Induction of Classifications from Linguistic Data

speaker: Wiebke Petersen

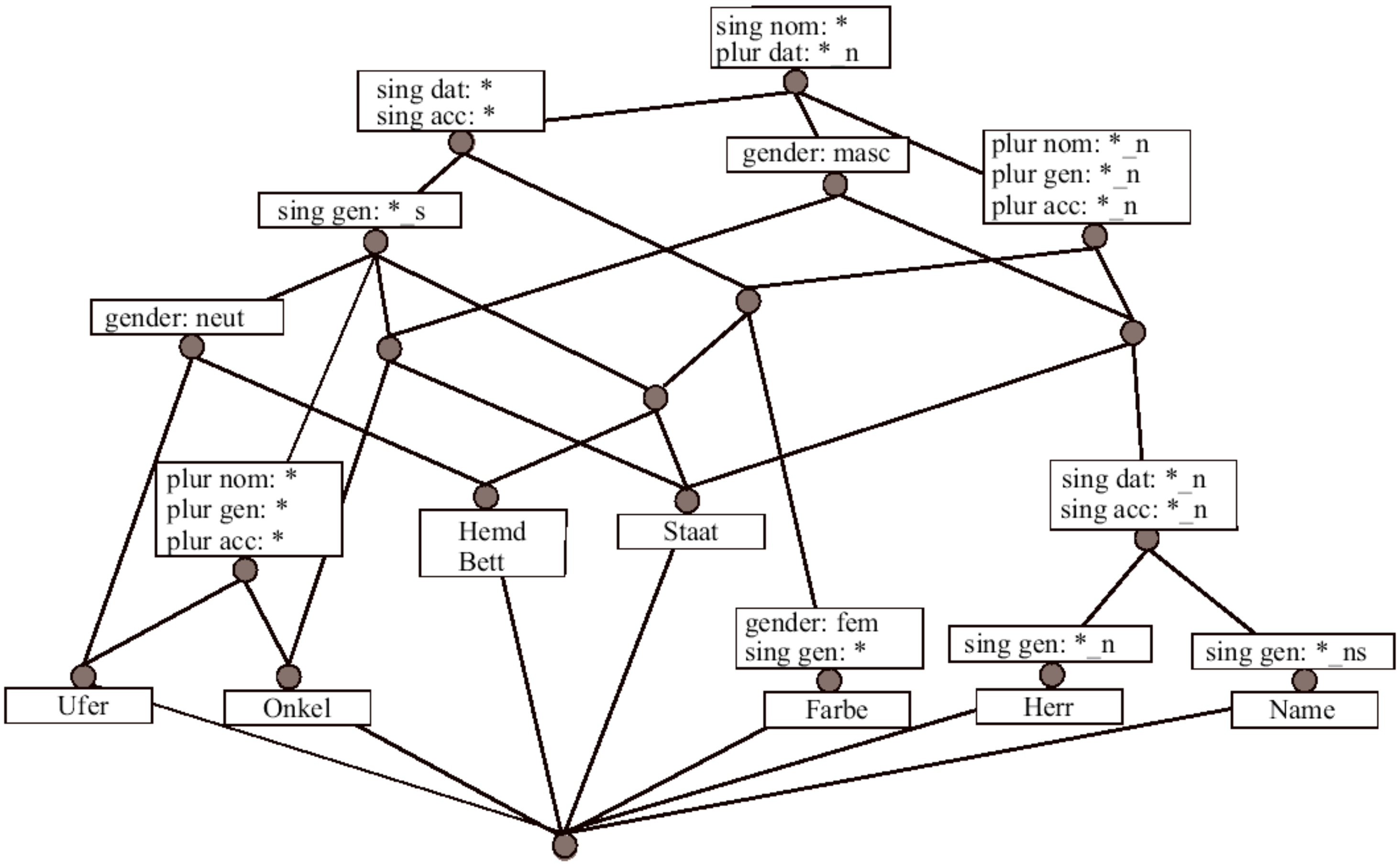
(Osswald/Petersen 2002)

Example context:
Inflectional paradigms of German nouns

	gender	sing nom	sing gen	sing dat	sing acc	plur nom	plur gen	plur dat	plur acc
Herr	masc	*	*_n	*_n	*_n	*_n	*_n	*_n	*_n
Name	masc	*	*_ns	*_n	*_n	*_n	*_n	*_n	*_n
Staat	masc	*	*_s	*	*	*_n	*_n	*_n	*_n
Hemd	neut	*	*_s	*	*	*_n	*_n	*_n	*_n
Farbe	fem	*	*	*	*	*_n	*_n	*_n	*_n
Bett	neut	*	*_s	*	*	*_n	*_n	*_n	*_n
Onkel	masc	*	*_s	*	*	*	*	*_n	*
Ufer	neut	*	*_s	*	*	*	*	*_n	*

