What feature co-occurrence restrictions have to do with type signatures

Wiebke Petersen
James Kilbury

FG/MOL 2005

overview

- introduction
  - feature co-occurrence restrictions in GPSG
  - type signatures in HPSG
  - formal concept analysis (FCA)
- FCA relates FCRs and type signatures
  - type signatures and concept lattices
  - FCRs and concept lattices
- outlook

unification-based grammar formalisms (UBG)

- **tool oriented** (according to Shieber 1986)
  - Functional Unification Grammar (**FUG**, Kay 1979)
  - **PATR-II** (Shieber et al. 1983)
  - Attribute Logic Engine (**ALE**, Carpenter 1992)
  - . . .
- **theory oriented**
  - Lexical-Functional Grammar (**LFG**, Bresnan & Kaplan 1982)
  - General. Phr.-Structure Gram.(**GPSG**, Gazdar & Pullum 1982)
  - Head-driven Phr.-Structure Gram. (**HPSG**, Pollard & Sag 1987)
  - Unification Categorial Grammar (**UCG**, Zeevat & Calder 1987)

feature co-occurrence restrictions (FCRs)

$[\mathsf{AGR}] \supset [-\mathsf{N}, +\mathsf{V}]$
'If something is specified for agreement, it is a verb.'

$[+\mathsf{INV}] \supset [+\mathsf{AUX}, \mathsf{FIN}]$
'If something is inverted, it is a finite form of an auxiliary [verb].'

G. Gazdar, E. Klein, G. Pullum, and I. Sag: *Generalized Phrase Structure Grammar* (1985)

FCRs in the UBG-family

- **tool oriented**
  - — FUG — no correspondence
  - — PATR-II — templates ??
  - — ALE, . . . — see HPSG
- **theory oriented**
  - — LFG — no correspondence (?)
  - — **GPSG** — **origin of FCRs**
  - — **HPSG** — conditional feature structures
    type signatures (*subject of the talk*)

Head-driven Phrase-Structure Grammar

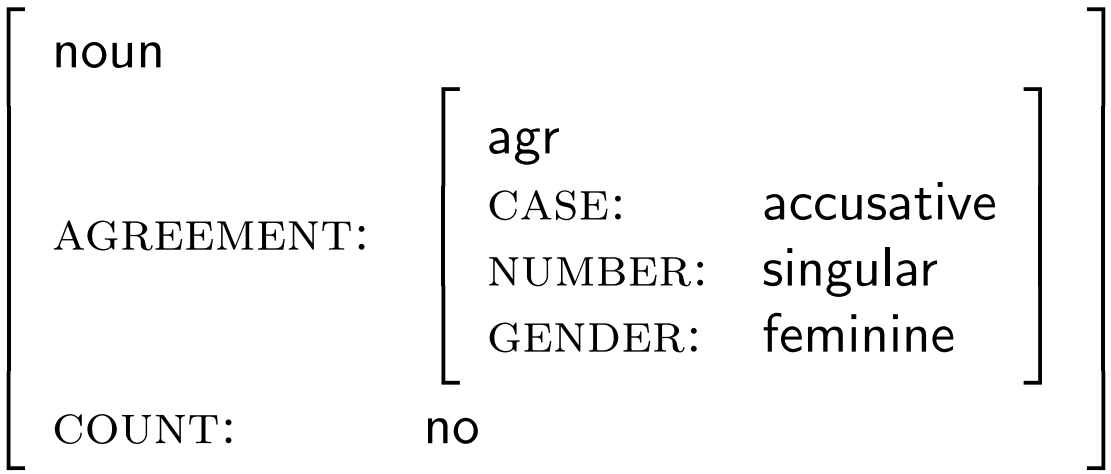
Carl Pollard & Ivan Sag:

Information-based Syntax and Semantics (1987)
Head-driven Phrase-Structure Grammar (1994)

Bob Carpenter:

The Logic of Typed Feature-Structures (1992)

