Yerevan, October 8th

# XIII UNL School

## Day #2

# Day #2

- ~~Nlization (UGO)~~
- UNLization (CORNELIA)

# UNLization

# The UNL System


C2 cotidiano TERÇA-FEIRA, 23 DE JANEIRO DE 2007 FOLHA DE S.PAULO

RUBEM ALVES

## O aluno perfeito

Era uma vez um jovem casal que estava muito feliz. Ela estava grávida, e eles esperavam com ansiedade o filho que ia nascer.

Transcorridos os nove meses de gravidez, ele nasceu. Ela deu à luz um lindo computador! Que felicidade ter um computador como filho! Era o filho que desejavam ter! Por isso eles não se importaram muito durante toda a gravidez, chegando mesmo a fazer exames...

O batizado foi uma festança. Deram-lhe o nome de Memorioso, porque julgavam que uma memória perfeita é o essencial para uma boa educação. Educação é memorização. Crianças com memória perfeita vão bem na escola e não têm problemas para passar no vestibular.

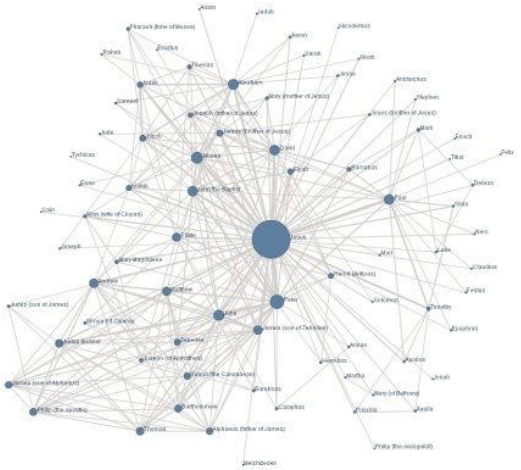
E foi isso mesmo que aconteceu. Memorioso memorizava tudo aquilo que os professores ensinavam. Mas tudo mesmo. E não reclamava. Seus computadores reclamavam, diziam que aquelas coisas que lhes eram ensinadas não faziam sentido. Suas inteligências recusavam-se a aprender. Travavam. Sofriam ruins. Ficavam de recuperação.

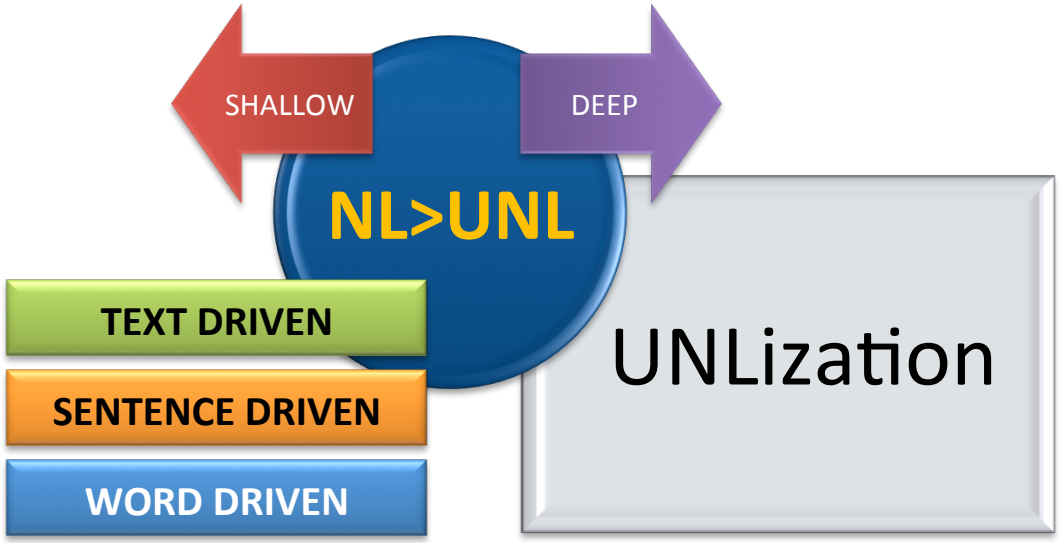
Isso não aconteceu com Memorioso. Ele memorizava com a mesma facilidade a maneira de extrair raiz quadrada, reações químicas, fórmulas de física, acidentes geográficos, população de países longínquos, datas de eventos históricos, nomes de rios, montanhas, regras de gramática, nomes, endereços, descobridores, cientistas, palavras novas, regras de gramática, livros inteiros, línguas estrangeiras. Sabia de cor todas as informações sobre o mundo cultural.