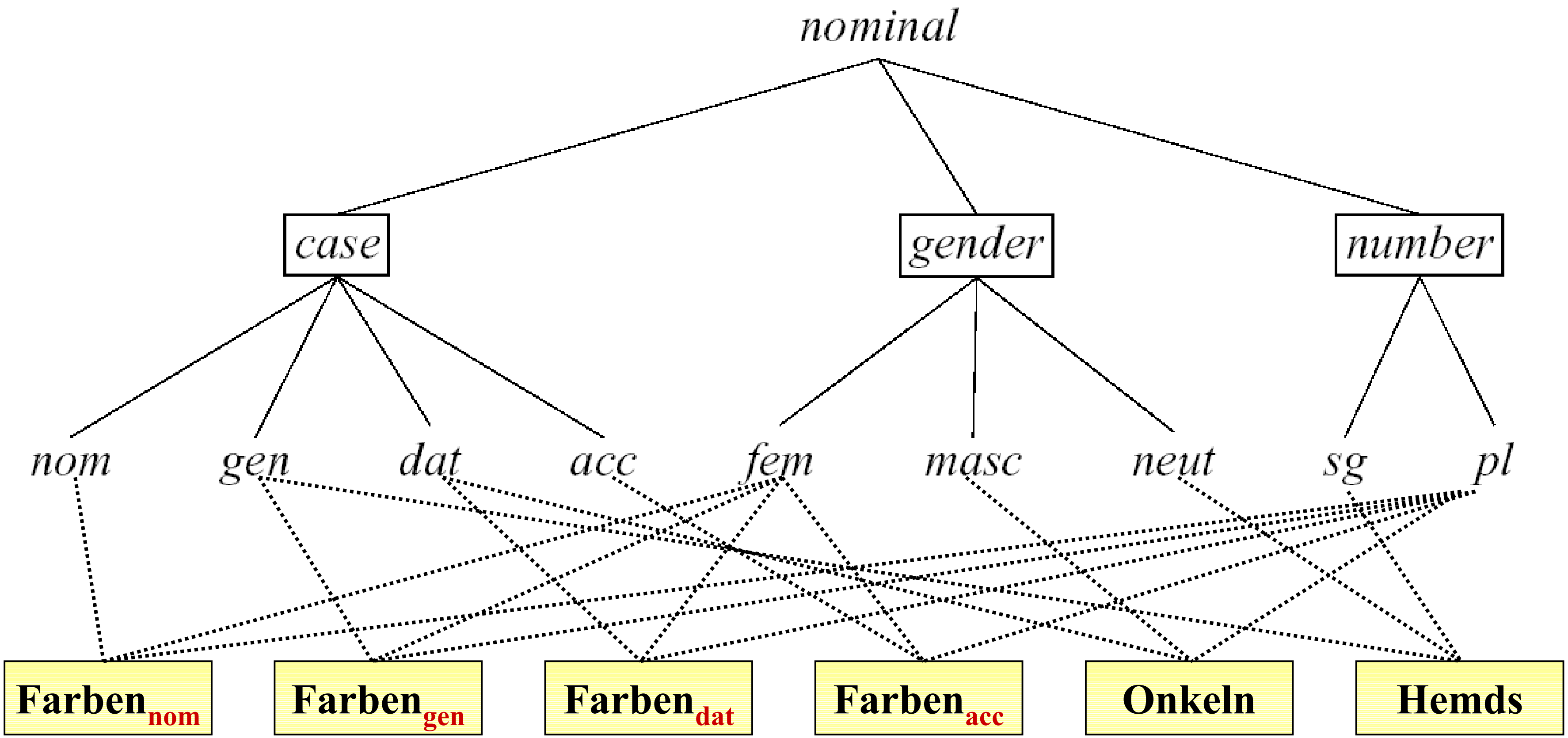
Example Concept Lattice:

Inflectional paradigms of German nouns



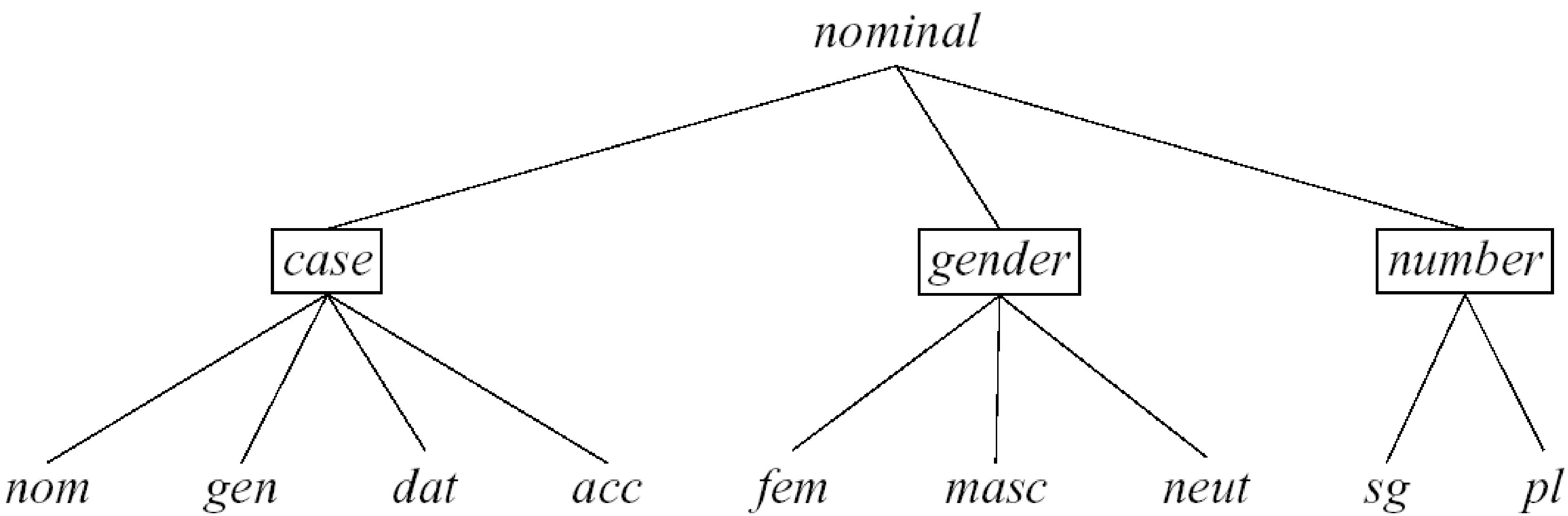
Examples of linguistic classification

And/Or trees (e.g. Koenig, 1999)



Classification as (First Order) Theory

And/OR trees (e.g. Koenig, 1999)



$A \subseteq B$ stands for $\forall x(Ax \rightarrow Bx)$

Λ stands for $\lambda x(x \neq x)$

V stands for $\lambda x(x = x)$

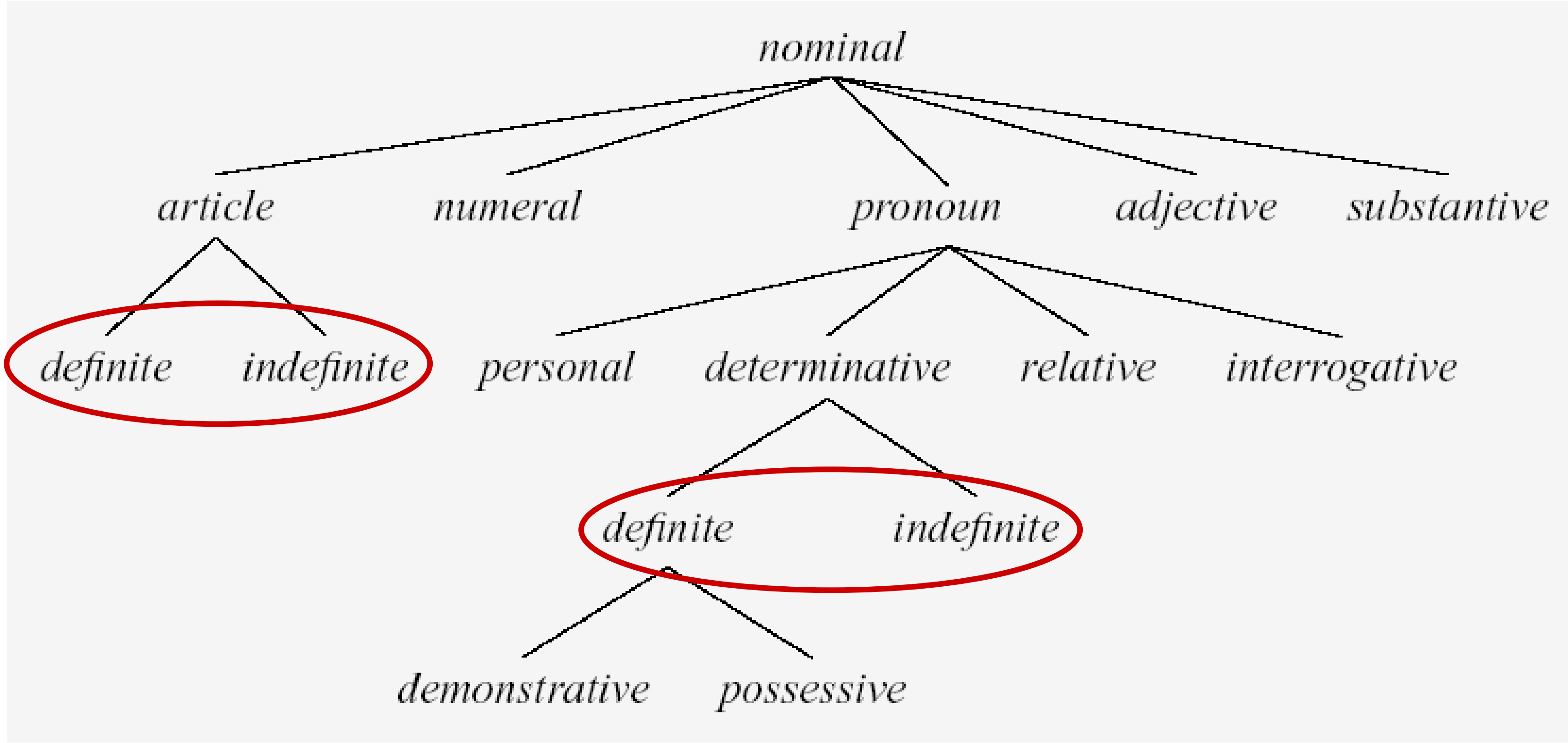
nom $\wedge$ gen $\subseteq \Lambda$, ... , sing $\wedge$ plur $\subseteq \Lambda$ (ISNOTA)

