

XV UNL School

Geneva, July 21-25, 2014

Program

- July 21st
 - Introduction
 - Normalization
- July 22nd
 - Tokenization
 - corpus_org.txt
 - ngrammar.txt
 - dic.txt
 - dgrammar.txt
 - morphology.txt
- July 23rd
 - UNLization
- July 24th
 - NLization
- July 25th
 - Evaluation & Discussion



Day #3

Morning

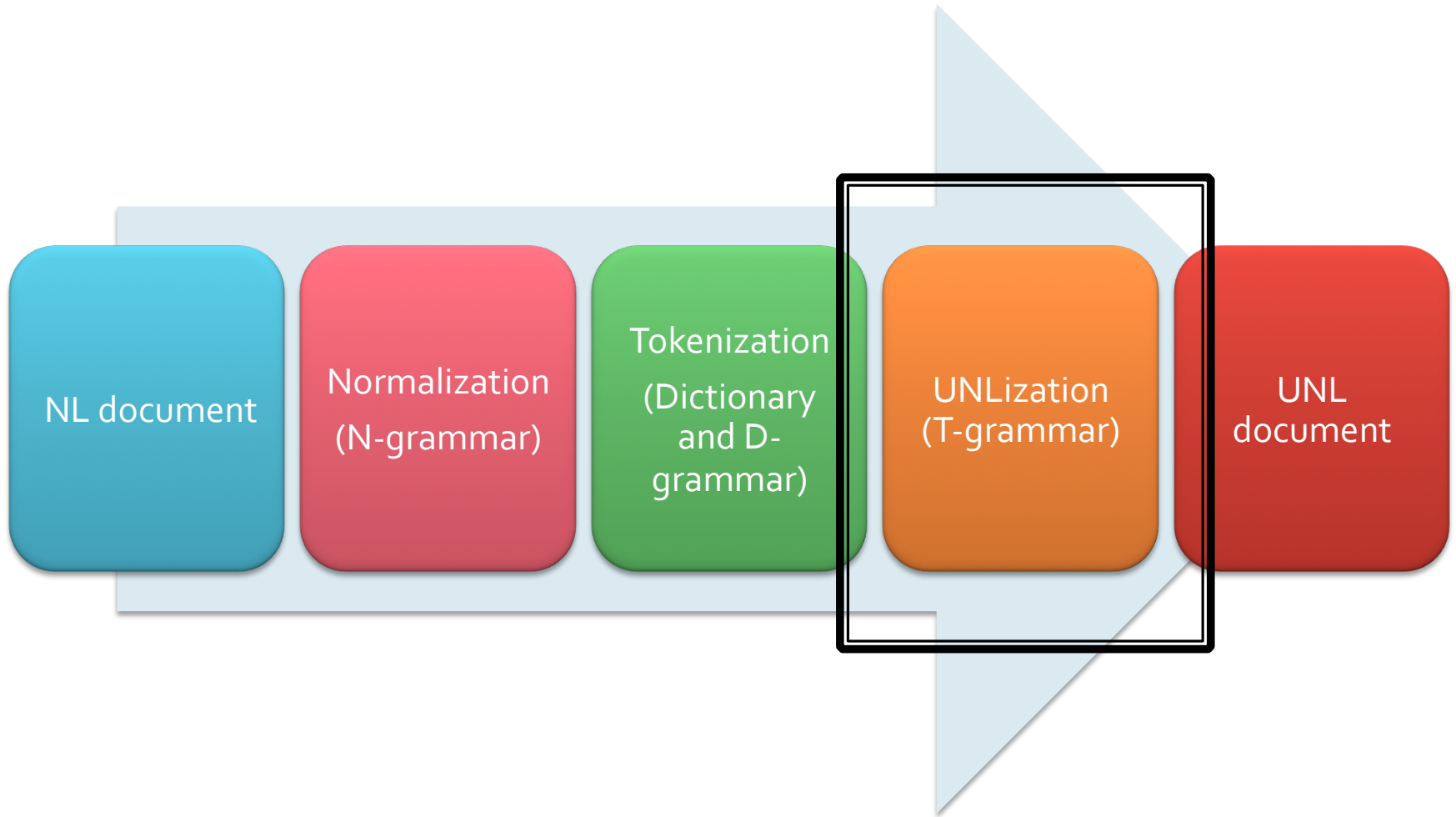
- Syntax

Lunch break

Afternoon

- Semantics

Goal



Syntax

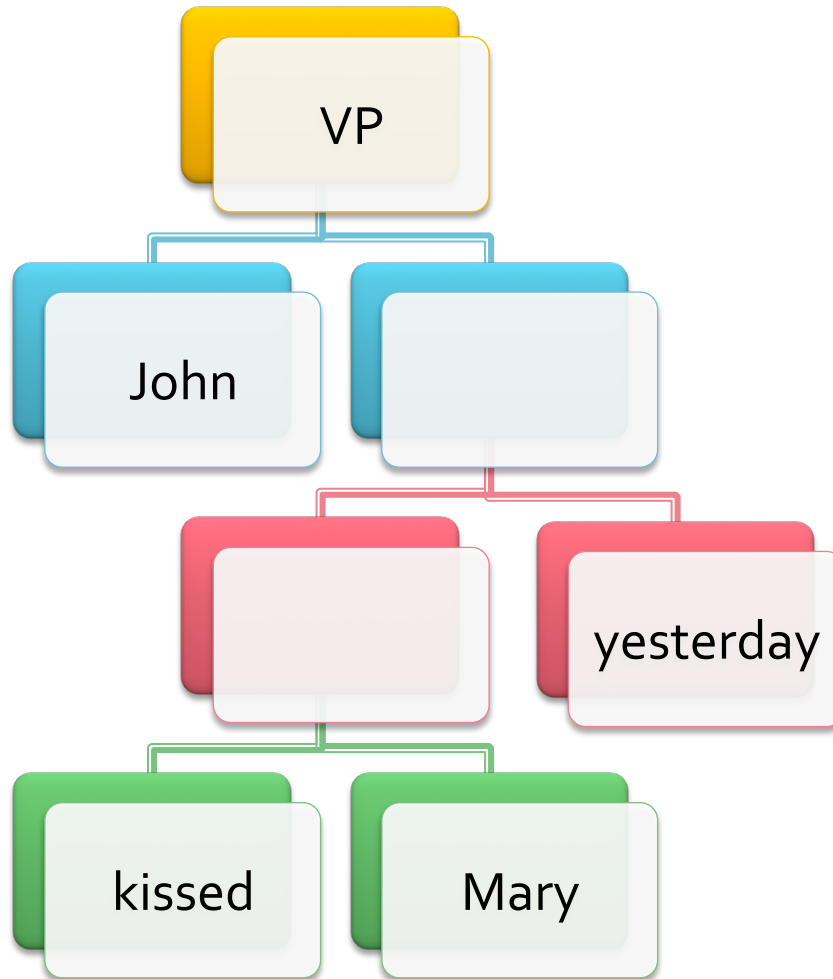
Grammar frameworks

- DCG (Definite Clause Grammar)
