Yerevan, October 7th

XIII UNL School

Day #1

Day #1

- Welcome
- Context
  - The UNL
  - The UNL System
  - The UNDL Foundation Road Map
- GENERATION

Welcome

Participants

- Aikaterini Tsiouma (Greek)
- Ali Safari (Azeri)
- Anahit Sargsyan (Armenian)
- Araksia Hakobyan (Armenian)
- Grega Milharcic (Slovenian)
- Lilit Paremuzian (Armenian)
- Maryam Faal Hamedanchi (Persian)
- Mihaela Ilioaia (Romanian)
- Oleg Kapanadze (Georgian)
- Olga Vartzioti (Latin)
- Ronaldo Martins (UNL)
- Sameh Alansary (Arabic)
- Somdev Kar (Bengali)
- Tatevik Manukyan (Armenian)

Observers

- Avaq Avaqyan (Armenian)
- Karen Khachatryan (Armenian)
- Levon Hakobyan (Armenian)
- Samvel Hovsepyan (Armenian)
- Setrak Grigoryan (Armenian)

The UNL

The Universal Networking Language (UNL)

UNL

translation
knowledge representation

☼1996

UNITED NATIONS
UNIVERSITY

Universal
Networking
Language

Commitments

1. The UNL must represent information

   The UNL must represent "what was meant" (and not "what was said").

2. The UNL must be a language for computers

   The UNL must be computable.

3. The UNL must be self-sufficient

   The UNL representation must not depend on any implicit knowledge.

4. The UNL must be general-purpose

   The UNL must not be bound to translation.

5. The UNL must be independent from any particular natural language

   As a language of the UN, the UNL must be neutral.

Properties

- Non-Ambiguity
  - the boys saw the girl with the telescope
  - [[the boys] [[[saw(icl>perceive) [the girl]] [with the telescope]]]]
- Non-Redundancy
  - Peter killed Mary ≅ Mary was killed by Peter ≅ Peter caused Mary to die
- Compositionality
  - John devoured thousands of books = John read many books
- Declarativeness
  - Can you pass me the salt? = (you pass the salt to me).@request.@polite
- Completeness
  - The monkey took the banana and ate it
  - The monkey_i took the banana_j and the monkey_i ate the banana_j

Structure

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

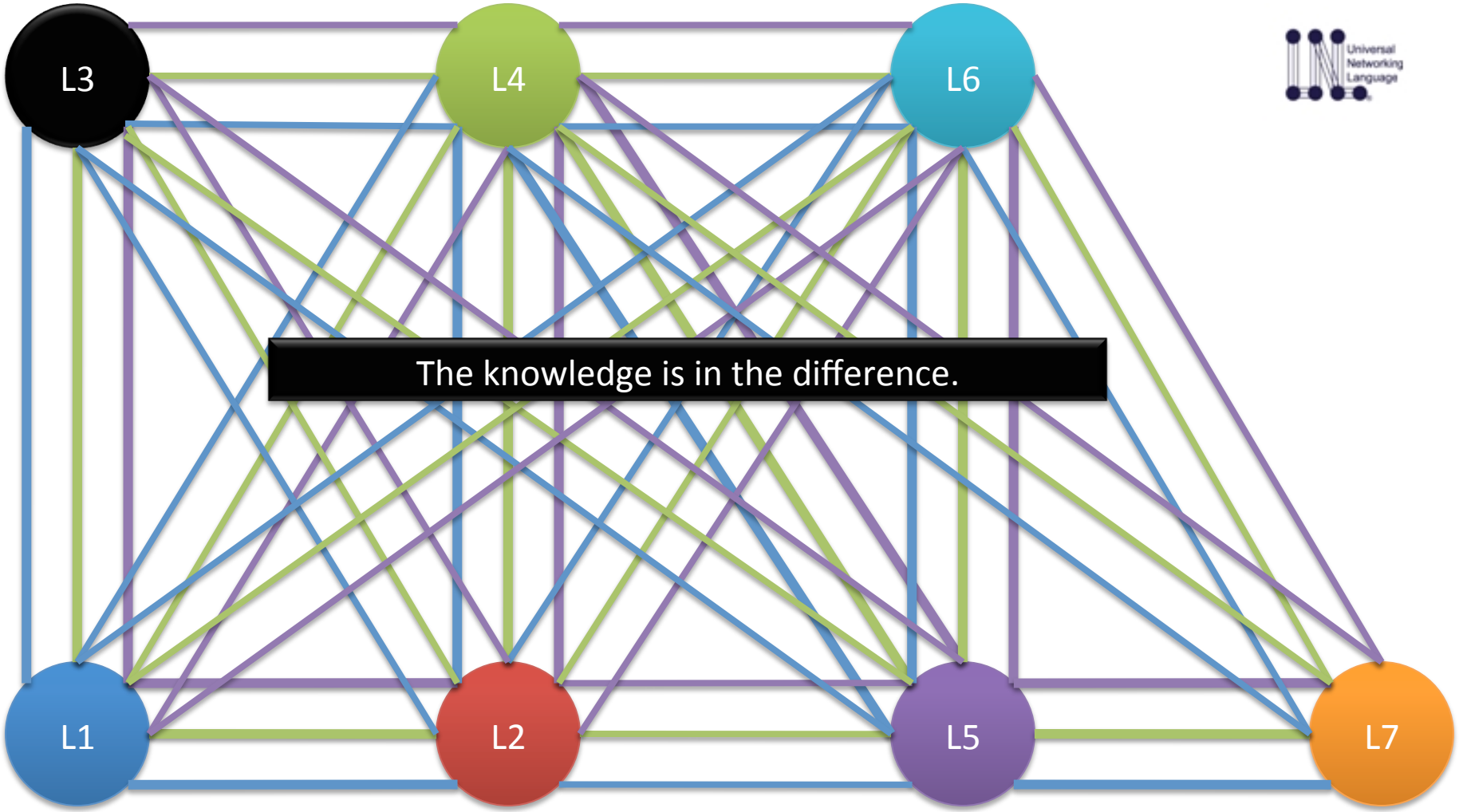
CONCEPTS = Universal Words (UWs)