typed feature structure

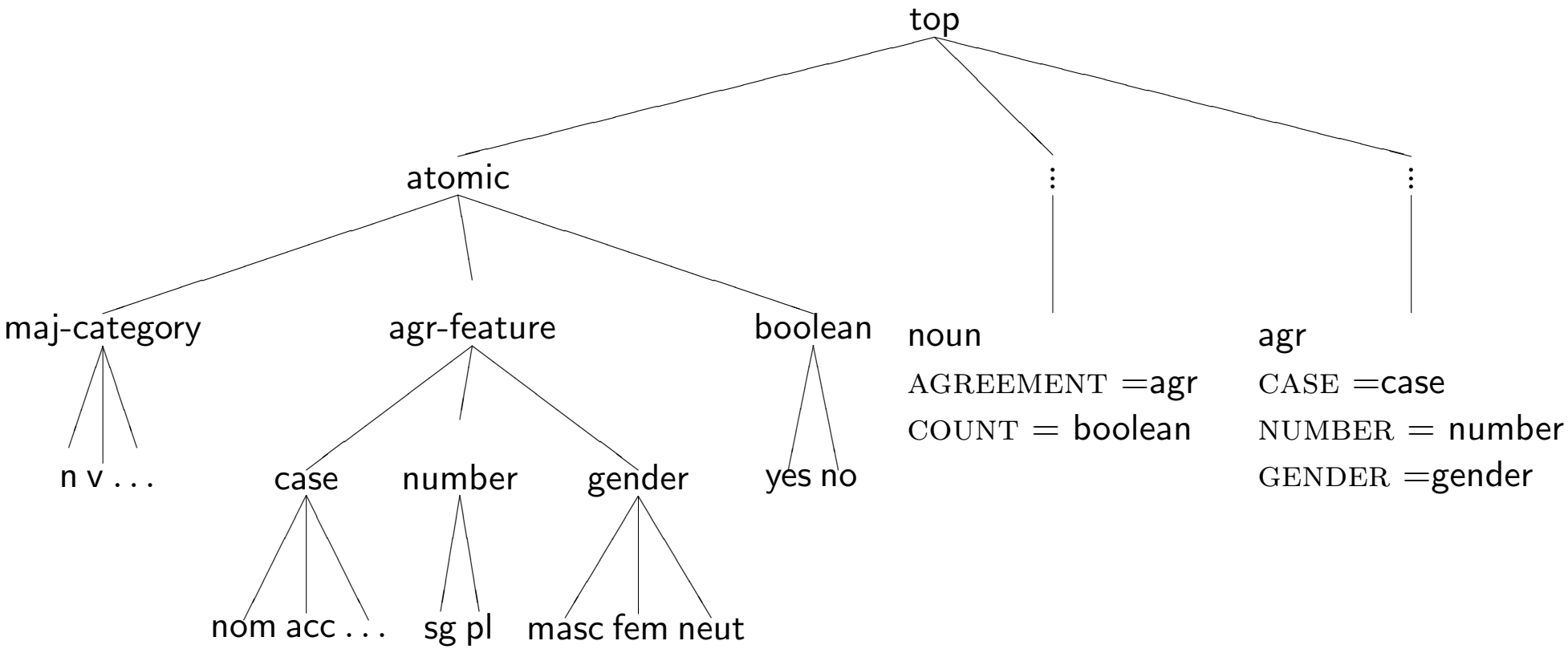


basic concepts:

- *well-typed*: "every feature that occurs is *appropriate* and takes an appropriate value" (Carpenter 1992)
- *totally well-typed*: "well-typed and "every feature which is appropriate must be present" (Carpenter 1992)
- *sort-resolved*: "every node is assigned a sort label that is maximal (i.e. most specific) in the sort ordering." (Pollard & Sag 1994)

"Feature structures that serve as models of linguistic entities [. . .] required to be both *totally well-typed* and *sort-resolved*." (Pollard & Sag 1994)

type signature



basic conditions (Carpenter 1992):

- meet-semi-latticehood
- unique feature introduction

formal concept analysis (FCA)

Bernhard Ganter & Rudolf Wille:

Formal Concept Analysis. Mathematical Foundations (1999)

basic concepts:

- formal context
- formal concept (lattice)
- attribute implication

lexemes classified w.r.t. their feature specifications

$$\text{sing}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{bse} \\ \text{aux} : & - \\ \text{inv} : & - \end{bmatrix} \qquad \text{sings}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{fin} \\ \text{aux} : & - \\ \text{inv} : & - \end{bmatrix} \qquad \text{sung}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{pas} \\ \text{aux} : & - \\ \text{inv} : & - \end{bmatrix}$$

$$\text{can}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{fin} \\ \text{aux} : & + \\ \text{inv} : & + \end{bmatrix} \qquad \text{can}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{fin} \\ \text{aux} : & + \\ \text{inv} : & - \end{bmatrix} \qquad \text{can}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & - \\ \text{vform} : & \text{bse} \\ \text{aux} : & + \\ \text{inv} : & - \end{bmatrix}$$

$$\text{child}: \begin{bmatrix} \text{v} : & - \\ \text{n} : & + \\ \text{nform} : & \text{norm} \end{bmatrix} \qquad \text{it}: \begin{bmatrix} \text{v} : & - \\ \text{n} : & + \\ \text{nform} : & \text{it} \end{bmatrix} \qquad \text{little}: \begin{bmatrix} \text{v} : & + \\ \text{n} : & + \end{bmatrix}$$

$$\text{with}: \begin{bmatrix} \text{v} : & - \\ \text{n} : & - \\ \text{pform} : & \text{with} \end{bmatrix}$$

feature-structure context

	v:+	v:−	n:+	n:−	vform:bse	vform:fin	vform:pas	aux:+	aux:−	inv:+	inv:−	nform:norm	nform:it	v:VAL	n:VAL	vform:VAL	aux:VAL	inv:VAL	nform:VAL	pform:with	pform:VAL
sing	×			×	×				×		×			×	×	×	×	×			
sings	×			×		×			×		×			×	×	×	×	×			
sung	×			×			×		×		×			×	×	×	×	×			
can1	×			×		×		×		×				×	×	×	×	×			
can2	×			×		×		×			×			×	×	×	×	×			
can3	×			×	×			×			×			×	×	×	×	×			
child		×	×									×		×	×				×		
it		×	×										×	×	×				×		
little	×		×											×	×						
with		×		×										×	×					×	×