nom $\vee$ gen $\vee$ dat $\vee$ acc $\subseteq$ nominal, ... (ISA)

nominal $\subseteq$ nom $\vee$ gen $\vee$ dat $\vee$ acc, ... (exhaustiveness)

Examples of linguistic classification

Taxonomic trees (after Eisenberg, 1999)



The canonical universe $C(\Gamma)$

For each observational theory Γ over a set of primitive predicates Σ, there is a canonical model $M(\Gamma)=(C(\Gamma), \models)$, where

$X \models p$ iff $p \in X$, for every $X \in C(\Gamma)$ and $p \in \Sigma$.

$\models$ is inductively extended to $T[\Gamma]$, the term algebra of observational predicates over Σ.

$C(\Gamma)$ consists of the Γ-closed consistent subsets of Σ.

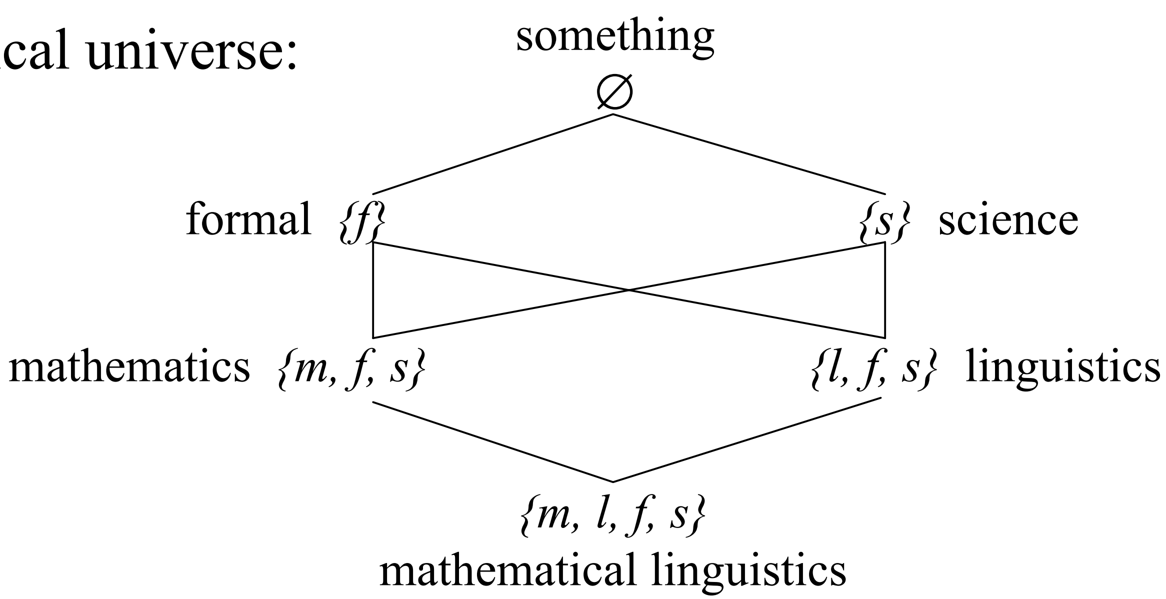
Construction of the canonical universe

Theory: (Mathematics and linguistics are the only formal sciences)

