

Coordination styles and transformations

Martin Popel

ÚFAL (Institute of Formal and Applied Linguistics)
Charles University in Prague

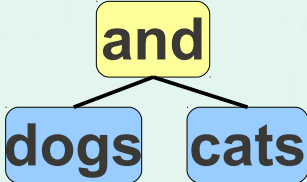
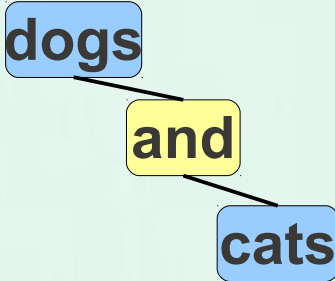


October 31st 2011, Prague



Outline

- Styles of annotating coordinations
 - Topological styles
 - Labeling styles
- Transformation of styles
- Evaluation



Participants of coordination

- **conjunct**
- **delimiter** (separates two conjuncts)
 - Coordinating conjunction
 - Comma or other punctuation (semicolon)
- **shared modifier** (modifies two or more conjuncts)

Examples:

- **lazy** **dogs** , **cats** **and** **rats** more than two conjuncts (“multi-conjunct c.”)
- **Mary** **came** **home** **and** **cried** *home* is a “private modifier”
- **John** **and** **Mary** **or** **Peter** embedded (nested) coordinations
- **big** **and** **cheap** **apples** **and** **oranges** coordinated shared modifier



Participants of coordination

- **conjunct**
- **delimiter** (separates two conjuncts)
 - Coordinating conjunction
 - Comma or other punctuation (semicolon)
- **shared modifier** (modifies two or more conjuncts)

Examples:

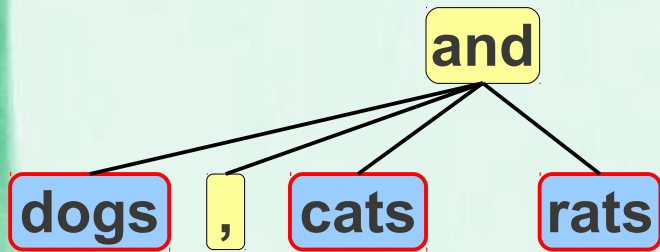
- **lazy** **dogs** , **cats** **and** **rats** more than two conjuncts (“multi-conjunct c.”)
- **Mary** **came** **home** **and** **cried** *home* is a “private modifier”
- **John** **and** **Mary** **or** **Peter** embedded (nested) coordinations
- **big** **and** **cheap** **apples** **and** **oranges** coordinated shared modifier
- Don't worry, be happy, keep smiling! As well as try hard to analyze it etc.



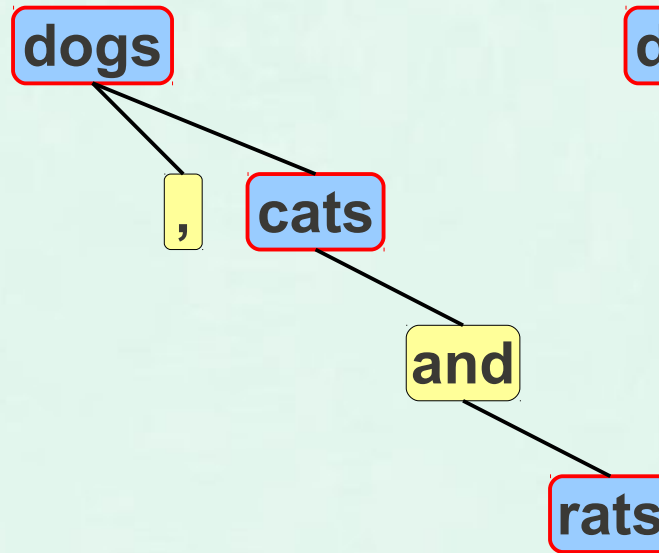
Topological styles (family)