Definition 1. *A* ***formal context*** K *is a triple* (G, M, I)*, consisting of a set of objects* G*, a set of attributes* M *and a binary* incidence relation $I \subseteq G \times M$*, where* $(g, m) \in I$ *is read as "the formal object* g *has the formal attribute* m*."*

feature-structure context

B	v:+	v:−	n:+	n:−	vform:bse	vform:fin	vform:pas	aux:+	aux:−	inv:+	inv:−	nform:norm	nform:it	v:VAL	n:VAL	vform:VAL	aux:VAL	inv:VAL	nform:VAL	pform:with	pform:VAL
sing	×			×	×				×		×			×	×	×	×	×			
sings	×			×		×			×		×			×	×	×	×	×			
sung	×			×			×		×		×			×	×	×	×	×			
can1	×			×		×		×		×				×	×	×	×	×			
can2	×			×		×		×			×			×	×	×	×	×			
can3	×			×	×			×			×			×	×	×	×	×			
child		×	×									×		×	×				×		
it		×	×										×	×	×				×		
little	×		×											×	×						
with		×		×										×	×					×	×

Definition 2. *For any subset of objects $A \subseteq G$, their set of common attributes is defined as $A' \stackrel{\text{def}}{=} \{m \in M : \forall g \in A : (g, m) \in I\}$. Analogously, the set of common objects for a subset $B \subseteq M$ of attributes is $B' \stackrel{\text{def}}{=} \{g \in G : \forall m \in B : (g, m) \in I\}$.*

feature-structure context

B'	v:+	v:−	n:+	n:−	vform:bse	vform:fin	vform:pas	aux:+	aux:−	inv:+	inv:−	nform:norm	nform:it	v:VAL	n:VAL	vform:VAL	aux:VAL	inv:VAL	nform:VAL	pform:with	pform:VAL
sing	×			×	×				×		×			×	×	×	×	×			
sings	×			×		×			×		×			×	×	×	×	×			
sung	×			×			×		×		×			×	×	×	×	×			
can1	×			×		×		×		×				×	×	×	×	×			
can2	×			×		×		×			×			×	×	×	×	×			
can3	×			×	×			×			×			×	×	×	×	×			
child		×	×									×		×	×				×		
it		×	×										×	×	×				×		
little	×		×											×	×						
with		×		×										×	×					×	×

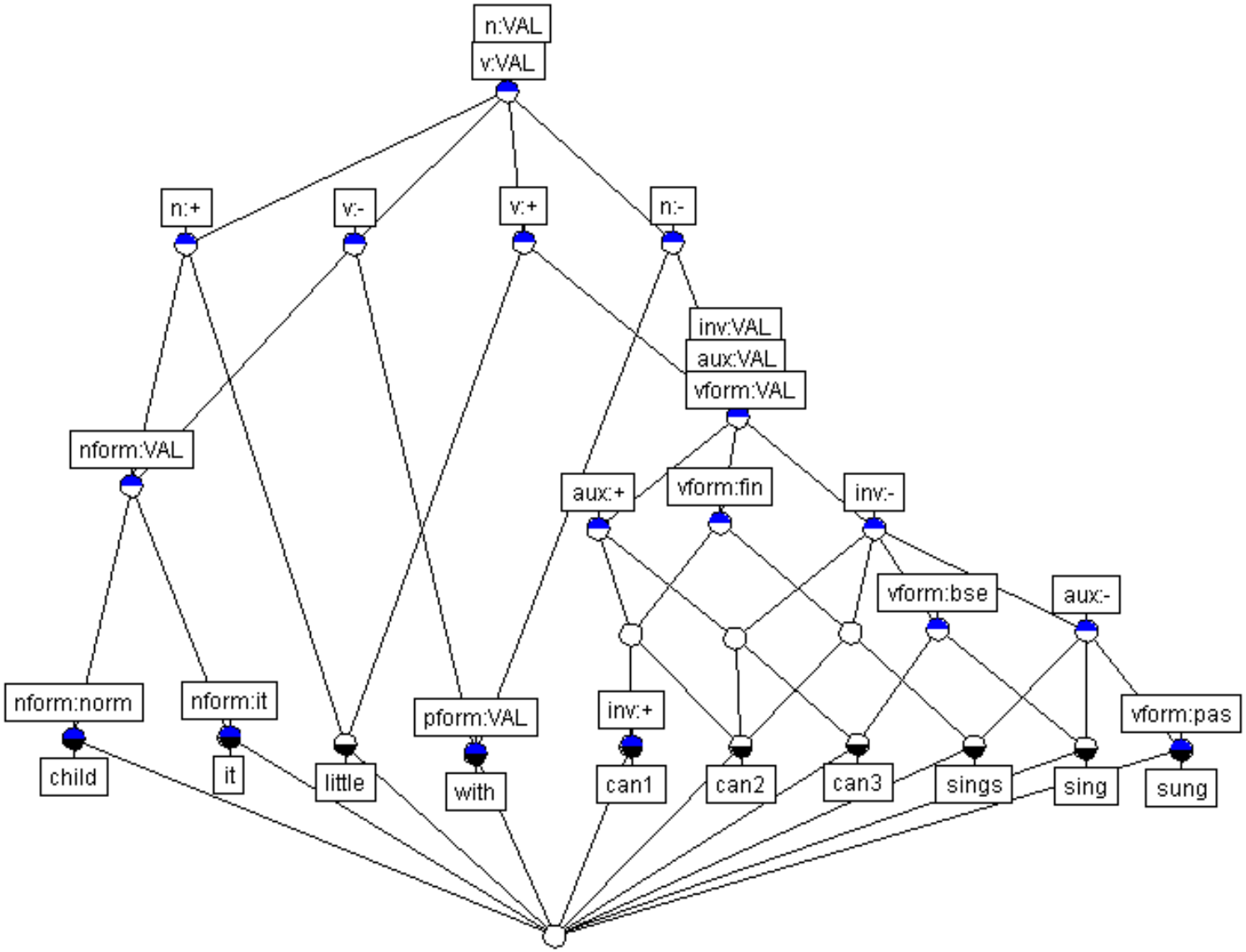
Definition 3. *For any subset of objects $A \subseteq G$, their set of common attributes is defined as $A' \stackrel{\text{def}}{=} \{m \in M : \forall g \in A : (g, m) \in I\}$. Analogously, the set of common objects for a subset $B \subseteq M$ of attributes is $B' \stackrel{\text{def}}{=} \{g \in G : \forall m \in B : (g, m) \in I\}$.*

