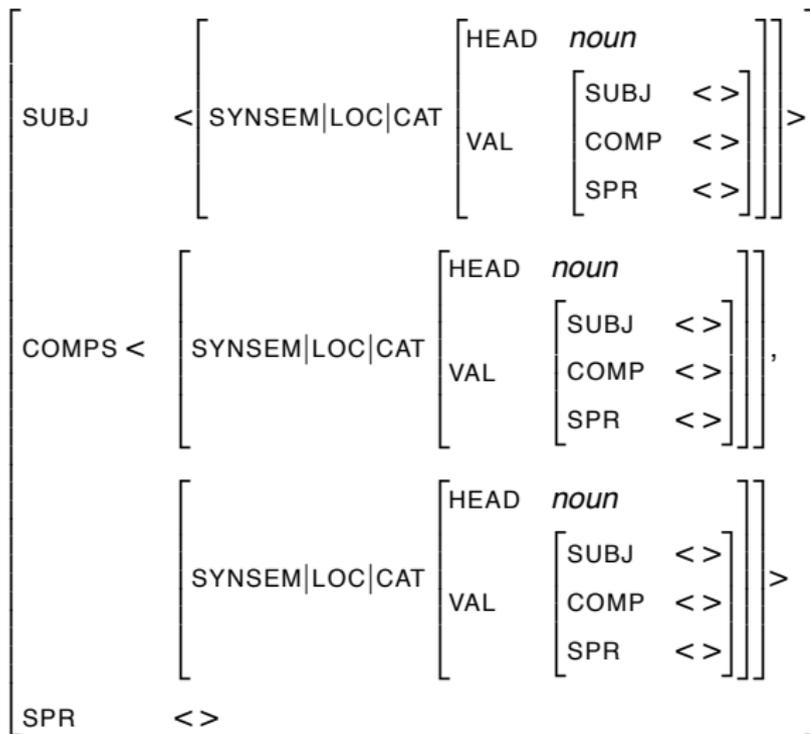
- Lexikalische Einträge spezifizieren ihre Subkategorizationseigenschaften in ihrer *valenz*
- SUBJ, COMPS und SPR haben alle Liste von *synsem* als Wert
- Weil sie die *synsem* subkategorisieren, können lexikalische Einträge die syntaktischen und semantischen Eigenschaften ihrer Argumente beeinflussen

Valenzwerte: intransitives Verb

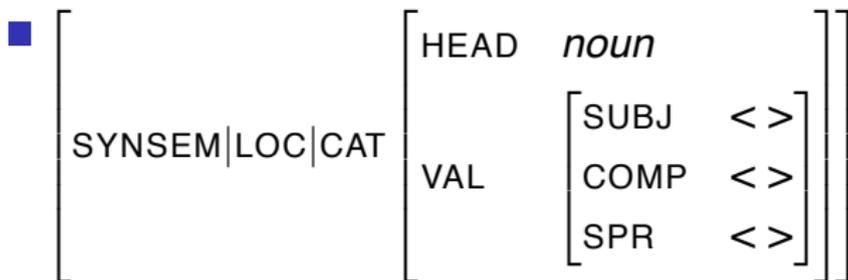
Valenz eines intransitiven Verbs:



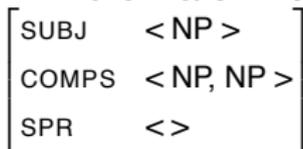
Valenzwerte: ditransitives Verb



- In den obenstehenden Beispielen sind die Argumente auf Nominal Phrasen Beschränkt: der Kopf muss *noun* sein, und alle Valenzlisten sind leer



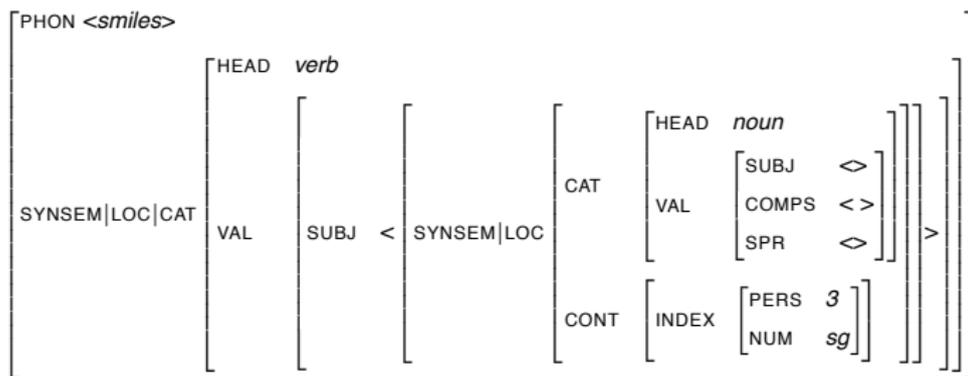
- Diese subkategorisierten *synsem* können abgekürzt als “NP” wiedergegeben werden
- z.B. die Valenz eines ditransitiven Verbs:



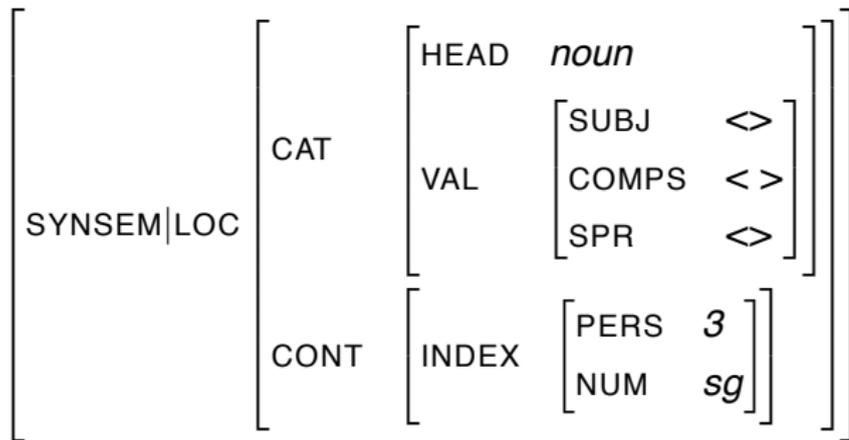
Kongruenz mit Valenz

- Kongruenz oder Übereinstimmung:
- Zum Beispiel: Verb-Subjekt Kongruenz
 - The girl smiles
 - * Girls smiles
- *smiles* kann nur mit einem 3e Person Singular Subjekt kombiniert werden
- Das Wort *smiles* enthält diese Information in dem subkategorisierten Subjekt

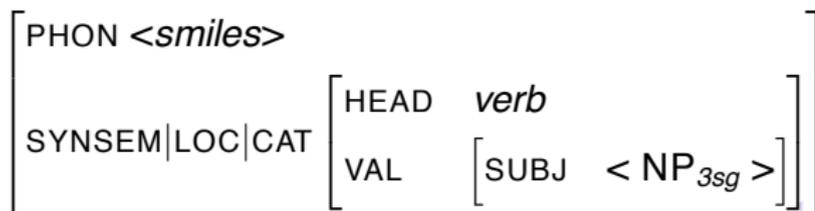
Intransitives Verb in 3e person singular