- GPSG (Generalized Phrase Structure Grammar)
- HPSG (Head-driven Phrase-Structure Grammar)
- TAG (Tree Adjoining Grammar)
- UG (Unification Grammar)
- CG (Categorial Grammar)
- FUG (Functional Unification Grammar)
- SFG (Systemic functional grammar (SFG))
- LFG (Lexical-functional Grammar)
- Generative Grammar
 - ST (Standard Theory)
 - EST (Extended Standard Theory)
 - X-bar
 - GB (Government and Binding)
 - PP (Principles and Parameters)
 - Minimalist Theory

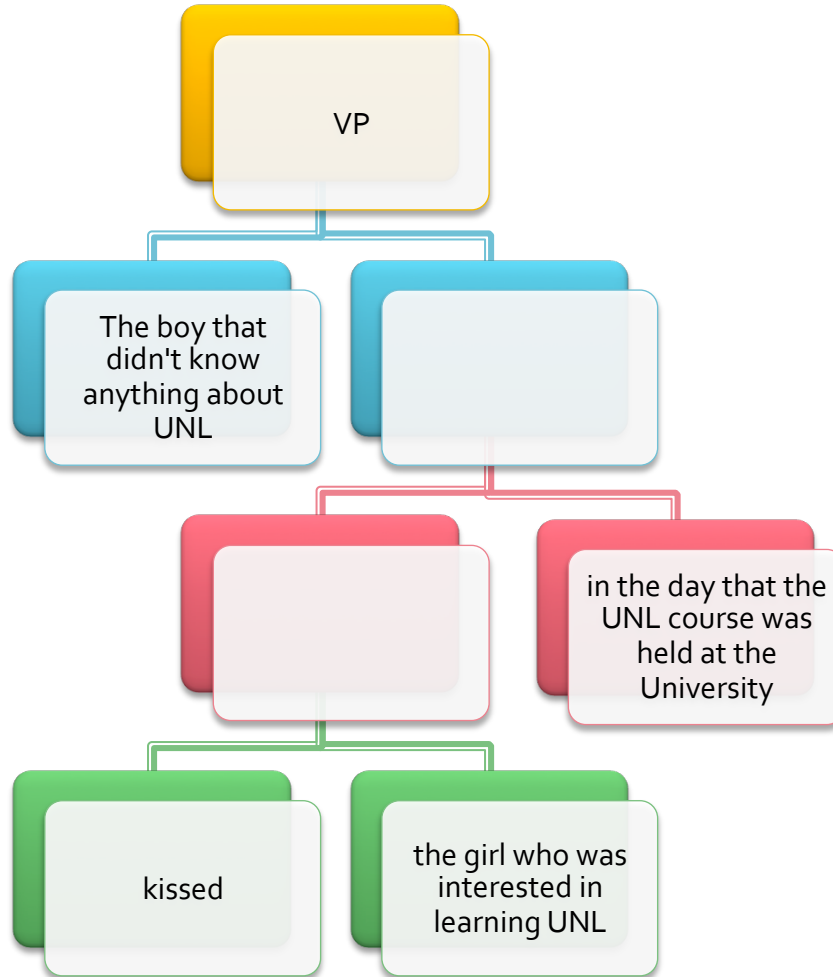
X-bar

(adapted and simplified)