Há perguntas para as quais a memória não tem respostas. E que tais perguntas não se encontram na memória. Encontram-se no coração, onde mora a emoção...

Memorioso ficou rouco. Os dedos de sua memória funcionavam com a velocidade de luz procurando a resposta. Mas aquilo não lhe fazia sentido. Seu rosto ficou vermelho. Começou a suar. Sua temperatura subiu. Seu coração disparou. Teve muitas dores, paradas, e resolveu, um dia de muito frio, ir ao hospital do computadores, verificarem que seu disco rígido estava irremediavelmente danificado.

Ele se chamava Memorioso, pois seus pais julgavam que a memória perfeita é essencial para uma boa educação






UNDL FOUNDATION

# Structure

Information can be represented by semantic networks made of three different types of discrete semantic entities:

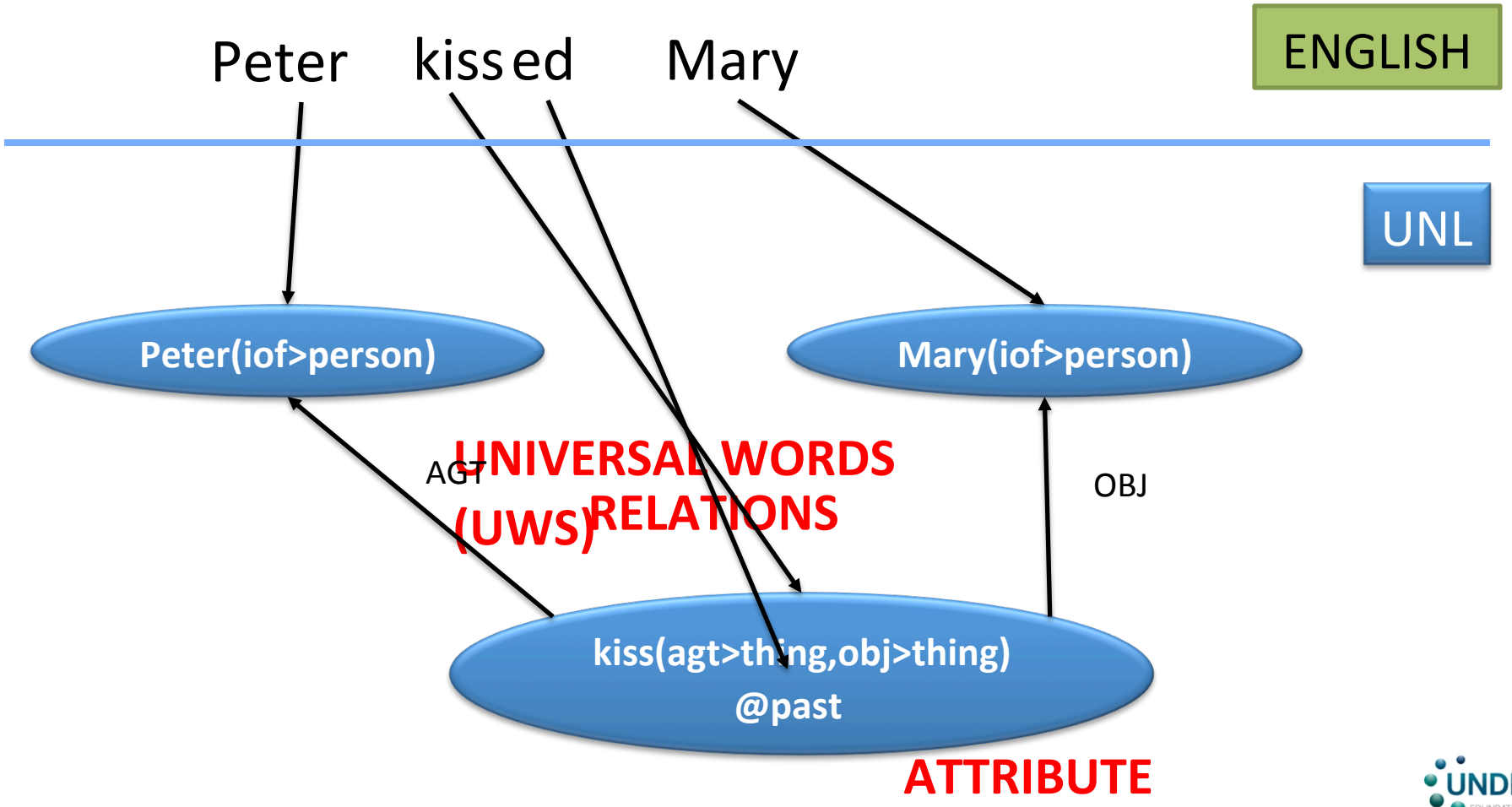
CONCEPTS = Universal Words (UWs)