Main “family” – configuration of conjuncts

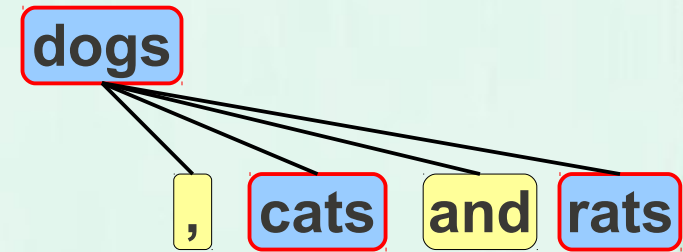
Prague



Moscow



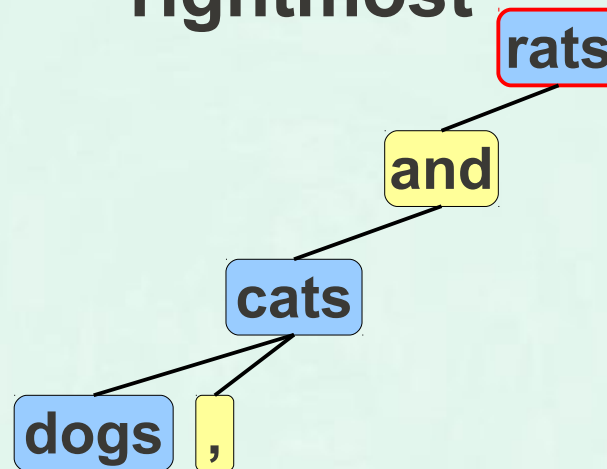
Stanford



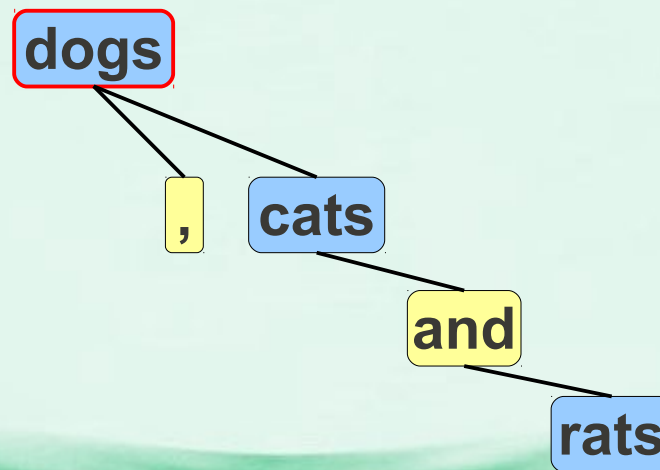
Topological styles (head)

Choice of head (which delimiter/conjunct to choose):

rightmost



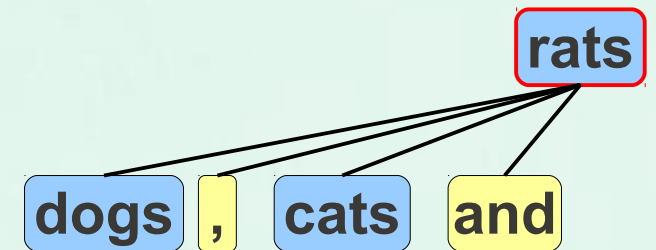
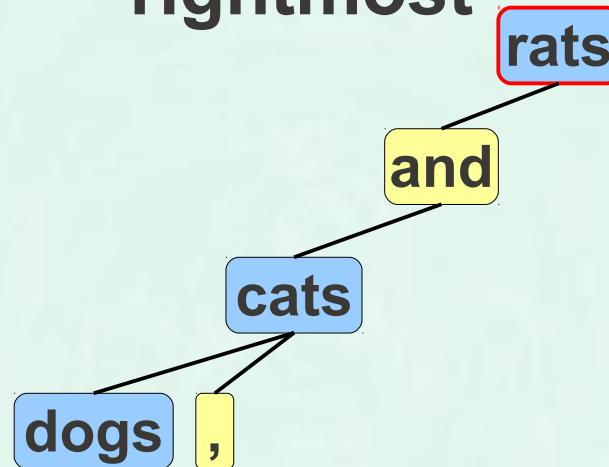
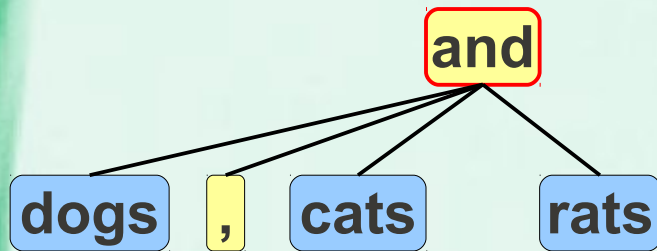
leftmost



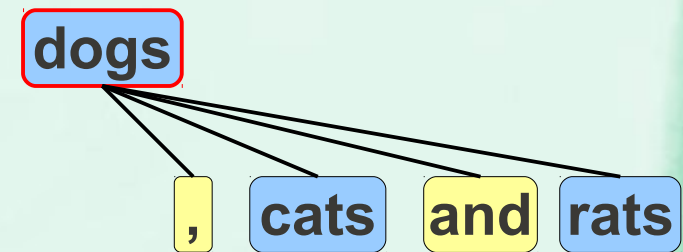
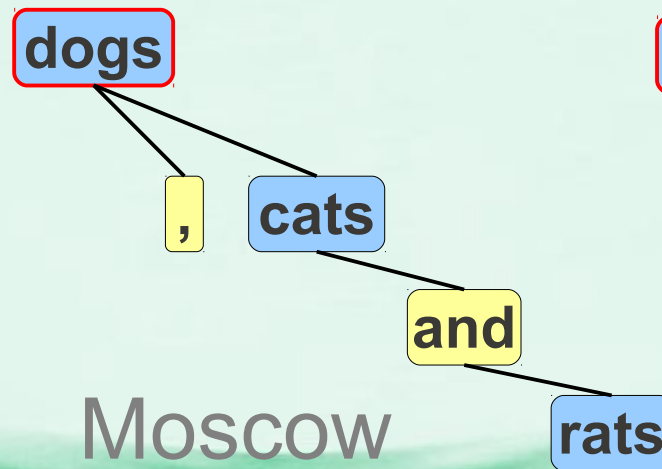
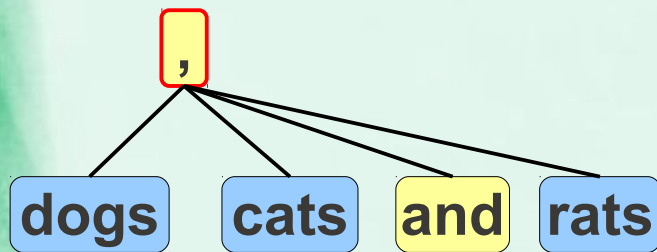
Topological styles (head)