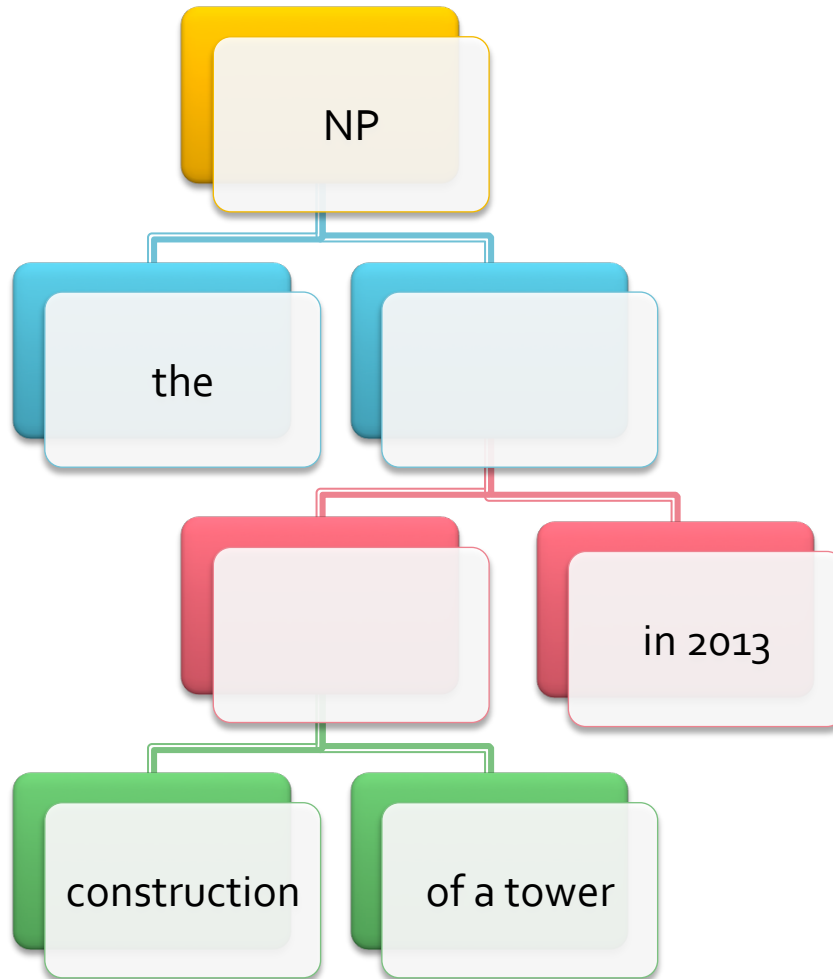
John kissed Mary yesterday



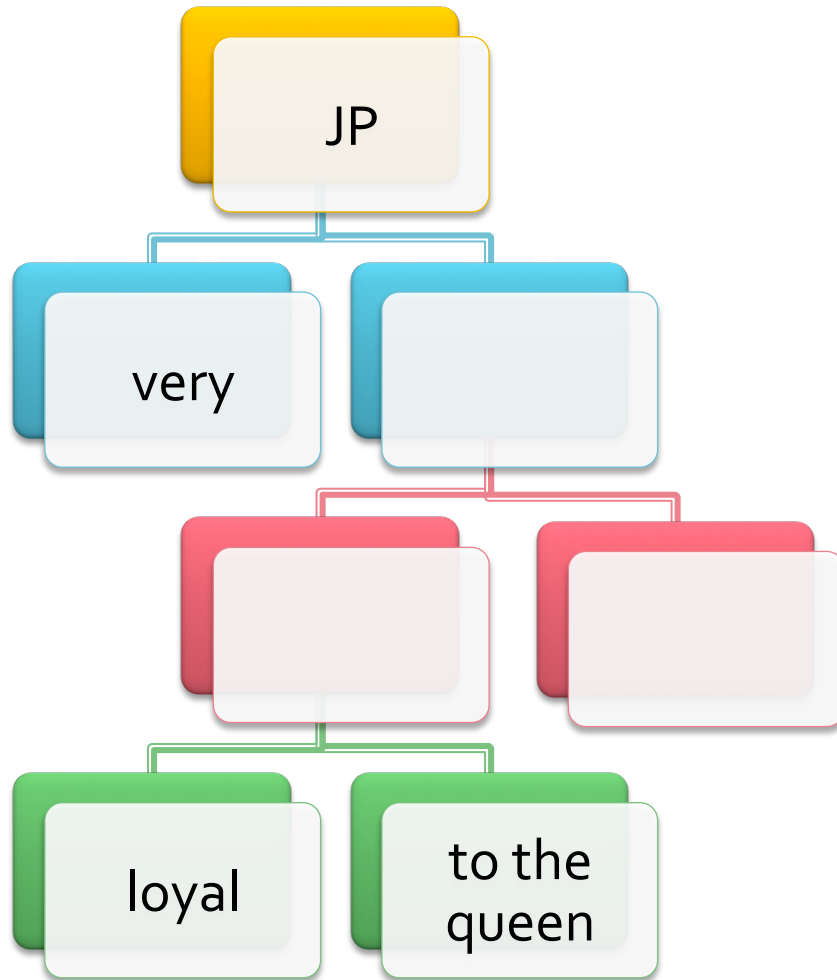
The boy that didn't know anything about UNL kissed the girl who was interested in learning UNL in the day that the UNL course was held at the University.