CONCEPT SPECIFIERS = Universal Attributes

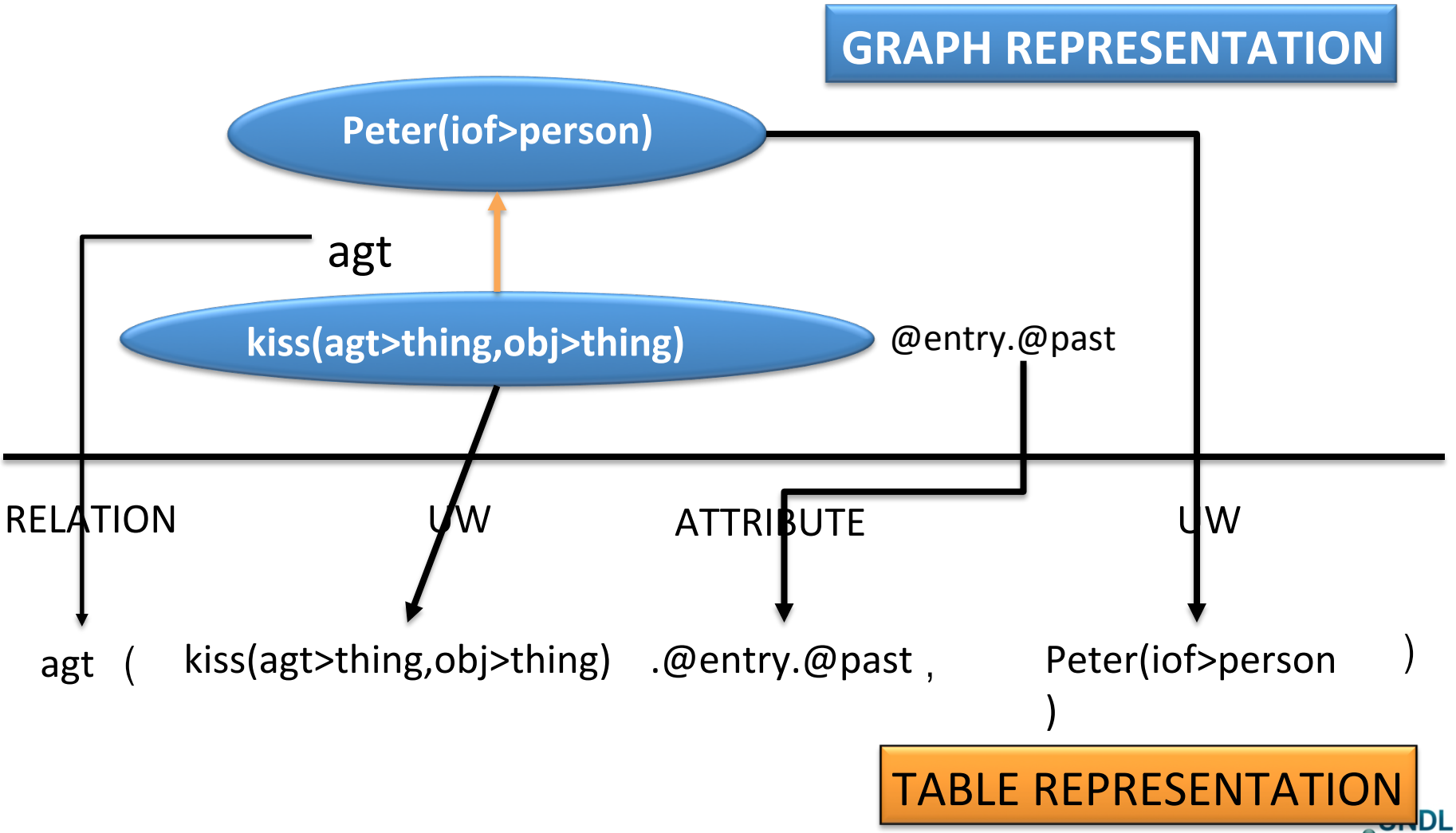
RELATIONS BETWEEN CONCEPTS = Universal Relations