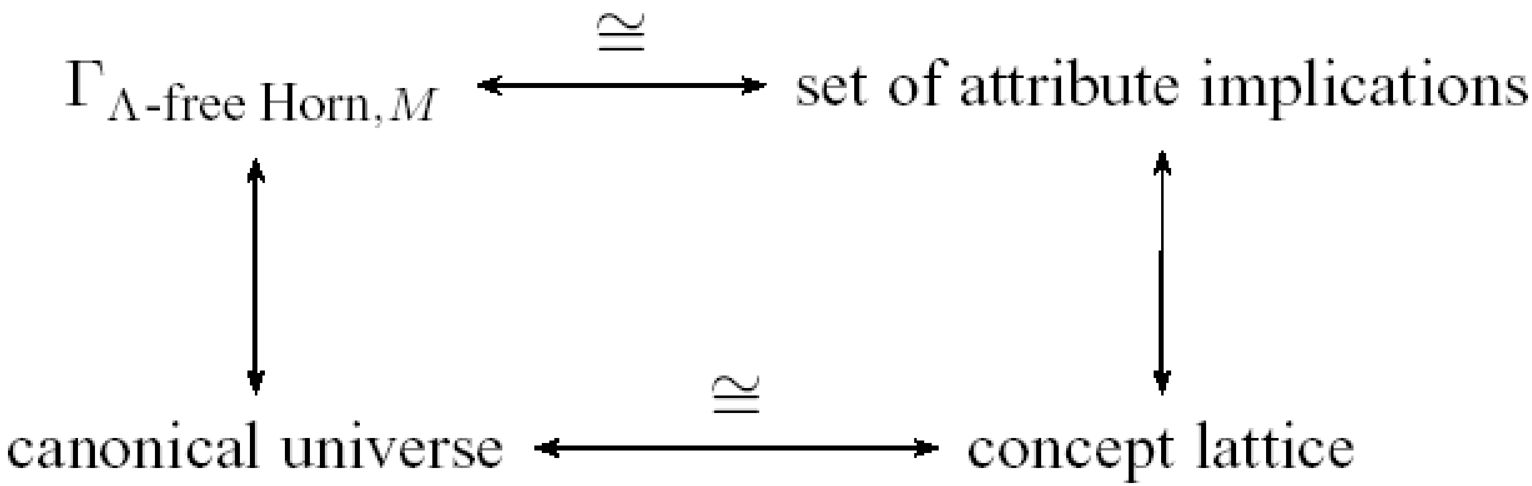
$\text{formal} \wedge \text{science} \subseteq \text{mathematics} \vee \text{linguistics}$

$\text{formal} \wedge \text{science} \supseteq \text{mathematics} \vee \text{linguistics}$

Canonical universe:



Relationship between Λ-free Horn theories and Concept Lattices



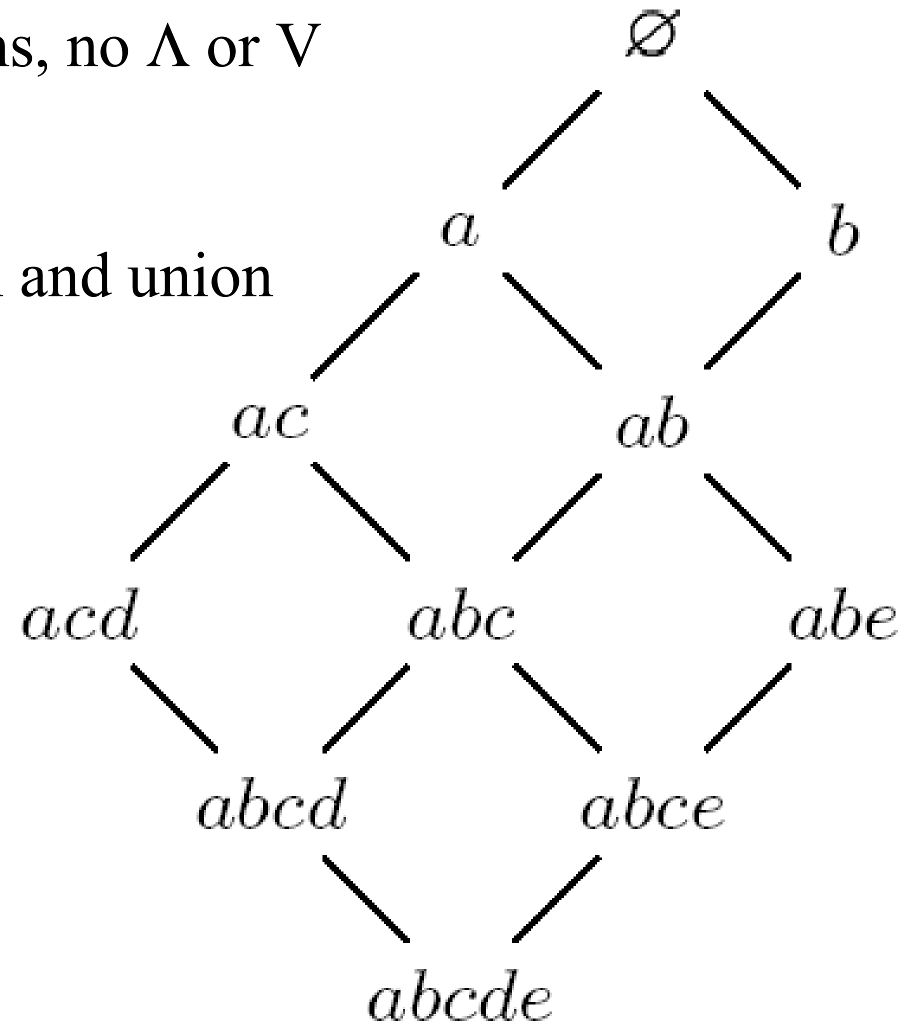
Simple inheritance

no disjunctions, no conjunctions, no Λ or V

$\Gamma=\{d \subseteq c, c \subseteq a, e \subseteq a, e \subseteq b\}$

$C(\Gamma)$: closed under intersection and union

	a	b	c	d	e
x_1	X	X			X
x_2	X				
x_3		X			
x_4	X	X	X	X	
x_5	X		X		
x_6	X	X	X	X	
x_7	X		X	X	