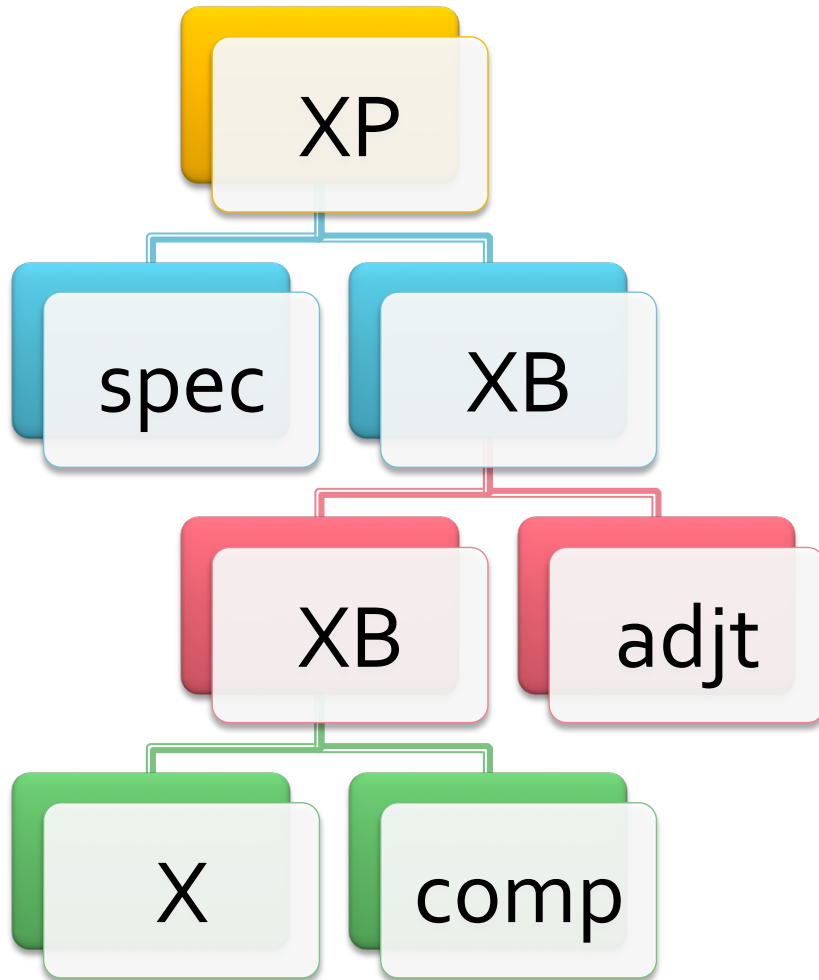
The construction of a tower in 2014



very loyal to the queen

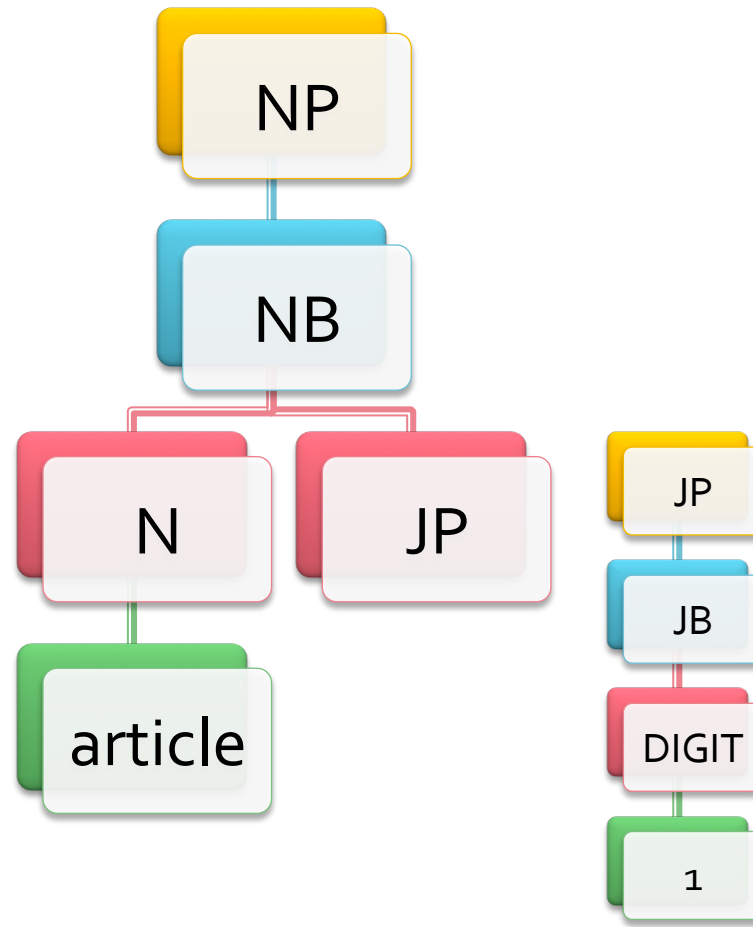


X-bar structure

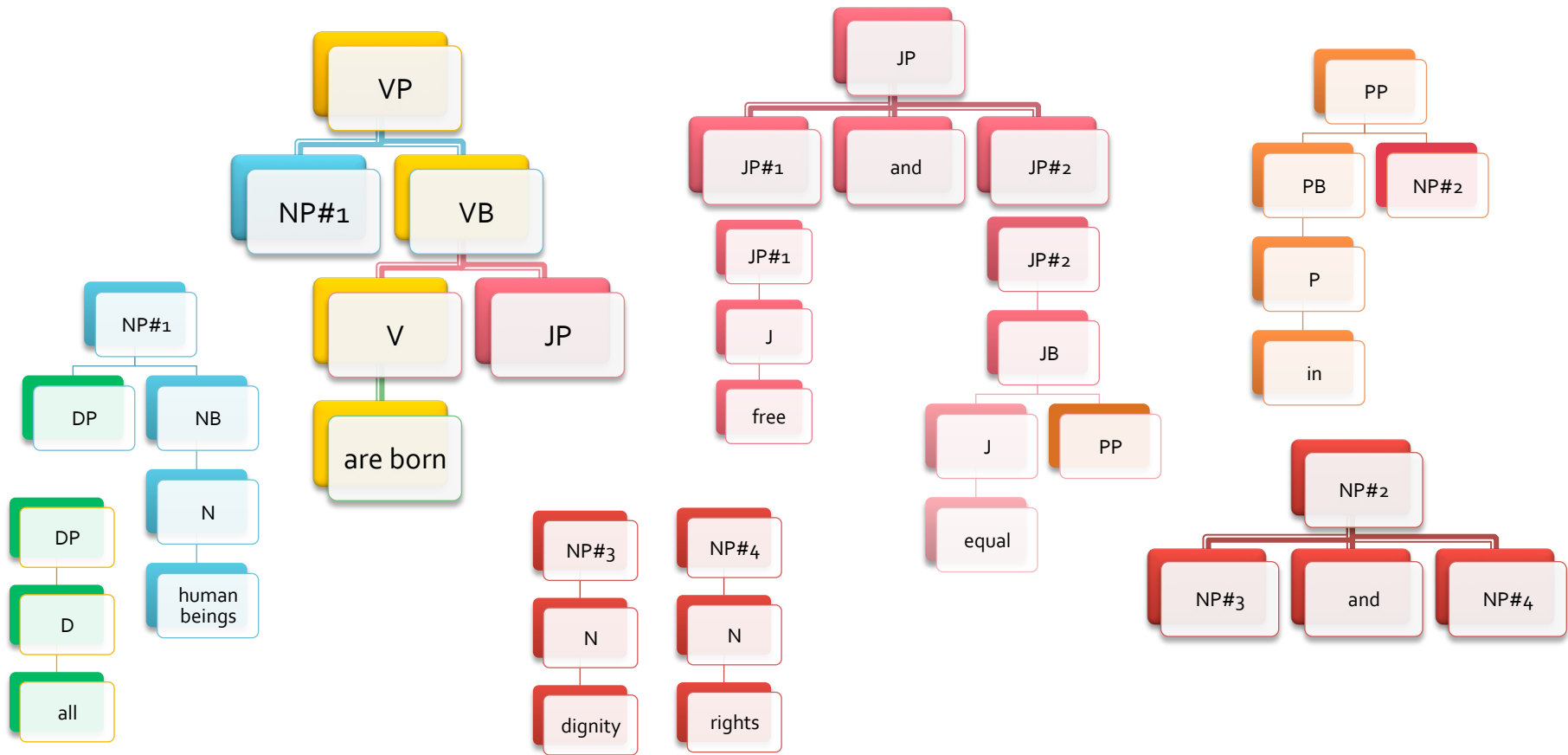


- Where:
 - XP = maximal projection
 - XB = intermediate projections
 - spec = specifier
 - adjt = adjunct
 - comp = complement