UNDL FOUNDATION

# Natural Language-to-UNL (UNL-ization)



# Syntax of UNL



# UNL document

**[D]**

**[S:01]**

{org:en}

Peter kissed Mary.

{/org}

{unl}

agt(kiss(agt>thing,obj>thing).@entry.@past, Peter(iof>person))

obj(kiss(agt>thing,obj>thing).@entry.@past, Mary(iof>person))

{/unl}

**[/S]**

**[/D]**

UNDL FOUNDATION

# Important!

- Nodes (UW's) must be open-class categories:
  - Nouns (common or proper) and pronouns (oo)
  - Verbs (except auxiliaries and copula)
  - Adjectives (≠ determiners)
  - Adverbs (except specifiers, such as "very" and "not")
  - Numerals (to be represented as digits)
- Nodes (UW's) must be represented by their lemmas (citation form)

# For instance:

- One (=1) machine can do the work of fifty (50) ordinary men (= man). No machine can do the work of one(=1) extraordinary man.
  - can do = do.@ability
  - the work = work.@def
  - work of man = agt(work,man)
  - No machine = machine.@no

# Questions?

Activity #2

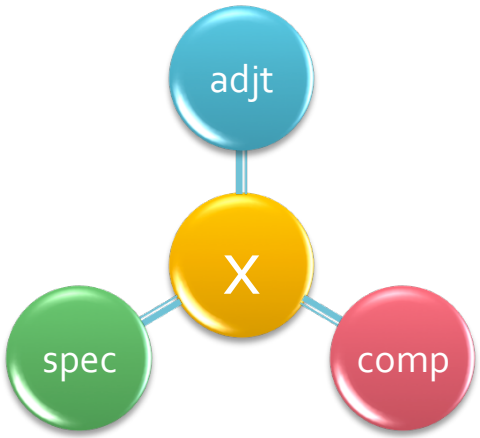
# CORNELIA

# CORNELIA

- **COR**pus for **N**atural languag**E** **I**nductive **A**nalysis
  - NL -> UNL
  - Goals
    - Training
    - Setting the standards for UNLization (grammar induction)
  - Structure
    - A1 = 250 simple NP's
    - A2 = 250 simple VP's
    - B1 = 250 complex NP's
    - B2 = 250 complex VP's
    - C1 = 250 full sentences
    - C2 = 250 full sentences
  - Requirements
    - CUP1,000
  - Languages = ALL