Person Numerus Kongruenz



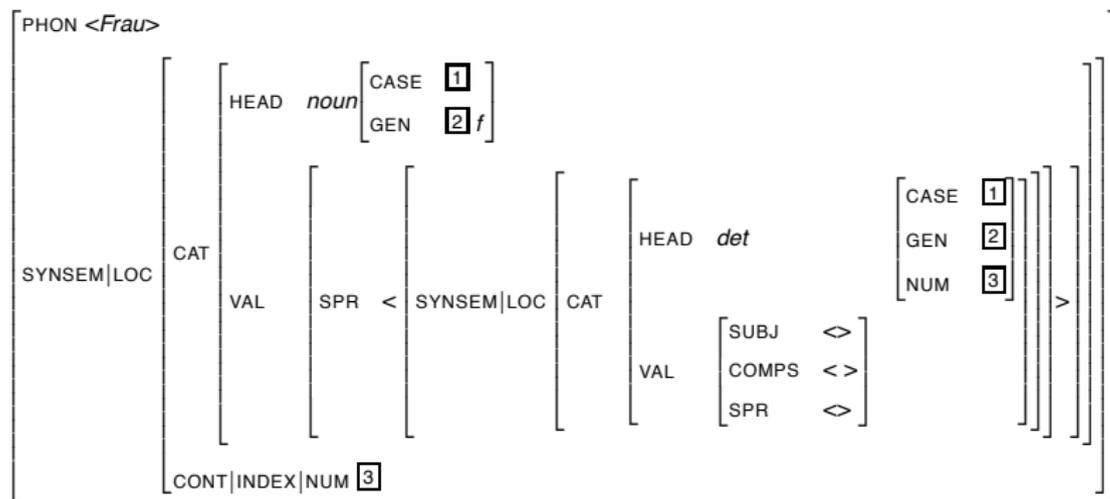
wird abgekürzt als NP_{3sg}



Nomen-Determinierer Kongruenz

- Deutsche Determinierer und Nomen stimmen überein in Kasus, Genus, und Numerus
- Die Übereinstimmung wird mittels des Valenzwerts etabliert
- CASE, NUMERUS und syntaktisches GENUS sind token-identisch zwischen Nomen und ihren Spezifizierern

Nomen



Outline

- 1 Einführung
- 2 Lexikalische Beschränkungen
- 3 Modifikation**
- 4 Sprachspezifische Schemata

Head-Adjunkt Phrasen

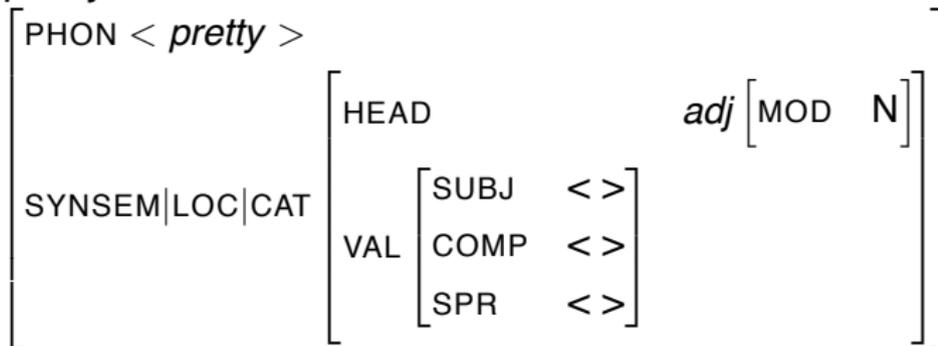
- Kopf-Argument Phrasen werden durch die Subkategorisierung beschränkt
- Das Subjekt, ein Komplement oder der Spezifizierer muss von dem Kopftochter subkategorisiert sein
- Kopf-Adjunkt Phrasen sind nach der jetzigen Definition noch unbegrenzt
- Folgende Beispiele müssen (u.a.) ausgeschlossen werden:
 - (1) * The a girl sleeps. (*a* oder *the* darf kein Modifizierer sein)
 - (2) * The beautifully girl sleeps. (*beautifully* kann keine Nomen modifizieren)
 - (3) * The girl sleeps beautiful. (*beautiful* kann keine Verben modifizieren)

Das Attribut MOD

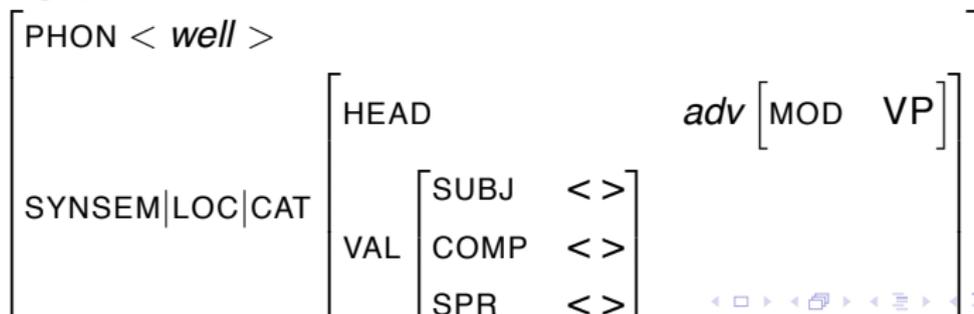
- Ob ein Wort als Adjunkt funktionieren kann, und welche Ausdrücke es modifizieren kann, hängt von seiner Wortart (part-of-speech) ab
- Diese Eigenschaften werden im Lexikon definiert
- Modifizierer haben ein Kopfmerkmal MOD
- Die Wert von MOD ist *synsem*
- MOD definiert welche Typen von *synsem* modifiziert werden können

Beispiele von Modifizierer

■ *pretty*:

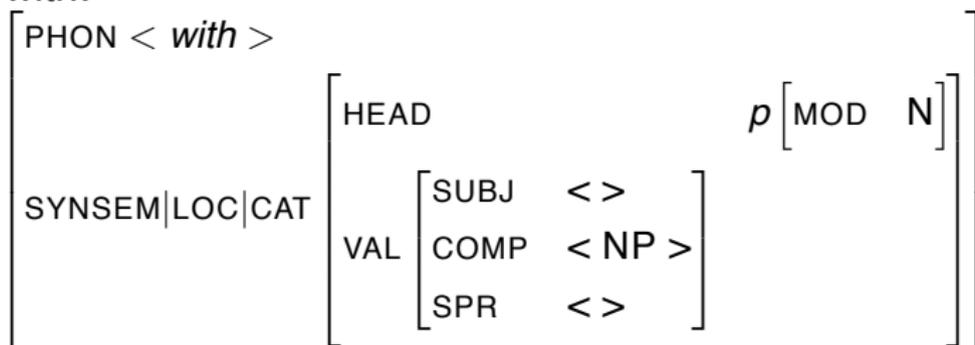


■ *well*:

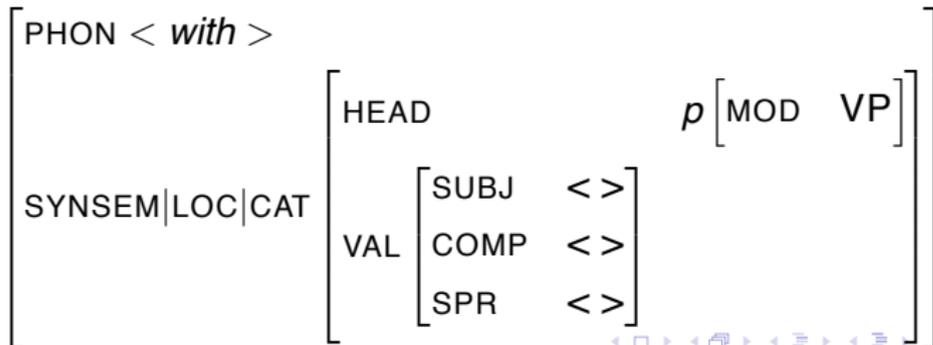


Beispiele von Modifizierer (2)

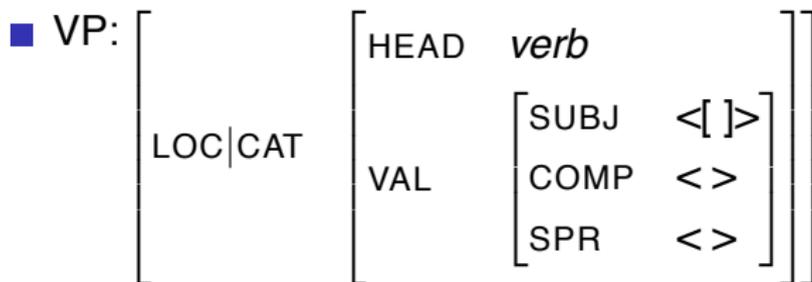
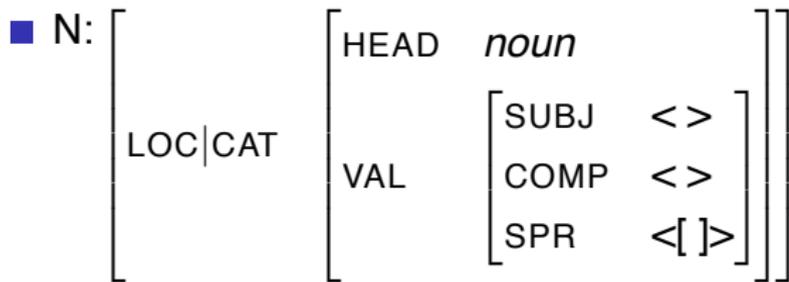
■ *with*:



■ *with*:

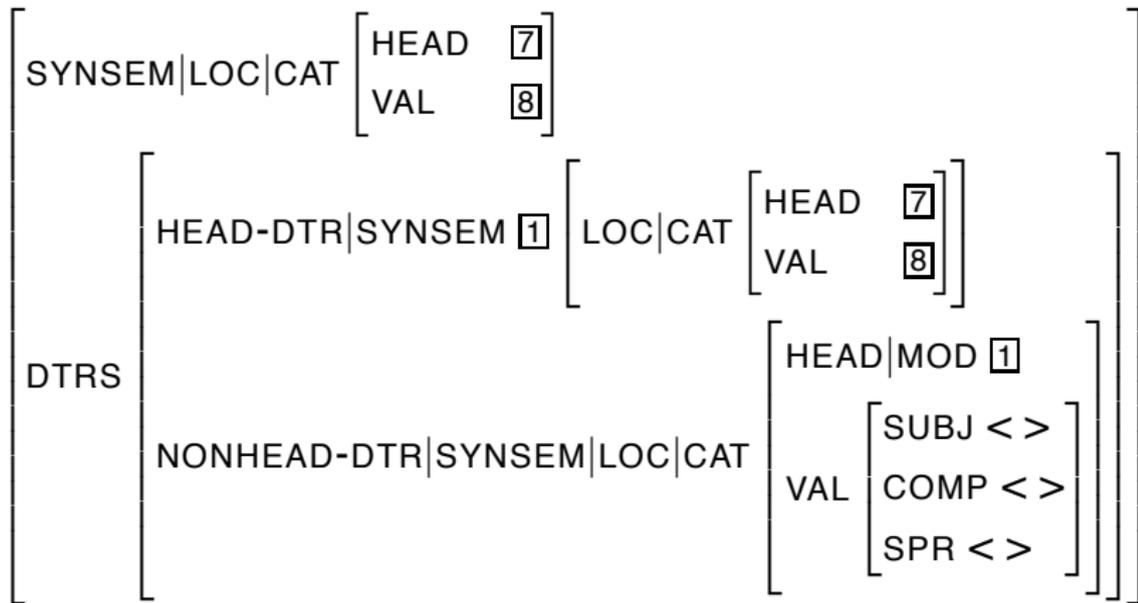


Abkürzungen in obenstehende Merkmalstrukturen



Das Kopf-Adjunkt Schema

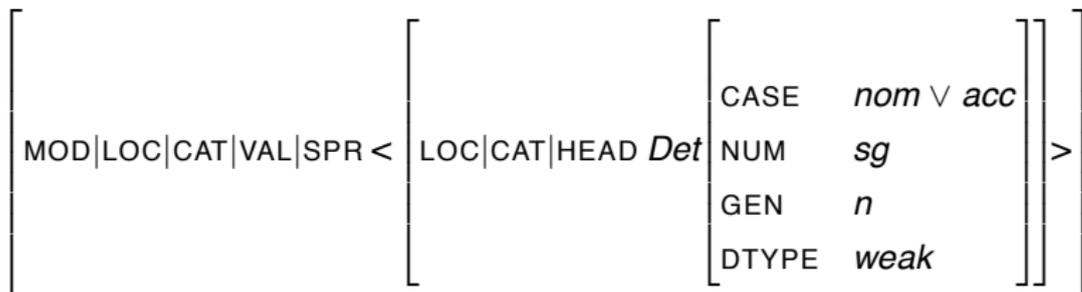
- *Head-Adjunkt-Schema* mit HFP, Valenzprinzip und MOD:



Adjektiv-Noun Kongruenz

- Genau wie bei den Werten von Valenzattributen, kann das MOD-Attribut syntaktische und semantische Beschränkungen einführen
- Zum Beispiel: deutsche Adjektive und Determinierer stimmen in Kasus, Numerus und syntaktischem Genus überein
- Adjektive können in einer bestimmten Form oft nur mit schwachen (z.B. “ein”) oder starken (z.B. “der”) Determinierern kombiniert werden
- Das Attribut DTYP wird verwendet um schwache von starken Determinierern zu trennen
- Die Kongruenz kann (z.B. nach Müller (2007)) zwischen Adjektiv und Determinierer etabliert werden