feature-structure context

B''	v:+	v:−	n:+	n:−	vform:bse	vform:fin	vform:pas	aux:+	aux:−	inv:+	inv:−	nform:norm	nform:it	v:VAL	n:VAL	vform:VAL	aux:VAL	inv:VAL	nform:VAL	pform:with	pform:VAL
sing	×			×	×				×		×			×	×	×	×	×			
sings	×			×		×			×		×			×	×	×	×	×			
sung	×			×			×		×		×			×	×	×	×	×			
can1	×			×		×		×		×				×	×	×	×	×			
can2	×			×		×		×			×			×	×	×	×	×			
can3	×			×	×			×			×			×	×	×	×	×			
child		×	×									×		×	×				×		
it		×	×										×	×	×				×		
little	×		×											×	×						
with		×		×										×	×					×	×

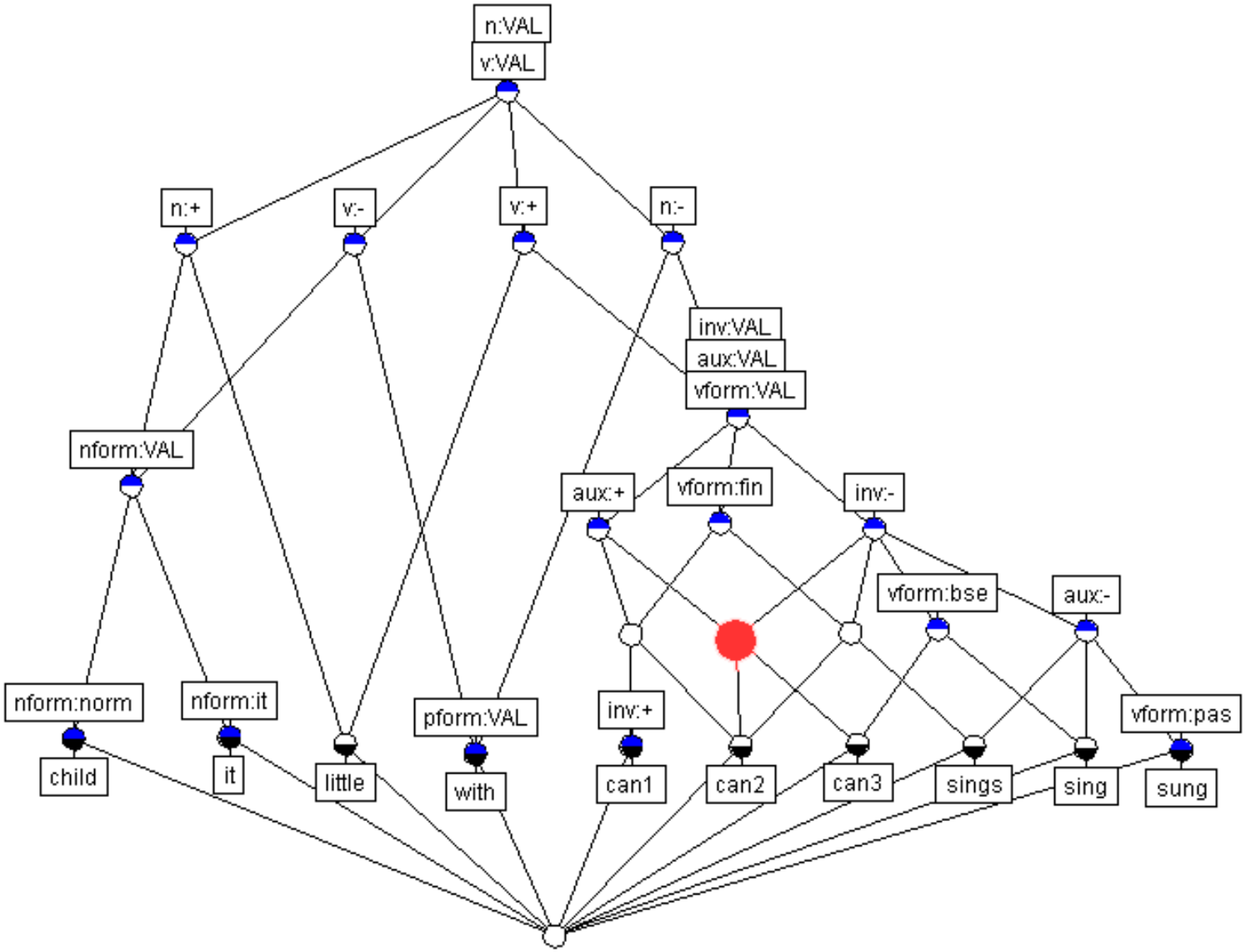
Definition 4. *A* ***formal concept*** *is a pair* (A, B)*, with the properties* $A = B'$ *and* $B = A'$*, where* A *is called the* ***extent*** *and* B *the* ***intent*** *of the concept.* (B', B'') *is a formal concept for all* $B \subseteq M$*.*