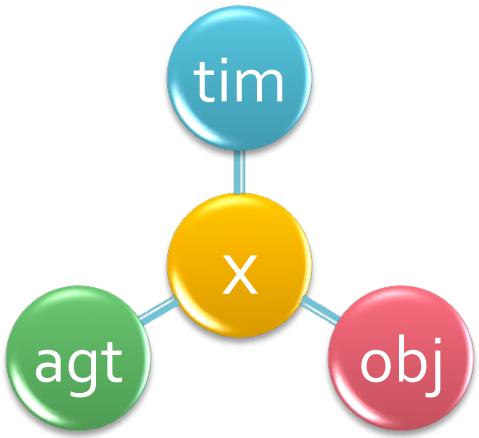
# Syntax x Semantics

**SYNTAX (SDS)**




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

**SEMANTICS (UNL)**




agt(X;spec)
tim(X;adjt)
obj(X;comp)

# UNL

# UNLization

## attributes

| NL | UNL |
|---|---|
| book | book |
| the book | book.@def |
| a book | book.@indef |
| some books<br>a few books<br>a couple of books | book.@paucal |
| many books<br>several books<br>plenty of books | book.@multal |
| too many books | book.@multal.@extra |
| actor | actor.@male |
| actress | actor.@female |

# UNLization

## relations

| NL | UNL |
|---|---|
| trip of John | agt(trip,John) |
| trip with John | ptn(trip,John) |
| trip in Yerevan | plc(trip,Yerevan) |
| trip to Yerevan | gol(trip,Yerevan) |
| trip from Yerevan | src(trip,Yerevan) |
| trip across Yerevan | via(trip,Yerevan) |
| trip in a plane | ins(trip,plane) |
| trip of my dreams | mod(trip,dream) |
| trip in a dream | lpl(trip,dream) |
| trip in the summer | tim(trip,summer) |
| trip since the summer | tmf(trip,summer) |
| trip until the summer | tmt(trip,summer) |

# Nlization

## attributes and relations

| NL | UNL |
|---|---|
| the very long trip | dur(trip.@def,long.@plus) |
| the very long trip of Ronaldo | dur(trip.@def,long.@plus)<br>agt(trip.@def,Ronaldo |
| the very long trip of Ronaldo from Geneva to Yerevan through Warsaw | dur(trip.@def,long.@plus)<br>agt(trip.@def,Ronaldo)<br>src(trip.@def,Geneva)<br>via(trip.@def,Warsaw)<br>gol(trip.@def,Yerevan) |
| the very long trip of Ronaldo from Geneva to Yerevan through Warsaw in the flight LOL761 by Polish Airlines | dur(trip.@def,long.@plus)<br>agt(trip.@def,Ronaldo)<br>src(trip.@def,Geneva)<br>via(trip.@def,Warsaw)<br>lpl(trip.@def,flight.@def)<br>cnt(flight.@def,"LOL416")<br>agt(flight.@def,Polish Airlines) |

# Important!

- You should encode ALL and ONLY the information available in the source language
  - teacher = teacher (~~teacher.@male~~)
- You should encode ONLY ONE MEANING, preferably the most conventional one
  - book of Paris
    - mod(book,Paris)
    - ~~src(book,Paris)~~
    - ~~nam(book,Paris)~~
    - ~~pos(book,Paris)~~
- Be careful of morphological and syntactic biases
  - I like books = book.@generic (and not book.@pl)
  - machine translation = agt(translation,machine)
  - automatic translation = ins(translation,automatic)
  - good translation = mod(translation,good)
  - past translation = tim(translation,past)