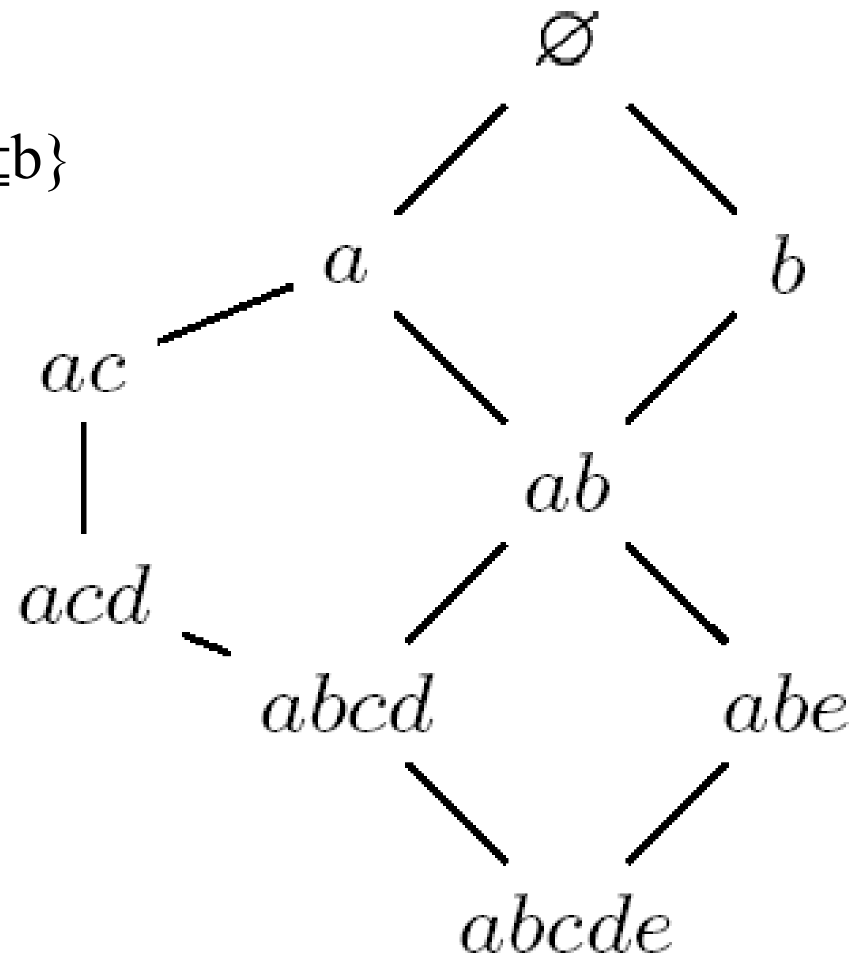
$\wedge$-free Horn Theory

no discjunctions

$\Gamma = \{b \wedge c \subseteq d, d \subseteq c, c \subseteq a, e \subseteq a, e \subseteq b\}$

$C(\Gamma)$: closed under intersection

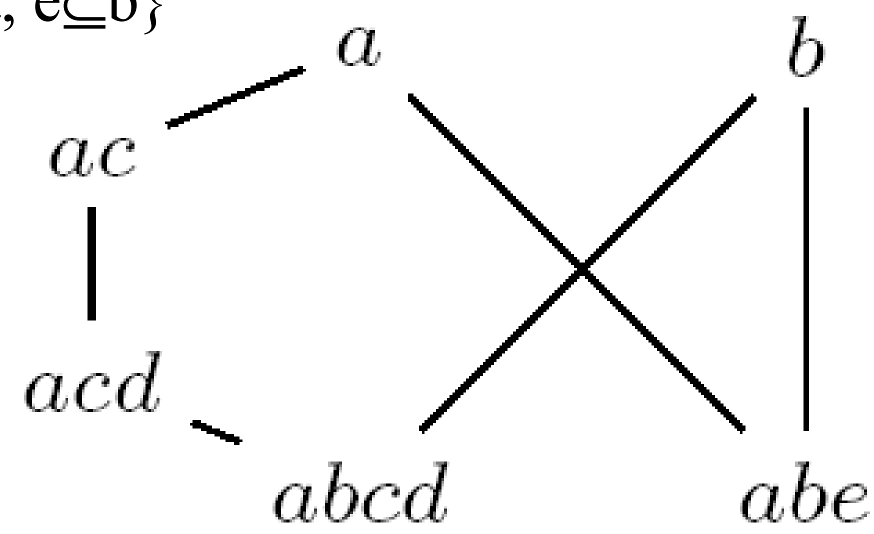
	a	b	c	d	e
x_1	X	X			X
x_2	X				
x_3		X			
x_4	X	X	X	X	
x_5	X		X		
x_6	X	X	X	X	
x_7	X		X	X	