Choice of head (which delimiter/conjunct to choose):

rightmost



leftmost



Prague

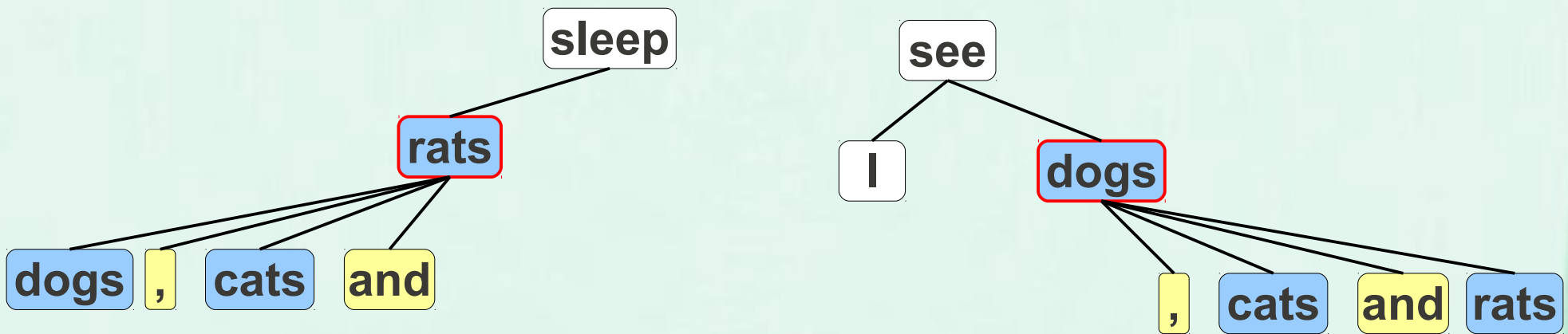
Moscow

Stanford



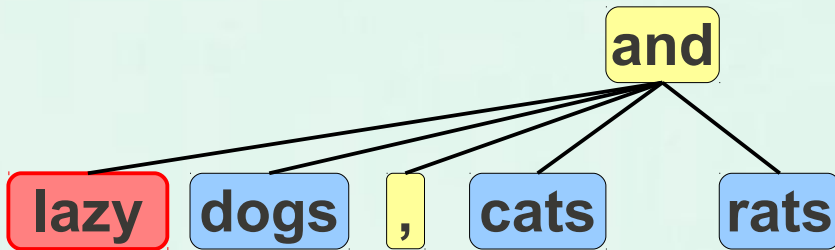
Topological styles (head)

Choice of head: leftmost, rightmost or
mixed

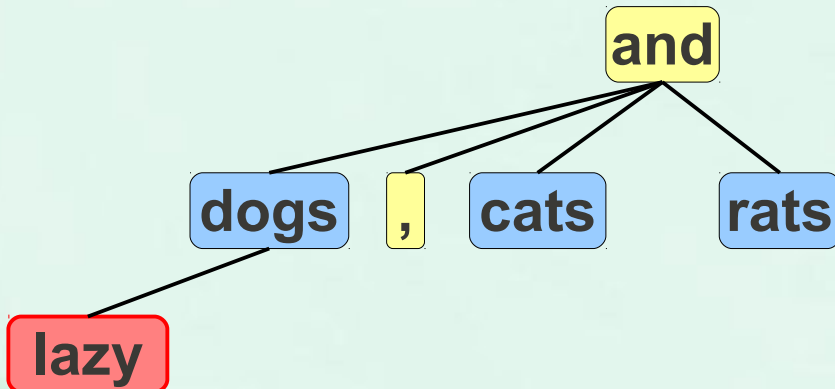


Topological styles (shared modifiers)