# Important!

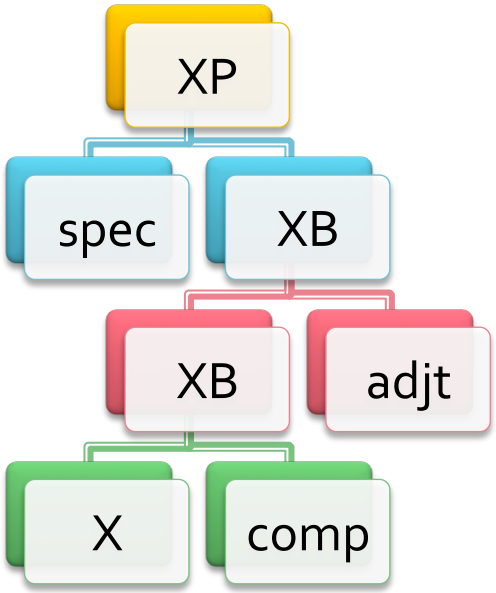
- The order of the attributes is not important
  - book.@def.@distal.@pl = book.@pl.@distal.@def
- Relations are not commutative
  - cnt(evidence,absence) = evidence of absence
  - cnt(absence,evidente) = absence of evidence
- The relation is always defined by the target
  - relation(source,target) = target is relation of source
    - pos(book,John) = John is possessor of book (John's book)
    - ben(book,John) = John is beneficiary of book (book for John)
    - cnt(book,John) = John is content of book (book about John)
    - mat(book,stone) = stone is the material of book (book of stone)
    - and(John, Mary) = Mary is coordinated to John (John and Mary)
- Be careful about multiword expressions (words must be analyzed only if compositional)
  - New York = New York (and not mod(York,New))

# SDS

## syntactic dependency structure

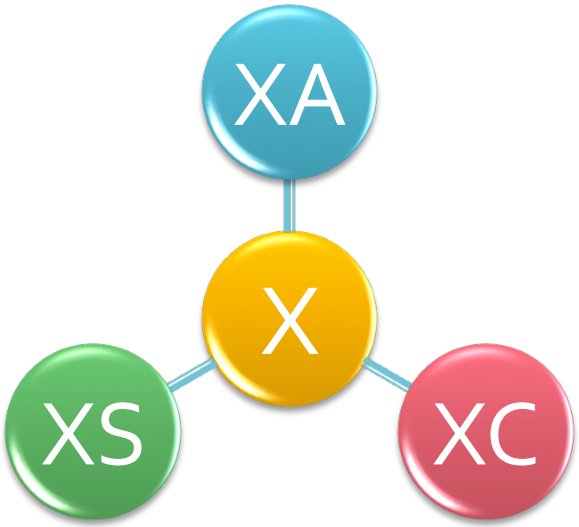
# X-bar structure

## SCS
**(SYNTACTIC CONSTITUENCY STRUCTURE)**




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

## SDS
**(SYNTACTIC DEPENDENCY STRUCTURE)**




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

# SDS
## heads

- HEAD (X)
  - N (noun) > NP (noun phrase)
    - she, girl, the girl, the beautiful girl, the beautiful girl from Yerevan, the beautiful girl from Yerevan that I saw yesterday, …
  - D (determiner) > DP (determiner phrase)
    - the, all, my, all my, all the, all these, …
  - J (adjective) > JP (adjective phrase)
    - loyal, very loyal, very loyal to the project, much more loyal to the project than anyone else , …
  - A (adverb) > AP (adverbial phrase)
    - differently, very differently, differently from the others, etc
  - P (preposition) > PP (prepositional phrase)
    - with her, with the girl, with the beautiful girl, with the beautiful girl from Yerevan, …
  - V (verb) > VP (verbal phrase)
    - Mary died, Peter killed Mary, Peter killed Mary with a knife, Peter killed Mary with a knife in the kicthen yesterday because of John, …
  - I (inflection) > IP (inflectional phrase)
    - it is raining, it has rained, it would have rained, …
  - C (conjunction) > CP (complementizer phrase)
    - if he comes, whether he comes, that he comes, …