concept lattice



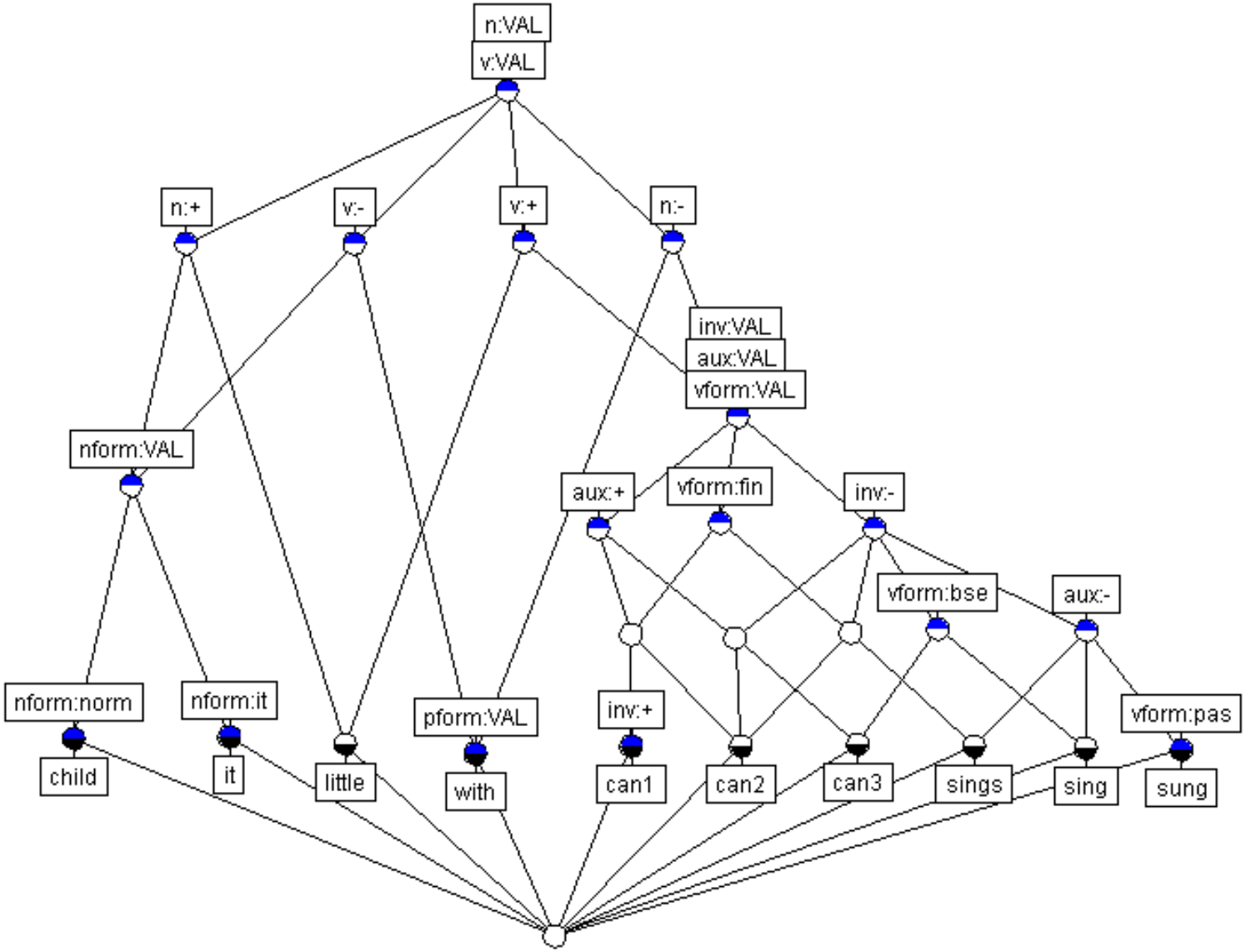
concept order: $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$.