Beispiel: Modwert von Adjektiv *schlaues* mit Kongruenz-Information



Outline

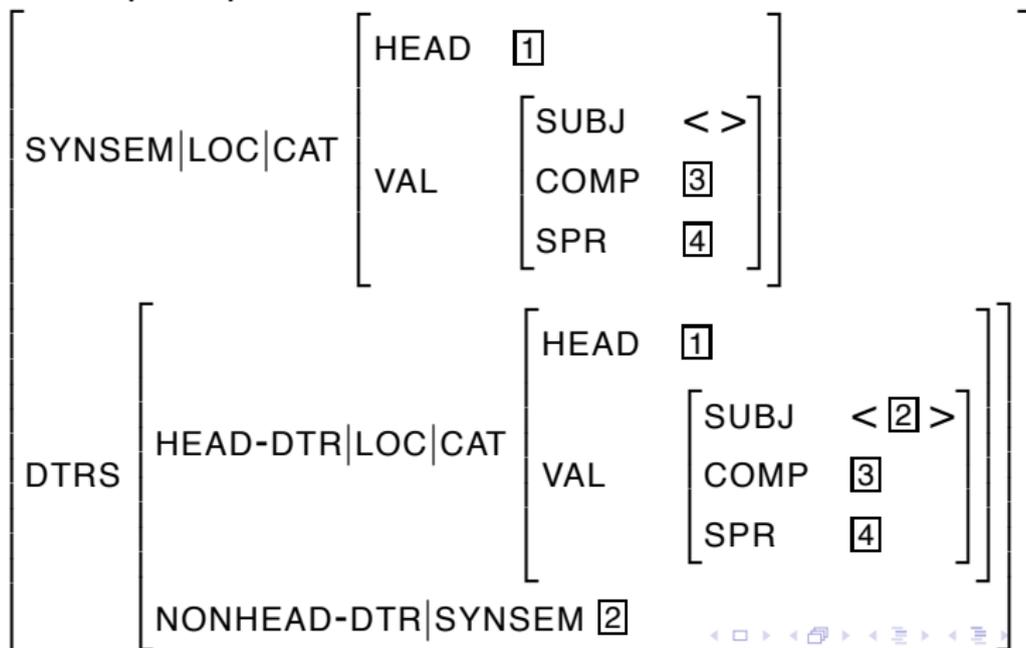
- 1 Einführung
- 2 Lexikalische Beschränkungen
- 3 Modifikation
- 4 Sprachspezifische Schemata**

Sprachspezifische Schemata

- Bisher haben wir allgemeine Definitionen von Schemata gesehen (eher sprachunabhängig)
 - Schemata können auch sprachspezifische Eigenschaften haben
 - Dazu werden sie oft von “linear precedence” Beschränkungen begleitet (immer sprachspezifisch)
- Wir werden jetzt die Schemata und Wortstellungsbeschränkungen für Englisch betrachten

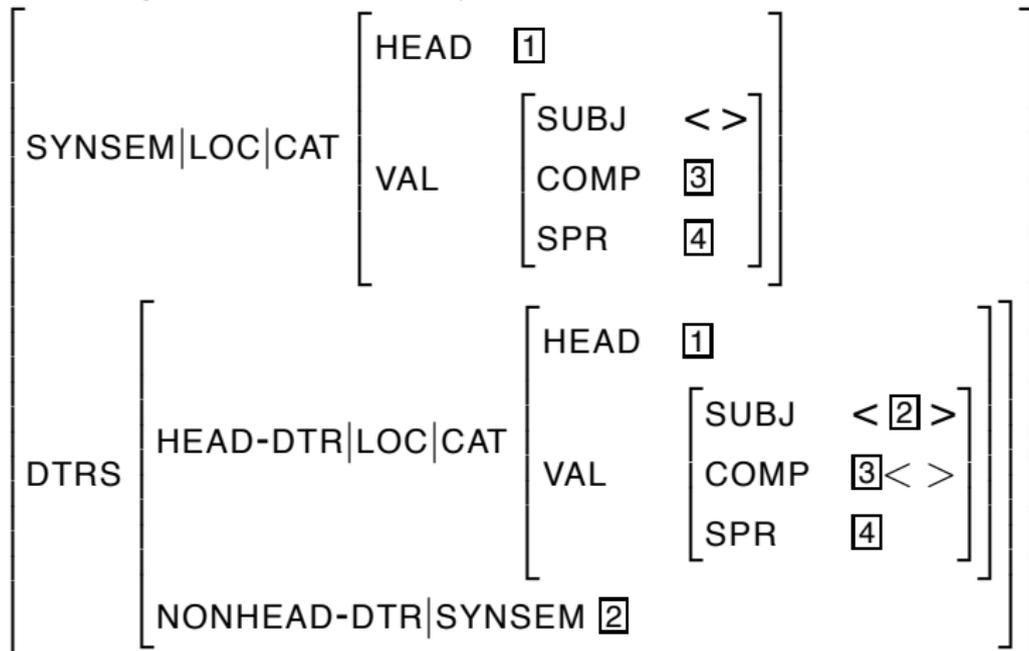
Head-Subject Schema

- Das Allgemeine *Head-Subject-Schema* mit HFP und Valenzprinzip:



Das englische Head-Subject Schema

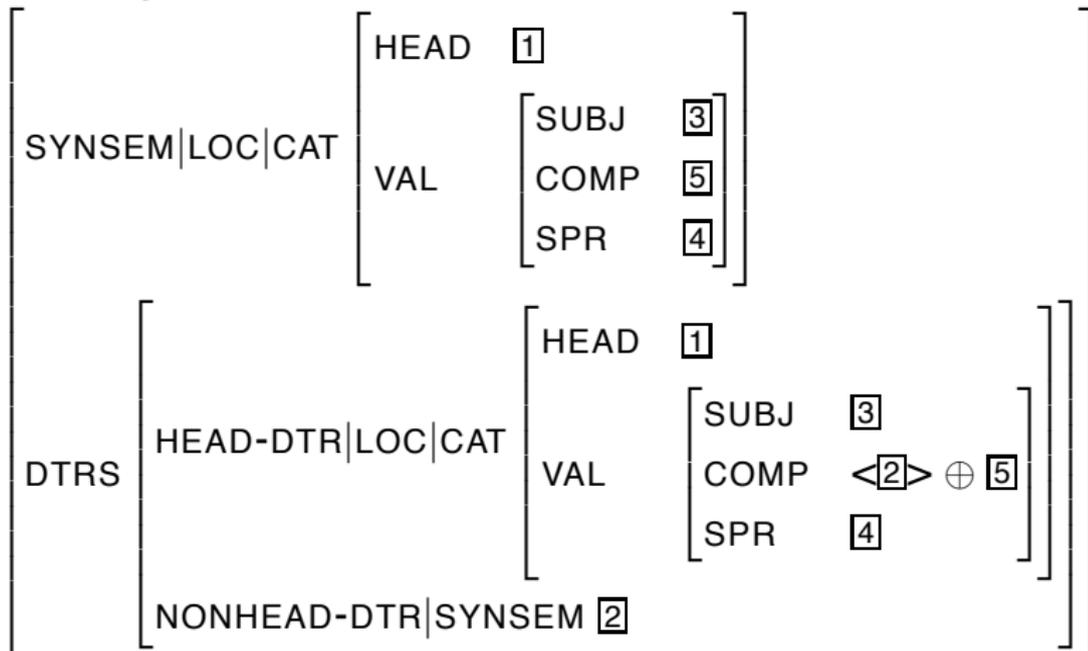
■ Das englische *Head-Subject-Schema*:



■ nonhead-dtr < head-dtr

Head-Complement Schema

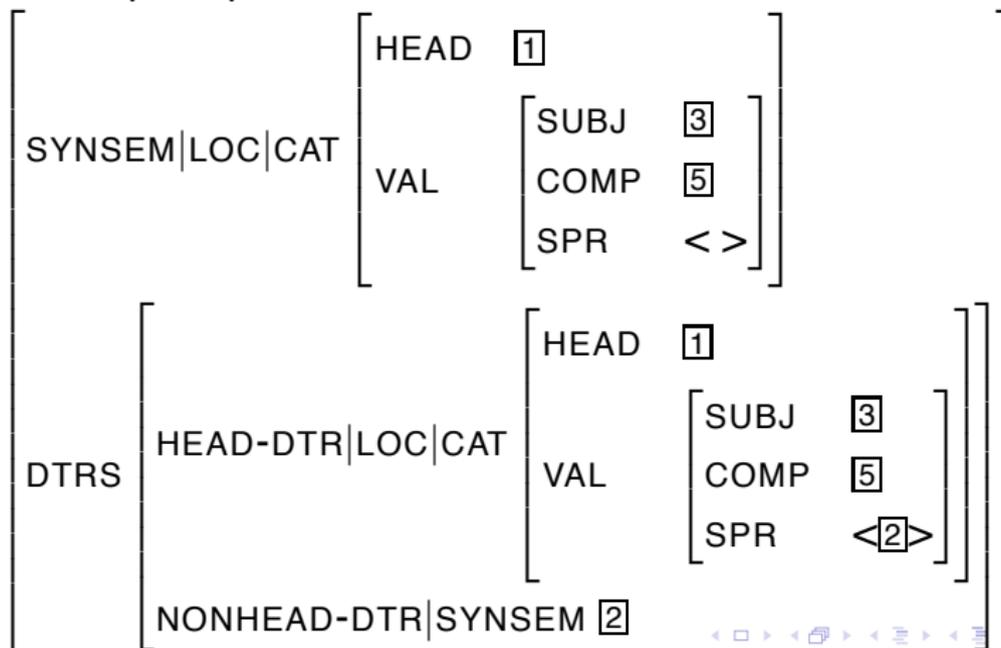
■ Das englische *head-compl-schema*:



■ head-dtr < nonhead-dtr

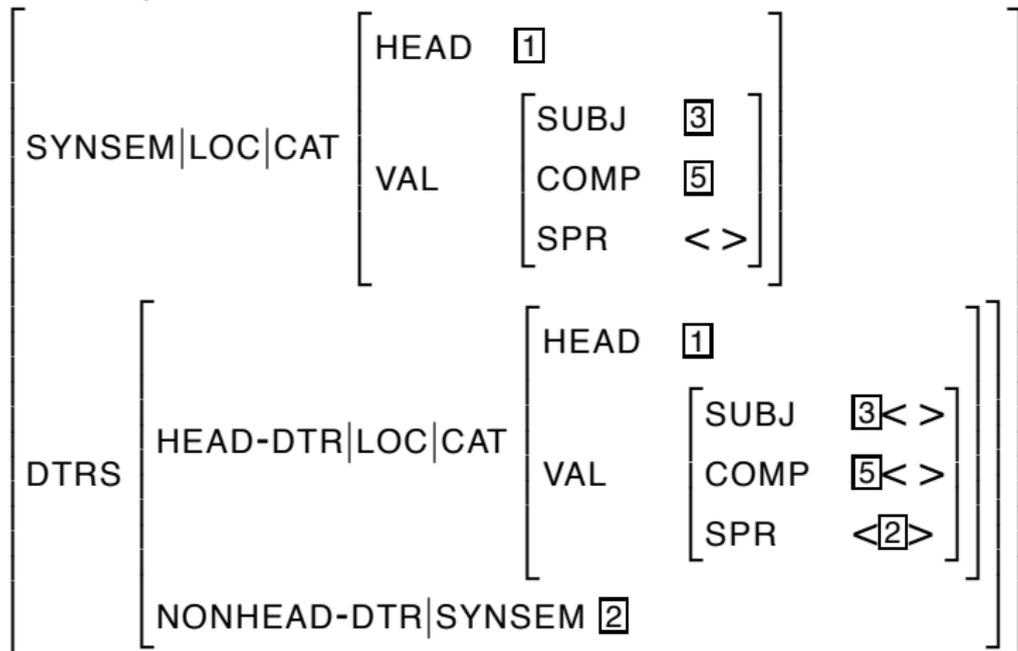
Kopf-Spezifizierer Phrasen

- Das allgemeine *head-spr-schema* mit HFP und Valenzprinzip:



Kopf-Spezifizierer Phrasen

■ Das englische *head-spr-schema*:



■ nonhead-dtr < head-dtr

Bibliography I

- Pollard, Carl and Ivan Sag (2000). *An Information-Based Syntax and Semantics*. Vol.1, Fundamentals. Stanford: CSLI Publications
- Sag, Ivan A. and Thomas Wasow. (1999) *Syntactic Theory: A formal Introduction*. CSLI Publications: Palo Alto, USA.
- **Sag, Ivan A., Thomas Wasow and Emily Bender (2003). *Syntactic Theory: A formal Introduction*. Second Edition. CSLI Publications: Palo Alto, USA.**
- Weibelhuth, Gert, Jean-Pierre Koenig and Andreas Kathol (1999). HPSG as a Theory of Grammar. Introduction to *Lexical and Constructional Aspects of Linguistic Explanation*. CSLI Publications: Palo Alto, USA.
- Pollard, Carl and Ivan A. Sag (1994). *Head-Driven Phrase Structure Grammar*. Chicago Press: Chicago, USA and CSLI Publications: Palo Alto, USA.

Bibliography II

- Müller, Stefan (2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Stauffenburg Einführungen. Stauffenburg Verlag: Tübingen: Germany.