# SDS
## specifiers

- SPECIFIERS (XS)
  - NS (specifiers in a NP) = DETERMINERS
    - (NONE) she, (NONE) girl, (the) girl, (the) beautiful girl, (all the) beautiful girl
  - DS (specifiers in a DP) = QUANTIFIERS
    - (NONE) the, (all) the, (all) my, …
  - JS (specifiers in a JP) = QUANTIFIERS (intensifiers, downtoners)
    - (NONE) loyal, (very) loyal, (much more) loyal, (not) loyal, …
  - AS (specifiers in a AP) = QUANTIFIERS (intensifiers, downtoners)
    - (NONE) differently, (very) differently, …
  - PS (specifiers in a PP) = QUANTIFIERS (intensifiers, downtoners)
    - (NONE) with her, (right) above the table, …
  - VS (specifiers in a VP) = SUBJECT
    - Mary died, Peter killed Mary, Peter killed Mary with a knife, Peter killed Mary with a knife in the kicthen yesterday because of John, …
  - IS (specifiers in an IP) = SUBJECT
    - (it) is raining, (it) has rained, (it) would have rained, …
  - CS (specifiers in a CP) = SUBJECT
    - if (he) comes, whether (he) comes, that (he) comes, …

# SDS
## complements (necessary arguments)

- COMPLEMENTS (XC)
  - NC (complements in a NP) (used with deverbals)
    - book (NONE), arrival (of John), construction (of a building), …
  - DC (complements in a DP)
    - DO NOT EXIST IN ENGLISH
  - JC (complements in a JP)
    - beautiful (NONE), loyal (to the queen), full (of people), …
  - AC (complements in a AP)
    - sadly (NONE), differently (from them), …
  - PC (complements in a PP)
    - with (her), of (John), …
  - VC (complements in a VP) = OBJECT AND PREDICATIVE
    - Mary killed (Peter), Mary laughed (at John), Mary is (beautiful), …
  - IC (complements in a IP) = VP
    - is (raining), would have (to go there), …
  - CC(complements in a CP) = VP or IP
    - if (he comes), whether (he is going or not), …

# SDS
## adjuncts (extra-nuclear)

- ADJUNCTS (XA)
  - NA (adjuncts in a NP)
    - book (NONE), (new) book, book (of John), …
  - DA (adjuncts in a DP)
    - DO NOT EXIST IN ENGLISH
  - JA (adjuncts in a JP)
    - beautiful (NONE), beatiful (for John), …
  - AA (adjuncts in a AP)
    - sadly (NONE), sadly (for John), …
  - PA (adjuncts in a PP)
    - DO NOT EXIST IN ENGLISH
  - VA (adjuncts in a VP)
    - John killed Mary (in the kitchen) (with a knife) (yesterday) (because of Peter), …
  - IA (adjuncts in a IP)
    - it is (not) raining, …
  - CA (adjuncts in a CP) = VP or IP
    - DO NOT EXIST IN ENGLISH

# Important!

- Be careful about prescriptive grammars biases:
  - John went to Paris
    - to Paris = complement (necessary argument)
  - John is beautiful
    - beautiful = complement (necessary argument)

# Exercise #2.1 (30 min)

- Create an assignment in the project CORNELIA-A1-SB (for English) for 10 entries (order by ORDER)
- Provide the corresponding NL and UNL graphs
- Discuss the problems and results

# Activities #2.2-2.5

- #2.2
  - Create assignments for the next 10 entries (11-20)
  - Discuss the results
- #2.3
  - Create assignments for the next 10 entries (21-30)
  - Discuss the results
- #2.4
  - Create assignments for the next 10 entries (31-40)
  - Discuss the results
- #2.5
  - Create assignments for the next 10 entries (41-50)
  - Discuss the results