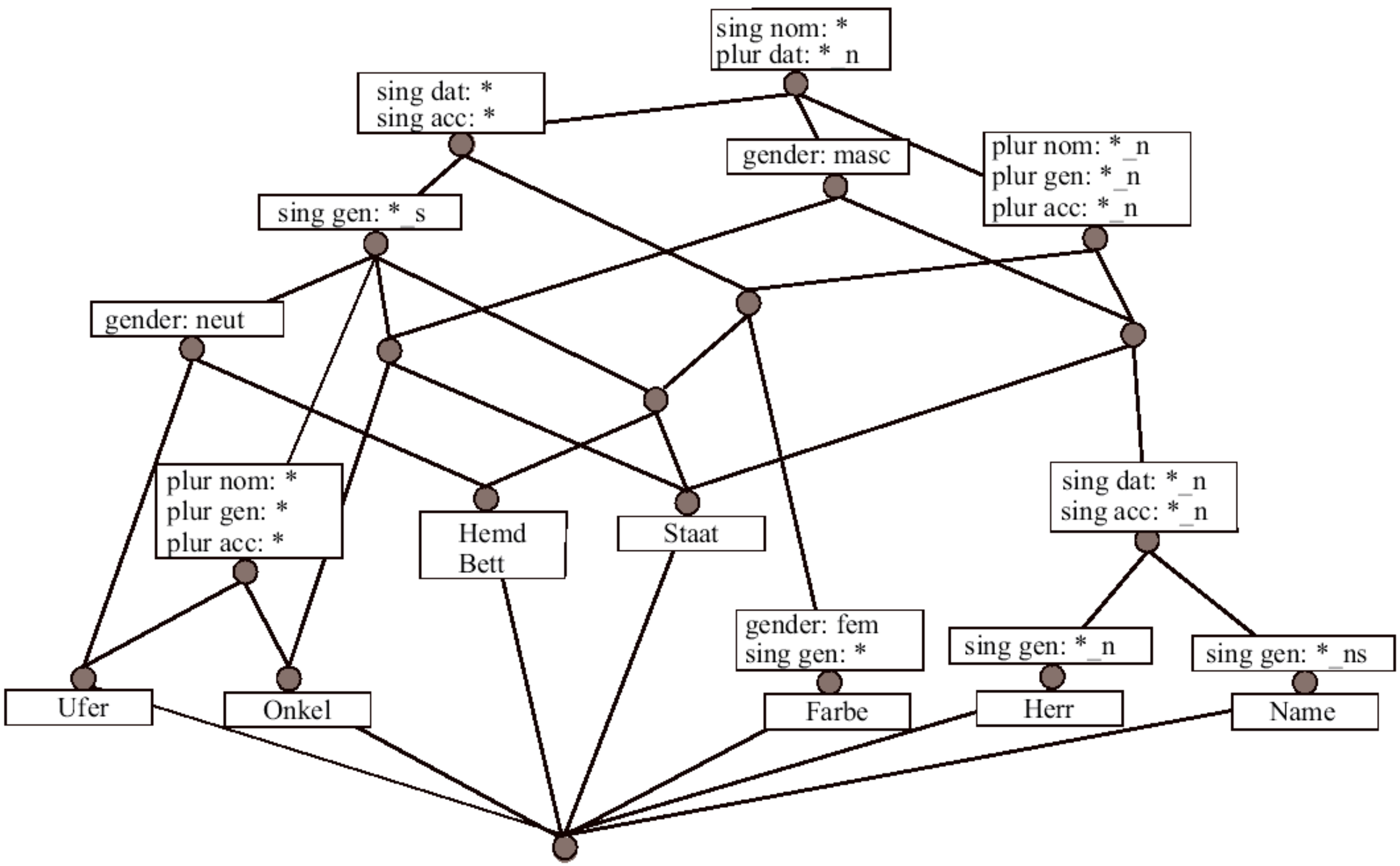
Observational Theory

$\Gamma=\{V\subseteq a\vee b,\ a\wedge b \subseteq c\vee e, c\wedge e \subseteq \Lambda,$
$b\wedge c \subseteq d,\ d\subseteq c,\ c\subseteq a,\ e\subseteq a,\ e\subseteq b\}$