concept lattice



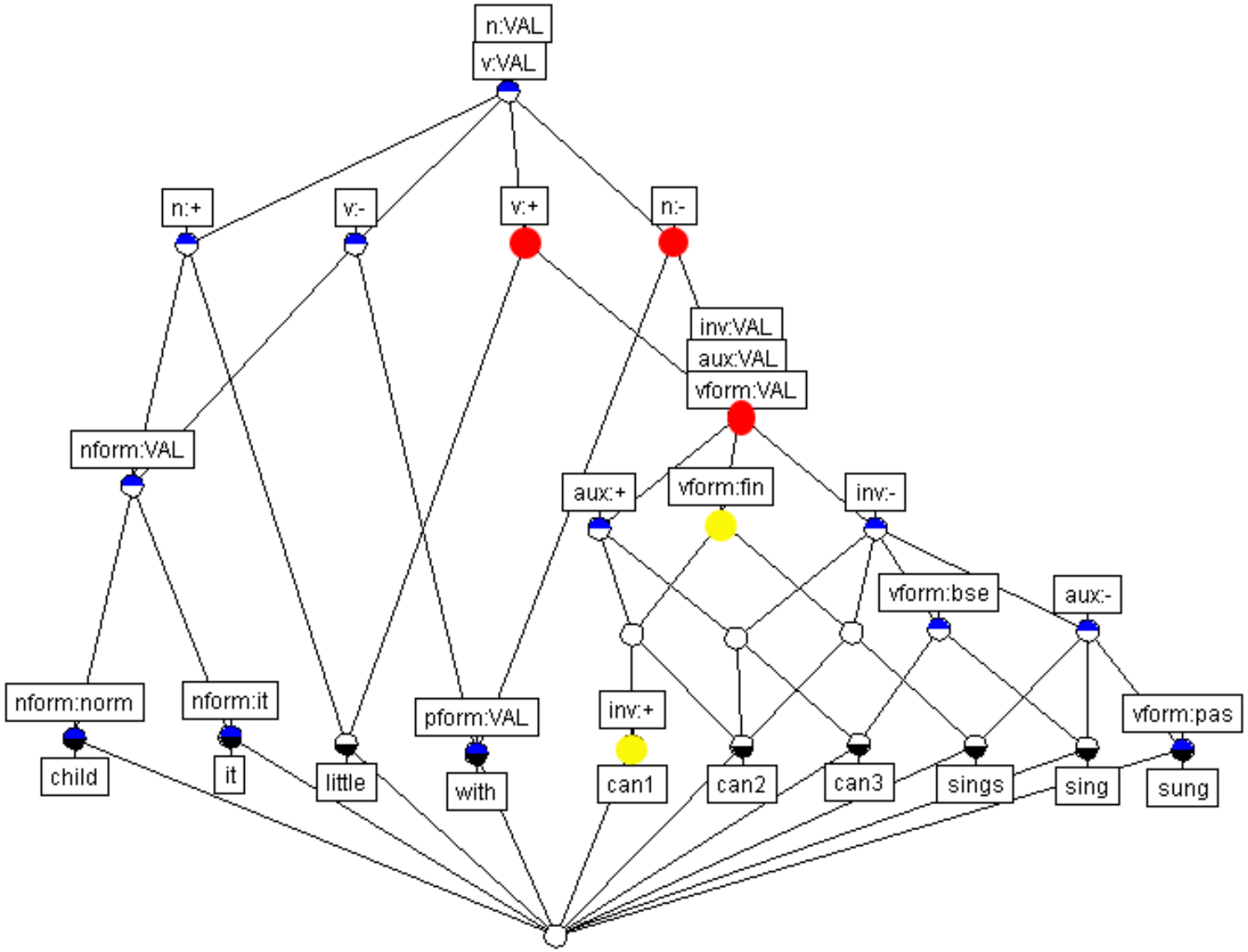
concept order: $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$.