Attachment of shared modifiers:
below **the head**



below **the nearest conjunct**



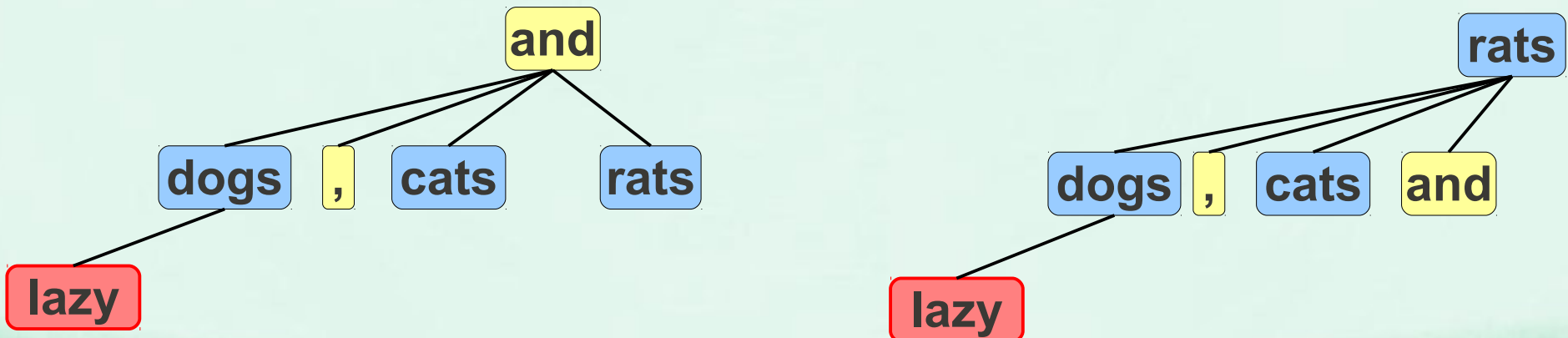
Topological styles (shared modifiers)

Attachment of shared modifiers:

below the head



below the nearest conjunct



Prague

Stanford



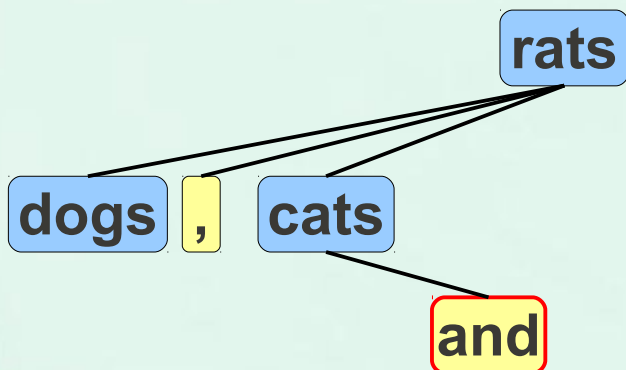
Topological styles (conjunction)

Attachment of coordinating conjunctions:

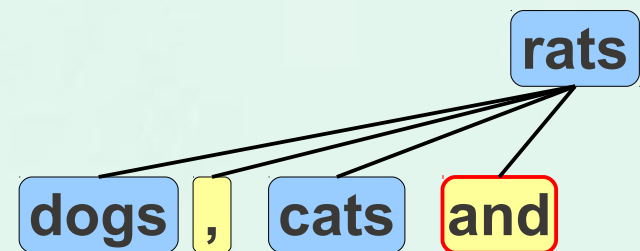
“between” conjuncts



below the previous conjunct



following conjunct



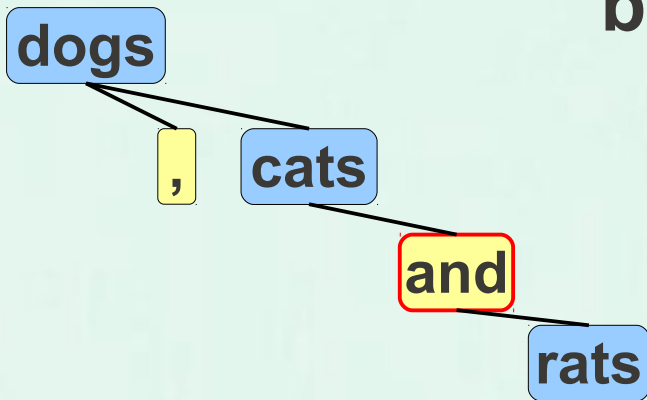
Stanford, head=rightmost



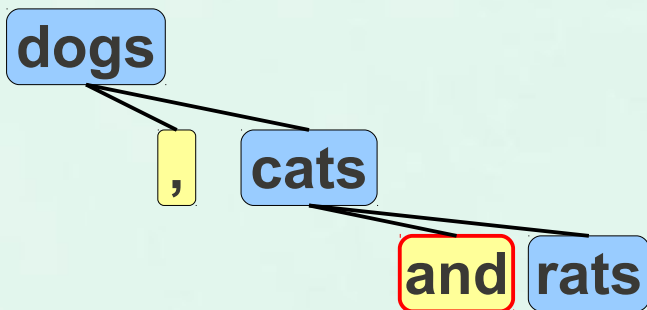
Topological styles (conjunction)

Attachment of coordinating conjunctions:

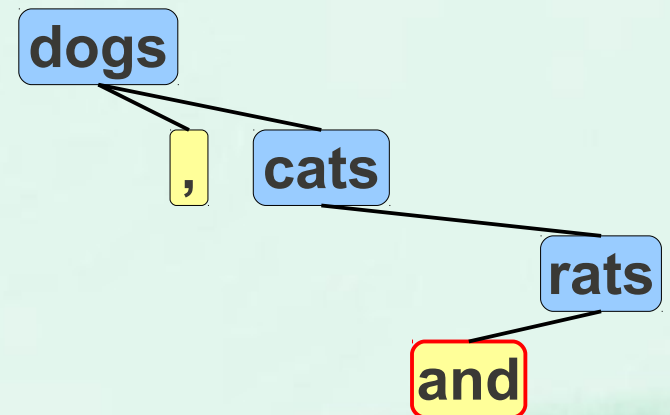
“between” conjuncts



below the previous conjunct



following conjunct



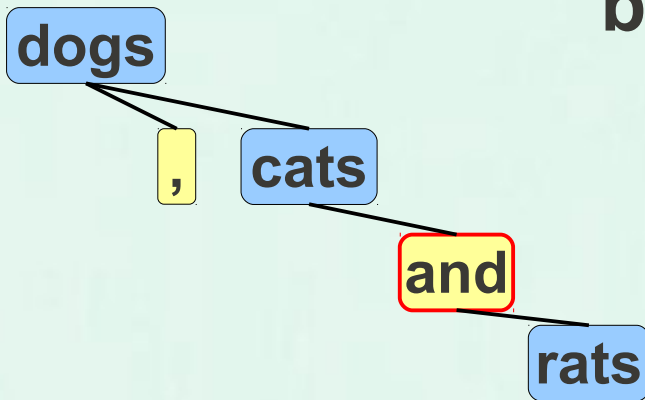
Moscow, head=leftmost



Topological styles (conjunction)

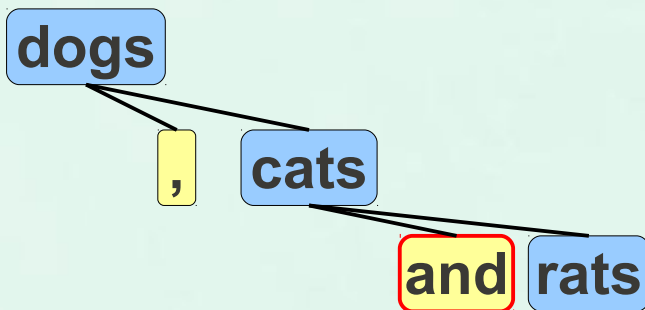
Attachment of coordinating conjunctions:

“between” conjuncts

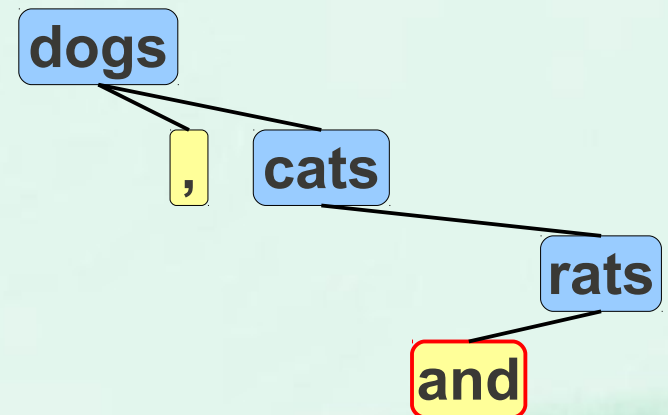


“as the head”
for Prague (the only applicable)

below the previous conjunct



following conjunct



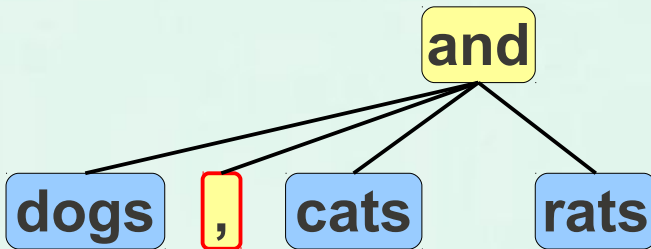
Moscow, head=leftmost



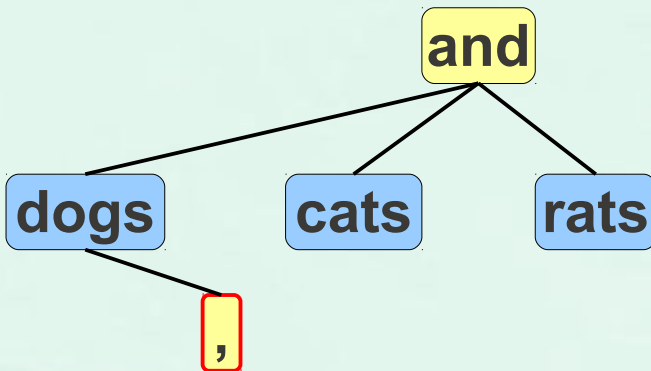
Topological styles (punctuation)

Attachment of punctuation delimiters:

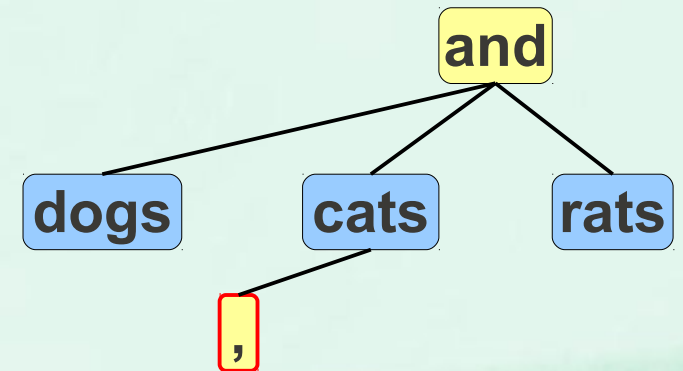
“between” conjuncts



below the **previous conjunct**



following conjunct



Prague



Topological styles – overview

How many treebanks use the given style?