 - X = head
 - N (noun)
 - V (verb)
 - J (adjective)
 - A (adverb)
 - D (determiner)
 - P (preposition)
 - C (conjunction)

Article 1



All human beings are born free and equal in dignity and rights.



Together

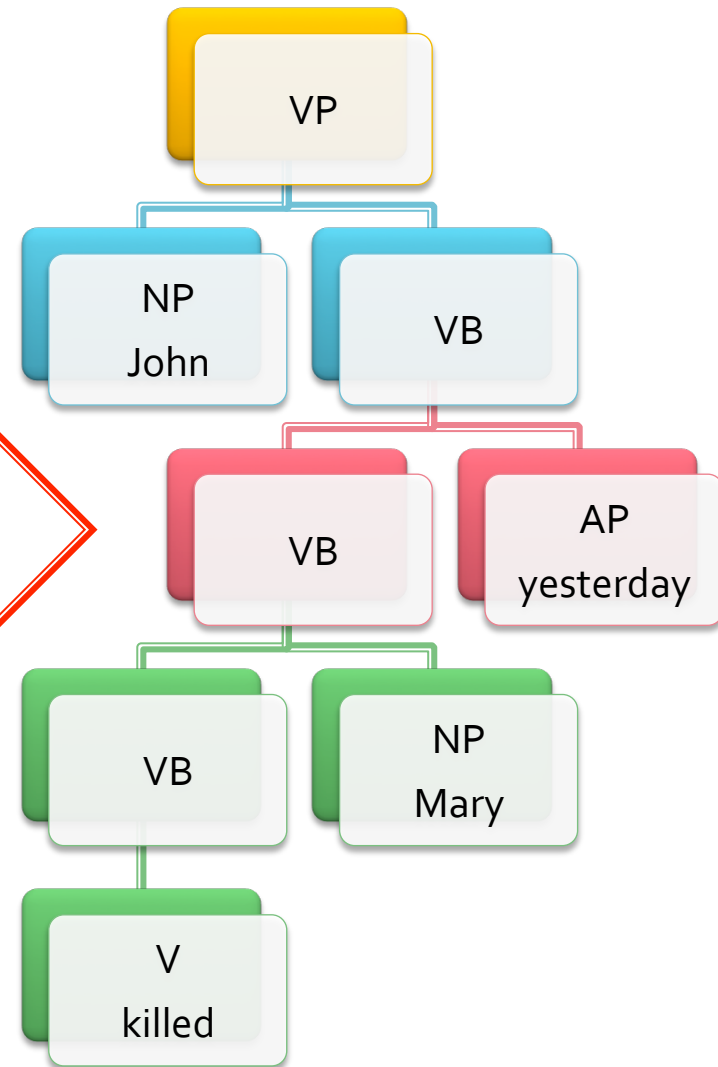
- They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