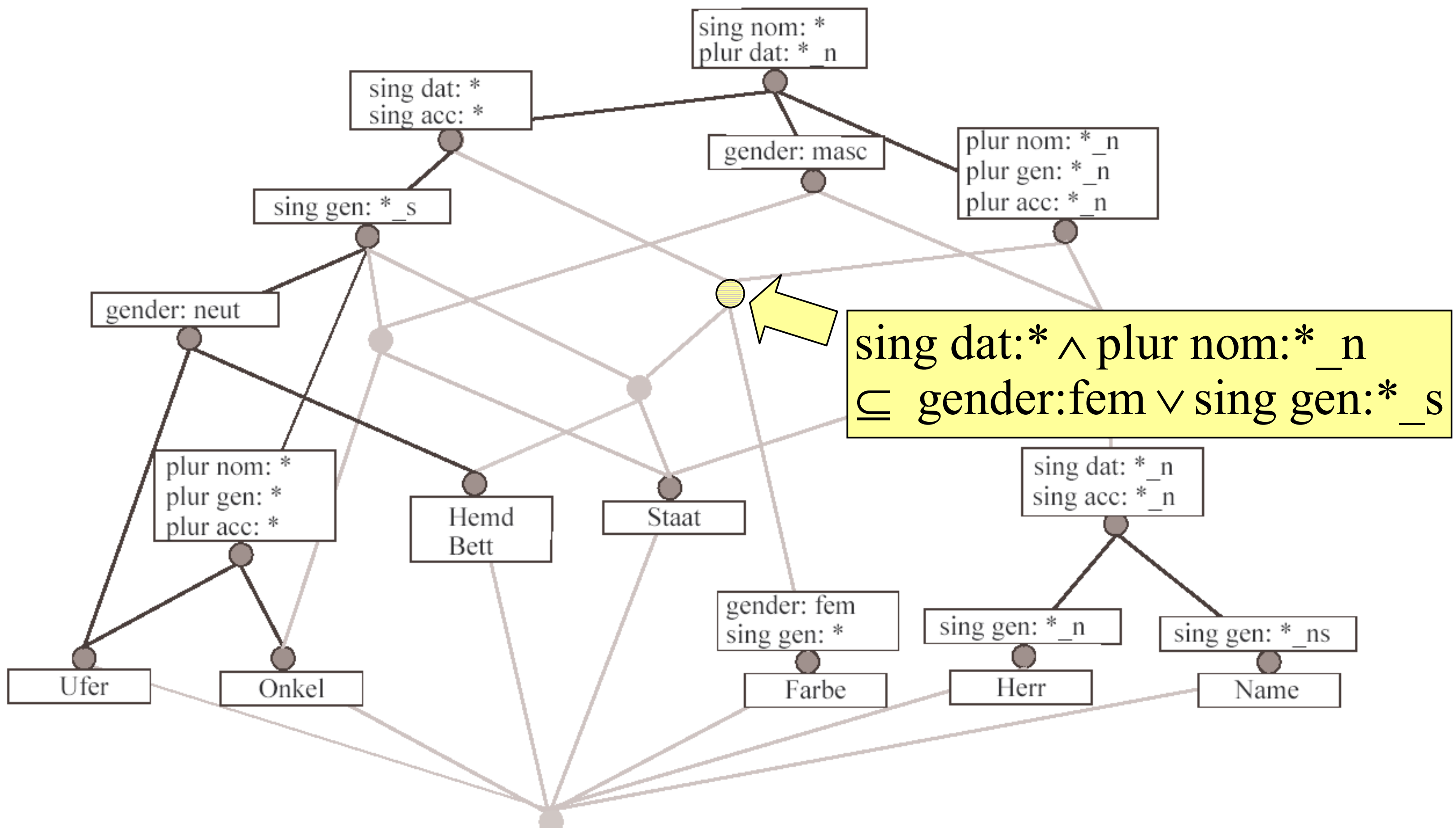
	a	b	c	d	e
x_1	X	X			X
x_2	X				
x_3		X			
x_4	X	X	X	X	
x_5	X		X		
x_6	X	X	X	X	
x_7	X		X	X	



Concept lattice ≅ Λ-free Horn theory



AOC-poset


sing nom: *
plur dat: *_n
sing dat: *
sing acc: *
gender: masc
plur nom: *_n
plur gen: *_n
plur acc: *_n
sing gen: *_s
gender: neut
sing dat:* ∧ plur nom:*_n
⊆ gender:fem ∨ sing gen:*_s
plur nom: *
plur gen: *
plur acc: *
Hemd
Bett
Staat
sing dat: *_n
sing acc: *_n
gender: fem
sing gen: *
sing gen: *_n
sing gen: *_ns
Ufer
Onkel
Farbe
Herr
Name

Relationship between a theory Γ and its canonical universe C(Γ)

Class of Γ	Closure properties of $C(\Gamma)$
observational	local membership
Horn	nonempty intersection $+$ directed union
Λ-free Horn	intersection $+$ directed union
simple inheritance	intersection $+$ union
exclusion	subsets $+$ finitely bounded union
simple inheritance $+$ exclusion	nonempty intersection $+$ finitely bounded union

Examples of linguistic classification

Systemic networks (e.g. Winograd, 1983)

$$\text{third} \wedge \text{sing} \subseteq \text{fem} \vee \text{masc} \vee \text{neut}$$
$$\text{third} \wedge \text{sing} \supseteq \text{fem} \vee \text{masc} \vee \text{neut}$$