- **Family** (Prague=13, Moscow=5, Stanford=6)
- **Head** (Leftmost=11, Rightmost=13, Mixed=0)
- **Shared m.** (below Head=7, Nearest conjunct=?)
- **Conjunctions** (Previous=2, Following=1, Between=8, as Head=13)
- **Punctuation** (Previous=7, Following=1, Between=14)

How many possible styles?

more than $2*3*2*3*3+1*3*2*1*3 = 126$

* labeling variants



- Parsers can produce labeled edges (nodes): Sb, Obj, Atr, ...
 - We can define special labels for coordinations (COORD, CC, ...)
 - or encode additional attributes (*is_member*) into the dependency label: Sb_M, Obj_M, Atr_M, ...
 - Some additional attributes can be deduced or guessed.
- Should we mark shared modifiers?
- Should we mark conjunct (except if head)?
- Should we mark nested coordinations (co-index modifiers)?
- In Prague styles, where to store the dependency relation of the whole coordination?
 - with conjuncts (they can have different labels in PDT)
 - with the head.



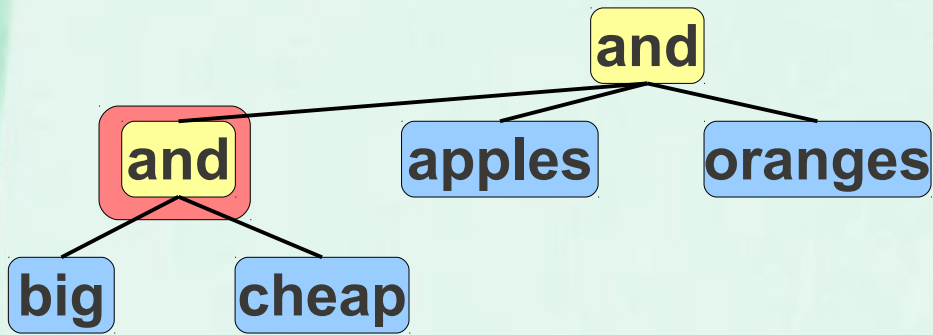
Subtasks

- 1) Detect coordinations in a sentence
(esp. boundaries of nested coordinations)
- 2) Classify participants of coordinations
(conjunct, commas, conjunctions, shared m.)
- 3) Transform each coordination to the target style