attribute implications



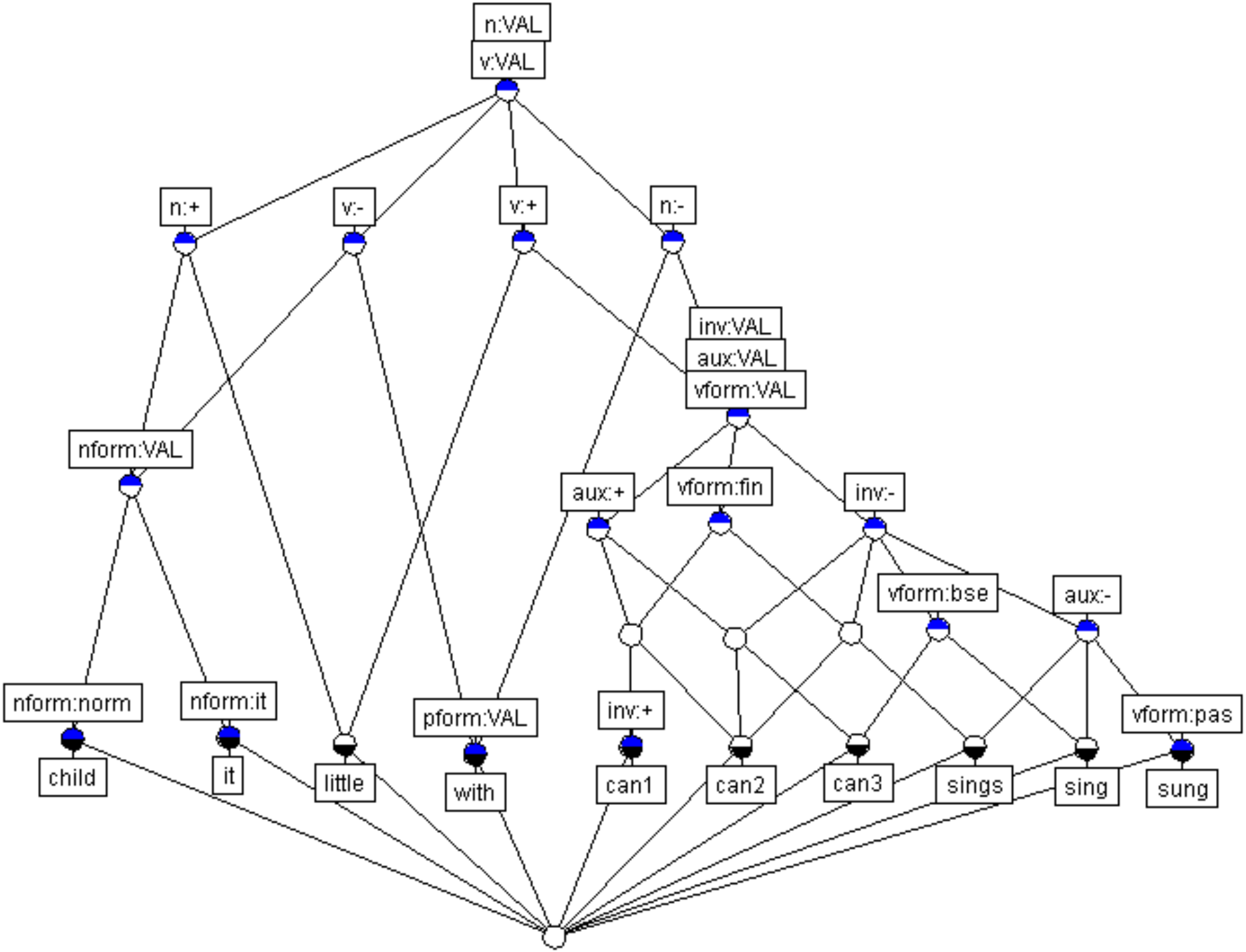
$$\{v:+, n:-\} \rightarrow \{vform:VAL\}, \quad \{inv:+\} \rightarrow \{vform:fin\}$$

attribute implications



$\{v : +, n : -\} \rightarrow \{vform : VAL\},$

$\{inv : +\} \rightarrow \{vform : fin\}$


n:VAL
v:VAL
n:+
v:-
v:+
n:-
inv:VAL
aux:VAL
vform:VAL
nform:VAL
aux:+
vform:fin
inv:-
vform:bse
aux:-
nform:norm
nform:it
pform:VAL
inv:+
vform:pas
child
it
little
with
can1
can2
can3
sings
sing
sung

basis of attribute implications

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

the first 4 FCRs from Gazdar et. al. 1985

$$\begin{array}{lrcr}
\text{FCR 1}: & [+\text{INV}] & \supset & [+\text{AUX}, \text{FIN}] \\
\text{FCR 2}: & [\text{VFORM}] & \supset & [+\text{V}, -\text{N}] \\
\text{FCR 3}: & [\text{NFORM}] & \supset & [-\text{V}, +\text{N}] \\
\text{FCR 4}: & [\text{PFORM}] & \supset & [-\text{V}, -\text{N}]
\end{array}$$

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$

sparse input data

$$
\begin{aligned}
\emptyset &\rightarrow \{v : VAL, n : VAL\} && (1)\\
\{v : -, n : -\} &\leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} && (2)\\
\{v : -, n : +\} &\leftrightarrow \{nform : VAL\} && (3)\\
\{v : +, n : -\} &\leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} && (4)\\
\{inv : +\} &\rightarrow \{vform : fin, aux : +\} && (5)\\
\{aux : -\} &\rightarrow \{inv : -\} && (6)\\
\{vform : pas\} &\rightarrow \{aux : -\} && (7)\\
\{vform : bse\} &\rightarrow \{inv : -\} && (8)\\
\{nform : it\} &\rightarrow \{nform : VAL\} && (9)\\
\{nform : nor\} &\rightarrow \{nform : VAL\} && (10)\\
\{inv : -\} &\rightarrow \{inv : VAL\} && (11)\\
\{aux : +\} &\rightarrow \{aux : VAL\} && (12)\\
\{vform : fin\} &\rightarrow \{vform : VAL\} && (13)\\
\{n : +, n : -\} &\rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot && (14)
\end{aligned}
$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$

value VAL

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$

excluding features

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$

gap: GPSG permits {[PAS], [+AUX], [−INV]}

$$\emptyset \rightarrow \{v : VAL, n : VAL\} \quad (1)$$

$$\{v : -, n : -\} \leftrightarrow \{pform : VAL\} \leftrightarrow \{pform : with\} \quad (2)$$

$$\{v : -, n : +\} \leftrightarrow \{nform : VAL\} \quad (3)$$

$$\{v : +, n : -\} \leftrightarrow \{vform : VAL\} \leftrightarrow \{aux : VAL\} \leftrightarrow \{inv : VAL\} \quad (4)$$