Exercise #8 (60 min)

- Goal: To create the tree bank for the training corpus
- Deliverable:
 - *trees*
- Activities:
 - www.unlweb.net/wiki/XV_UNL_School

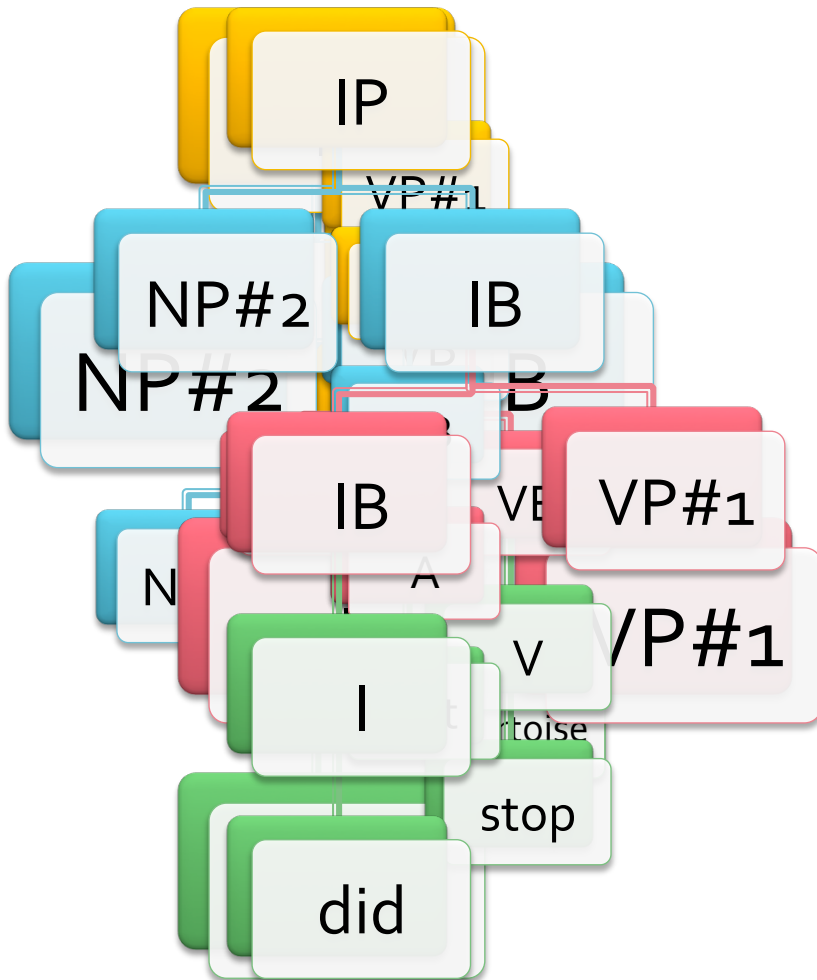
Parsing

Parsing list > tree



Building a grammar (I)

building a tree bank



Building a grammar (III)

inverting the rules

(NP):=(NP)("and")(NP);
(NP):=(DP)(NB);
(NB):=(N);
(DP):=(DB);
(DB):=(D);
(IP):=(NP)(IB);
(IB):=(VP);
(VP):=(VB);
(VB):=(V);
(IB):=(IB)(VP);
(IB):=(I);
(VB):=(AP)(VB);
(AP):=(AB);
(AB):=(A);

(NP)("and")(NP):=(NP);
(DP)(NB):=(NP);
(N):=(NB);
(DB):=(DP);
(D):=(DB);
(NP)(IB):=(IP);
(VP):=(IB);
(VB):=(VP);
(V):=(VB);
(IB)(VP):=(IB);
(I):=(IB);
(AP)(VB):=(VB);
(AB):=(AP);
(A):=(AB);

Building a grammar (IV)

ordering the rules (longest first)

(NP)("and")(NP):=(NP);
(DP)(NB):=(NP);
(N):=(NB);
(DB):=(DP);
(D):=(DB);
(NP)(IB):=(IP);
(VP):=(IB);
(VB):=(VP);
(V):=(VB);
(IB)(VP):=(IB);
(I):=(IB);
(AP)(VB):=(VB);
(AB):=(AP);
(A):=(AB);