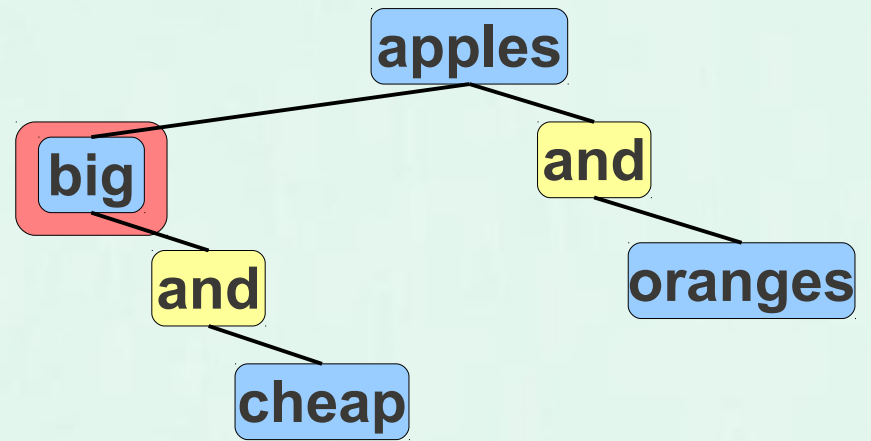


Problematic cases

big and cheap apples and oranges



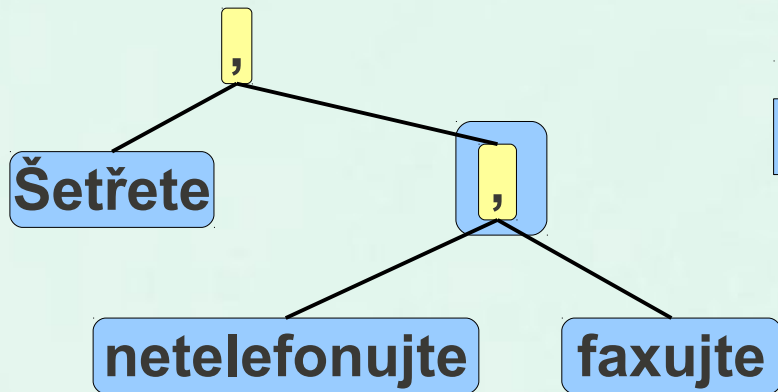
Prague



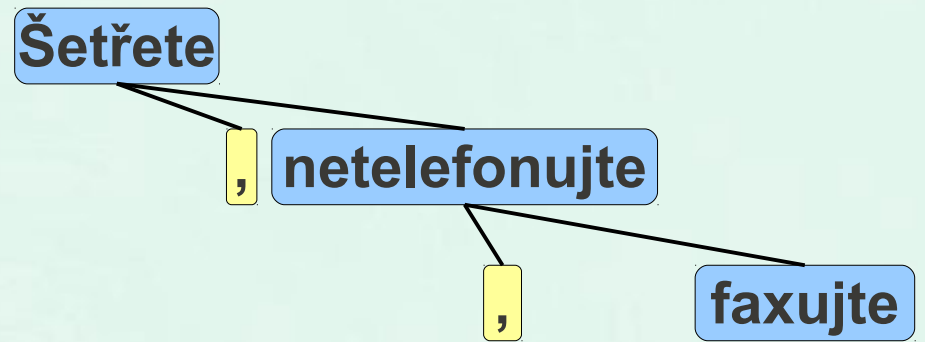
Moscow



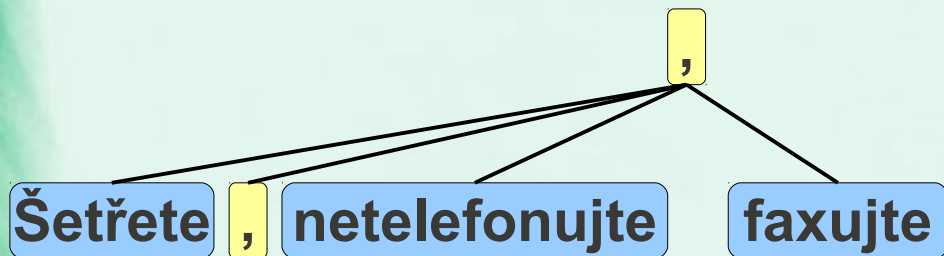
Problematic cases



Prague



Moscow



"Save money, don't phone, use fax."

PDT 2.0



Evaluation

- Transform both train and test set
 - train the parser on the transformed train set
 - evaluate on the transformed test set
- Transform only the train data
 - train the parser on the transformed train set
 - apply inverse transformation
 - evaluate on the original test set