$$\{inv : +\} \rightarrow \{vform : fin, aux : +\} \quad (5)$$

$$\{aux : -\} \rightarrow \{inv : -\} \quad (6)$$

$$\{vform : pas\} \rightarrow \{aux : -\} \quad (7)$$

$$\{vform : bse\} \rightarrow \{inv : -\} \quad (8)$$

$$\{nform : it\} \rightarrow \{nform : VAL\} \quad (9)$$

$$\{nform : nor\} \rightarrow \{nform : VAL\} \quad (10)$$

$$\{inv : -\} \rightarrow \{inv : VAL\} \quad (11)$$

$$\{aux : +\} \rightarrow \{aux : VAL\} \quad (12)$$

$$\{vform : fin\} \rightarrow \{vform : VAL\} \quad (13)$$

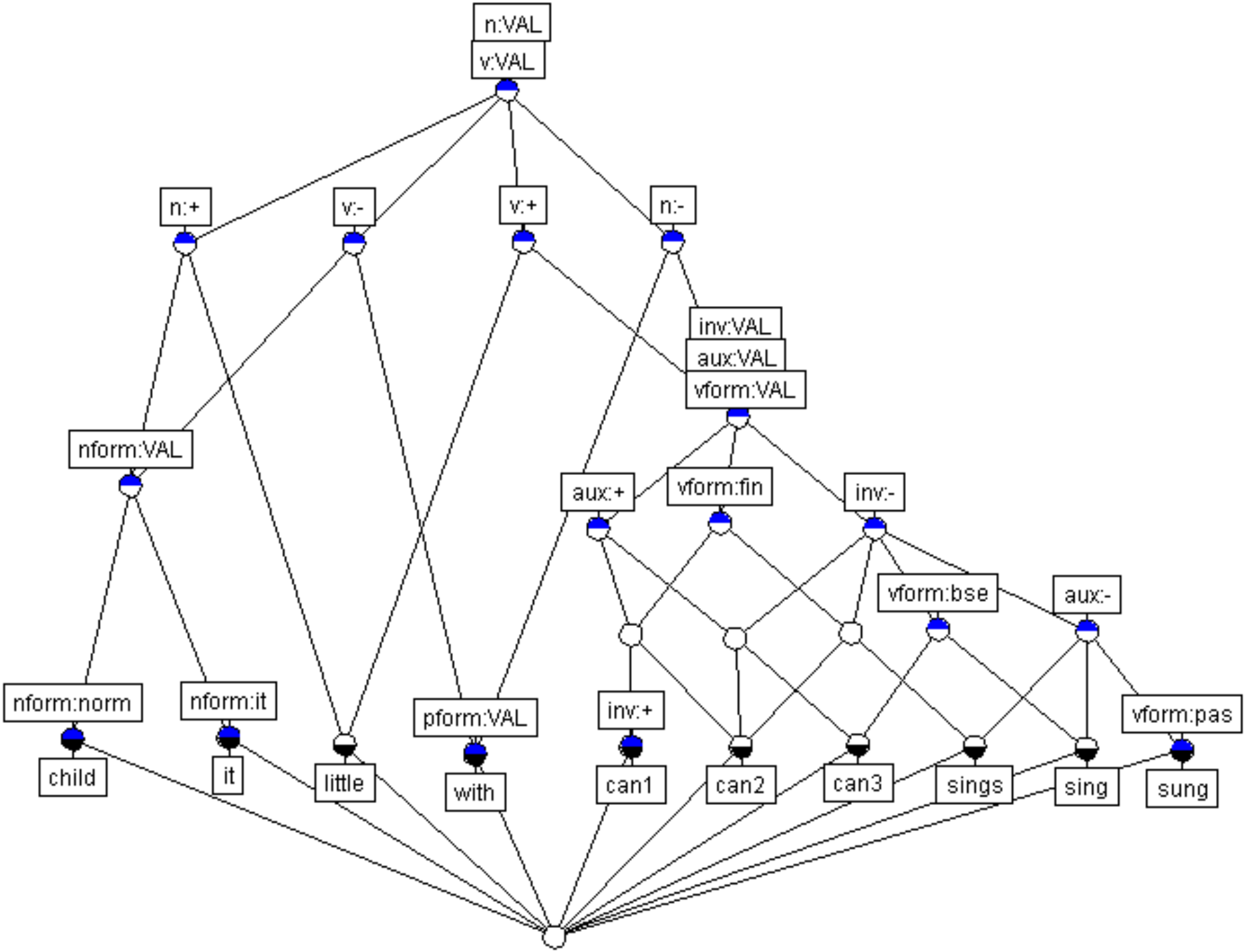
$$\{n : +, n : -\} \rightarrow \bot, \quad \ldots \qquad \{v : +, v : -\} \rightarrow \bot \quad (14)$$

FCR 1 : $[+INV] \supset [+AUX, FIN]$

FCR 2 : $[VFORM] \supset [+V, -N]$

FCR 3 : $[NFORM] \supset [-V, +N]$

FCR 4 : $[PFORM] \supset [-V, -N]$


n:VAL
v:VAL
n:+
v:-
v:+
n:-
inv:VAL
aux:VAL
vform:VAL
nform:VAL
aux:+
vform:fin
inv:-
vform:bse
aux:-
nform:norm
nform:it
pform:VAL
inv:+
vform:pas
child
it
little
with
can1
can2
can3
sings
sing
sung