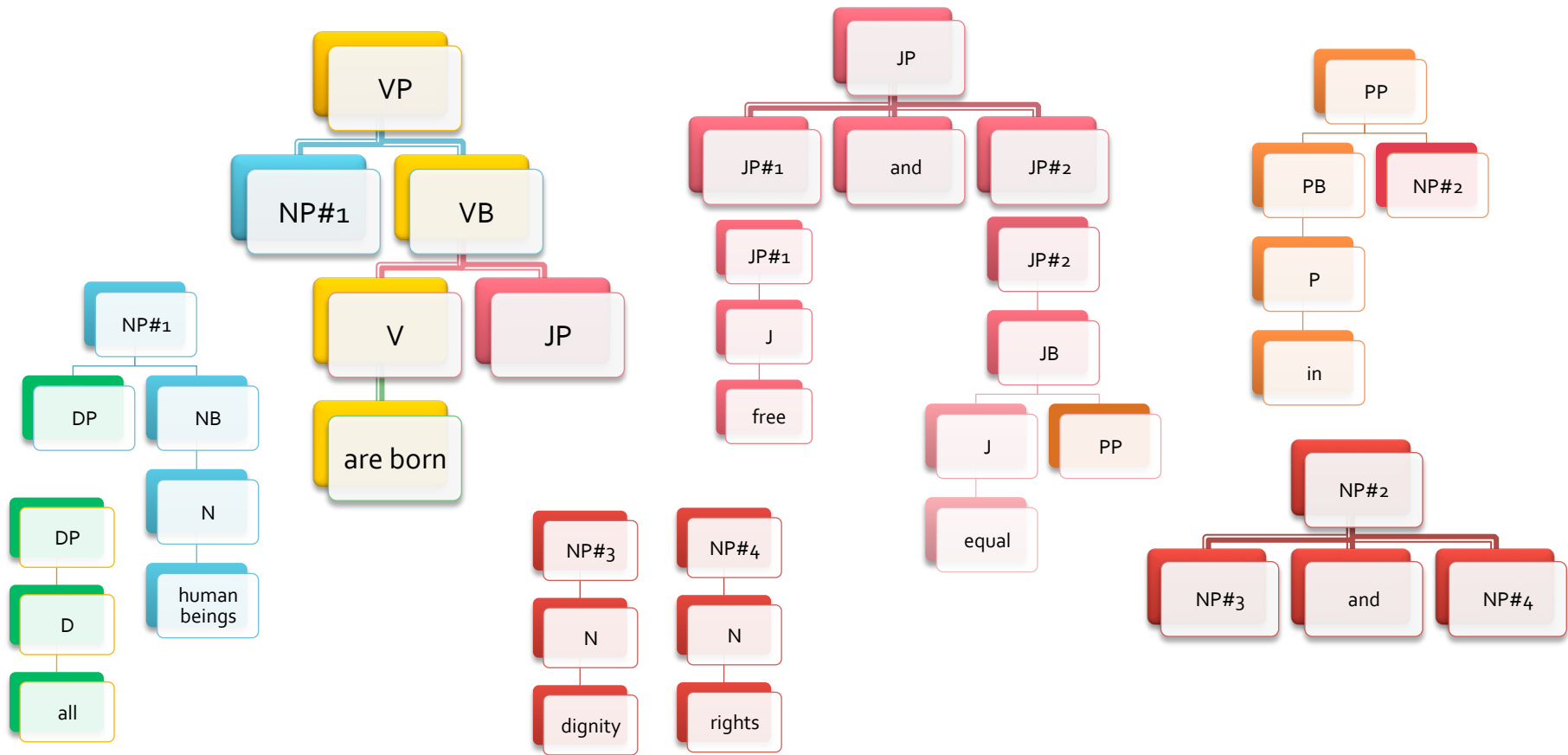
(NP)("and")(NP):=(NP);
(AP)(VB):=(VB);
(DP)(NB):=(NP);
(IB)(VP):=(IB);
(NP)(IB):=(IP);
(AB):=(AP);
(DB):=(DP);
(VB):=(VP);
(VP):=(IB);
(A):=(AB);
(D):=(DB);
(I):=(IB);
(V):=(VB);
(N):=(NB);

Building the grammar (V)

Creating the parsing module

SOURCE	RULE
(NP)("and")(NP):=(NP);	(NP,%np1)("and")(NP,%np2):=(and(%np2;%np1),+XP=NP,+LEX=N);
(AP)(VB):=(VB);	(AP,%a)(VB,%v):=(VB(%v;%a,+adjt),+XB=VB,+LEX=V);
(DP)(NB):=(NP);	(DP,%d)(NB,%n):=(NP(%n;%d,+spec),+XP=NP,+LEX=N);
(IB)(VP):=(IB);	(IB,%i)(VP,%v):=(IB(%i;%v,+comp),+XB=IB,+LEX=I);
(NP)(IB):=(IP);	(NP,%n)(IB,%i):=(IP(%i;%n,+spec),+XP=IP,+LEX=I);
(AB):=(AP);	(AB,^AP):=(+XP=AP);
(DB):=(DP);	(DB,^DP):=(+XP=DP);
(VB):=(VP);	(VB,^VP):=(+XP=VP);
(VP):=(IB);	(VP,^IB):=(+XB=IB);
(A):=(AB);	(A,^AB):=(+XB=AB);
(D):=(DB);	(D,^DB):=(+XB=DB);
(I):=(IB);	(I,^IB):=(+XB=IB);
(V):=(VB);	(V,^VB):=(+XB=VB);
(N):=(NB);	(N,^NB):=(+XB=NB);

All human beings are born free and equal in dignity and rights.





Any questions?