cats
and
dogs

dogs
and
cats

Preliminary results

Lang.	orig	fP (PDT) hR sH cH pB	fM hL sN cB pP	fM hM sN cB pP	fM hR sH cB pP	fP hR sN cB pP	fS hL sN cB pP	fS hM sN cB pP	fS hR sH cB pP	fS hR sN cB pP	
ar	72.50	69.30±1.50	2.00	1.00	0.90	1.60	-0.10	1.30	1.60	1.10	1.20
	72.20	72.20±1.80	0.80	0.20	0.60	0.60	-0.40	-0.70	-0.70	1.20	1.20
bg	88.10	80.50±1.30	-0.50	-0.80	0.20	0.40	0.10	-1.10	-2.00	-0.40	-0.30
	86.30	79.00±1.90	0.50	-0.10	0.20	0.30	-0.40	0.20	-0.20	0.90	0.90
bn	80.30	78.50±2.80	-0.40	0.20	0.50	-0.60	1.50	1.00	-0.20	-0.80	0.30
	81.60	81.10±2.80	-0.60	-0.50	0.10	-0.50	-0.90	-0.40	-0.40	-0.10	0.50
cs	75.40	75.10±1.90	-1.20	-2.40	-1.00	0.00	-2.40	-3.80	-4.50	-2.40	-1.70
	68.60	68.90±2.70	1.70	0.90	-0.40	0.30	-2.20	-0.10	-0.20	0.40	1.10
da	88.10	81.40±1.50	-0.60	-1.30	-0.60	0.00	-1.30	-2.30	-2.90	-1.70	-1.10
	84.30	75.70±1.40	1.40	0.40	-1.00	-0.90	-1.40	-0.20	-0.50	-0.40	-0.20
de	88.50	82.90±0.70	-0.30	-0.90	-0.30	0.00	-0.40	-0.90	-1.50	-0.60	-0.70
	81.50	74.90±0.80	1.20	0.50	0.30	0.40	-0.40	0.50	-0.10	0.40	0.50
el	73.60	74.10±1.80	0.40	-0.10	0.40	0.40	-1.50	-1.60	-1.80	-1.10	-0.50
	72.80	72.40±1.40	1.20	0.10	0.20	0.70	-0.70	-0.60	-0.80	0.50	0.70
en	90.90	85.80±0.90	-0.90	-1.40	-1.00	-1.00	-1.20	-1.80	-2.20	-1.50	-1.30
	86.20	79.40±1.00	0.30	-0.20	-0.20	-0.20	-1.00	-0.10	-0.50	0.40	0.40
es	88.00	84.20±0.80	-0.70	-1.20	-0.80	-0.80	-1.00	-2.10	-2.40	-1.30	-1.20
	83.90	79.10±1.00	0.70	0.40	-0.30	-0.30	-1.50	-0.40	-0.70	0.10	0.00
eu	76.20	66.00±1.40	-1.30	-2.50	-1.70	-1.80	-2.60	-2.70	-3.50	-3.00	-3.30
	71.80	60.10±1.60	0.50	-0.70	0.30	0.30	-2.70	-0.80	-1.50	-0.50	-0.70
fi	72.20	69.00±1.20	-1.70	-2.50	-2.00	-1.90	0.40	-3.80	-4.50	-2.40	-3.50
	70.00	64.80±1.80	2.60	1.40	0.20	0.00	-0.20	1.30	0.70	0.50	0.50
grc	56.20	55.10±1.60	-1.60	-0.90	-0.60	-0.60	-1.10	-1.70	-2.00	-1.20	-1.20
	42.50	43.40±1.80	2.40	2.00	2.70	2.60	-1.20	2.10	2.60	1.60	2.10
hi	76.90	71.40±1.60	1.10	1.70	0.90	0.60	0.00	1.40	1.70	0.80	0.60
	86.60	81.90±1.90	0.20	0.00	0.10	0.00	-1.00	-0.20	-0.10	-0.30	-0.40
hu	80.40	76.10±1.90	-1.50	-2.00	-1.80	-1.70	-1.70	-2.40	-2.50	-1.80	-1.70
	76.10	71.50±1.90	0.00	-0.50	-0.70	-0.70	-0.70	-0.70	-0.60	-0.50	-0.40
it	85.00	79.60±2.40	-1.20	-1.50	-1.40	-1.20	-1.60	-2.40	-2.40	-1.60	-1.90
	83.20	76.30±2.20	0.30	-0.20	-0.40	-0.20	-0.80	-0.10	-0.50	0.30	0.30
la	56.30	54.80±2.30	1.60	0.70	1.50	2.90	-0.60	-0.50	-1.20	-0.60	0.50
	44.90	42.40±1.70	5.30	2.60	3.30	5.50	-0.70	3.80	3.10	2.00	2.40
nl	83.80	78.60±1.50	-0.90	-1.60	-0.60	-0.30	-1.30	-3.80	-3.90	-1.60	-1.20
	75.10	70.00±2.00	0.50	-0.70	-0.40	-0.30	-1.70	-1.40	-2.20	-0.60	-0.70
pt	87.80	82.00±1.40	-0.30	-0.50	0.00	-0.30	-0.90	-1.00	-1.10	-0.20	-0.40
	85.40	77.80±2.10	0.10	-0.20	-0.20	-0.40	-1.60	-0.40	-0.40	-0.20	-0.40
ro	88.30	88.80±1.60	-0.90	-0.10	-0.20	-0.20	0.00	-1.20	-1.60	0.00	0.00
	86.20	86.50±1.70	-0.10	0.00	-0.50	-0.50	-0.20	-0.50	-1.00	0.00	0.00
ru	?	78.10±1.60	1.80	0.40	0.60	0.60	0.70	1.40	1.70	1.00	1.30
	58.90	84.40±1.80	0.50	-0.10	-0.30	-0.30	-0.20	-0.30	0.50	-0.10	0.00
sl	75.30	74.10±1.50	-0.20	-0.70	-0.80	-0.30	-0.60	-2.30	-2.70	-1.70	-1.60
	71.50	68.60±1.60	2.10	1.70	1.20	1.50	-0.80	0.60	0.20	0.70	0.90
sv	87.10	78.50±1.70	-0.10	-0.80	-0.90	-0.70	-1.50	-2.20	-2.80	-1.90	-1.70
	88.20	76.60±1.50	0.20	-0.60	-1.00	-0.70	-2.40	-1.60	-1.70	-0.90	-1.20
ta	69.40	71.60±2.00	0.40	0.30	-0.60	0.30	0.00	-0.90	-0.40	0.70	0.30
	71.40	72.80±2.70	1.10	1.20	0.40	1.00	-0.10	0.20	1.30	1.10	1.60
te	86.90	87.20±3.70	1.20	-0.90	-1.20	-0.90	-0.10	0.10	-0.40	-0.80	-1.60
	87.30	88.00±3.50	2.60	2.30	0.40	0.50	0.00	2.30	1.60	0.50	0.70
tr	78.30	76.30±1.90	-1.70	-1.00	-1.00	-0.90	-1.00	-1.40	-1.30	-1.80	-1.90
	72.70	72.10±2.10	0.10	-0.30	-0.30	-0.30	-0.40	0.00	-0.40	-0.60	-0.50
Aver.	76.22	75.96	-0.30	-0.75	-0.46	-0.26	-0.73	-1.39	-1.71	-0.99	-0.90
	75.57	72.80	1.02	0.38	0.17	0.34	-0.94	0.10	-0.10	0.26	0.37
Significantly positive change	2	1	?	?	?	?	?	3	?	?	?
	5	3	2	2	?	2	2	2	1	2	2
Insignificant change	21	16	22	22	20	21	13	8	16	20	20
	20	22	23	23	23	22	22	21	24	23	23
Significantly negative change	2	8	3	3	4	12	14	9	5	?	?
	?	?	?	?	3	1	2	?	?	?	?

Questions?



Thank you.