Exercise #9 (90 min)

- Goal: To create the parsing module for the training corpus
- Deliverable:
 - *syntax_ana_<ID>.txt*
- Activities:
 - www.unlweb.net/wiki/XV_UNL_School

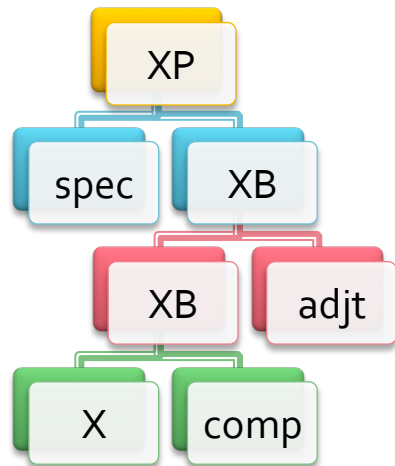


Lunch Break

Interpretation

Trees x Networks

TREE

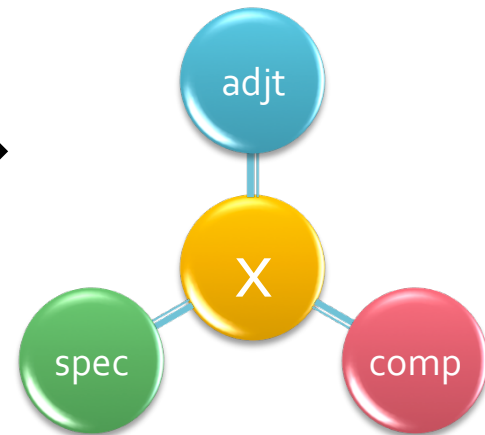


$XP(XB(XB(X;comp);adjt);spec)$

DEARBORISATION

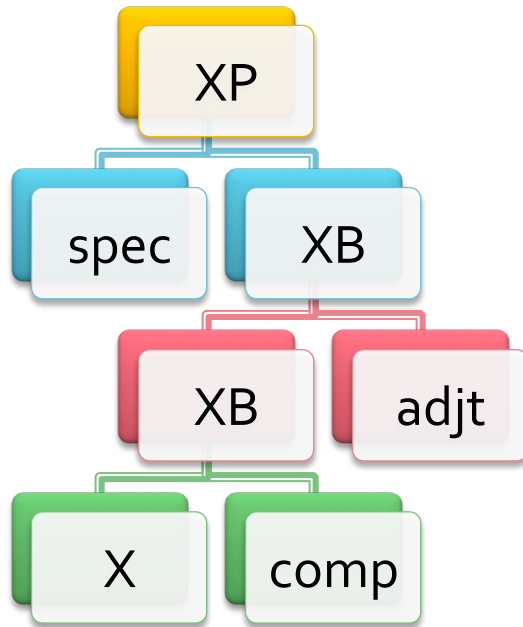
ARBORISATION

NETWORK



$XS(X;spec)$
 $XA(X;adjt)$
 $XC(X;comp)$

De-arborization



- Phrase-driven

$XP(XB(XB(X;comp);adjt);spec)$

- Head-driven

- $XC(X;comp)$

- $XA(X;adjt)$

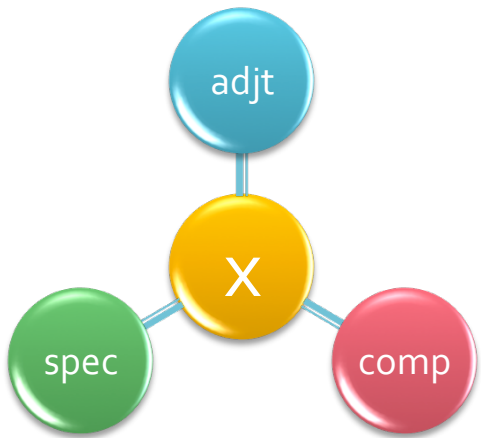
- $XS(X;spec)$

De-arborization rules

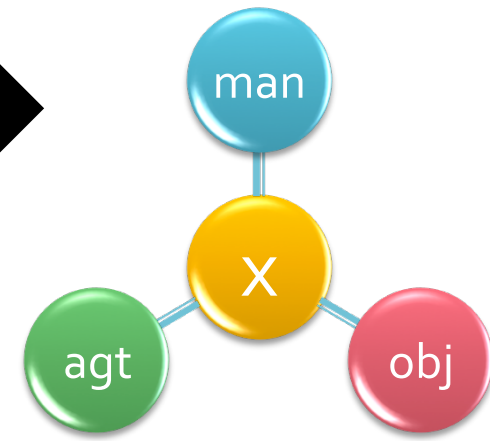
- $NP(NB(\%x;\%y);\%z) := NB(\%x;\%y)NS(\%x;\%z);$
- $NB(NB(\%x;\%y);\%z) := NB(\%x;\%y)NB(\%x;\%z);$
- $NS(NB(\%x;\%y);\%z) := NB(\%x;\%y)NS(\%x;\%z);$
- $NS(\%x;NB(\%y;\%z)) := NS(\%x;\%y)NB(\%y;\%z);$
- $NB(\%z;NB(\%x;\%y)) := NB(\%x;\%y)NB(\%z;\%x);$
- $NP(\%x;\%y) := NS(\%x;\%y);$

Interpretation

SYNTACTIC
NETWORK



SEMANTIC
NETWORK



Interpretation

- $VS(\%x;\%y) := \text{agt}(\%x;\%y);$
- $VC(\%x;\%y) := \text{obj}(\%x;\%y);$
- $VA(\%x;\%y) := \text{man}(\%x;\%y);$

Exercise #10 (90 min)

- Goal: To create the semantic module for the training corpus
- Deliverable:
 - *semantic_ana_<ID>.txt*
- Activities:
 - www.unlweb.net/wiki/XV_UNL_School



That's all, Folks!