CONCEPT SPECIFIERS = Universal Attributes

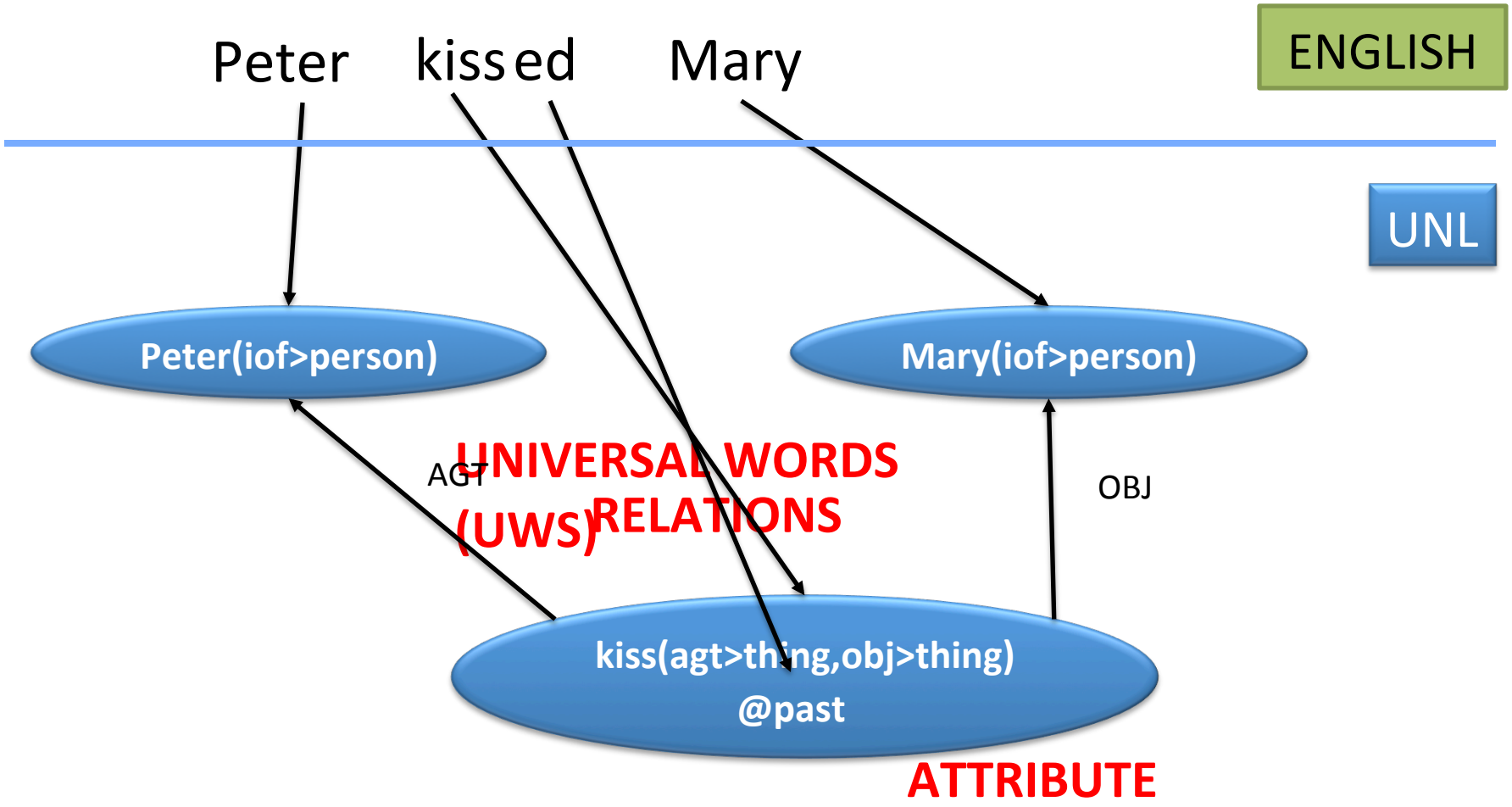
RELATIONS BETWEEN CONCEPTS = Universal Relations

UNDL FOUNDATION

The Universal NETWORKING Language



Natural Language-to-UNL (UNL-ization)



The UNL System

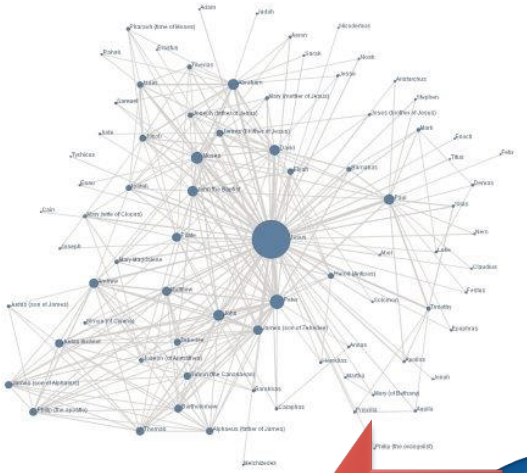
The UNL System


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

RUBEM ALVES

O aluno perfeito

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

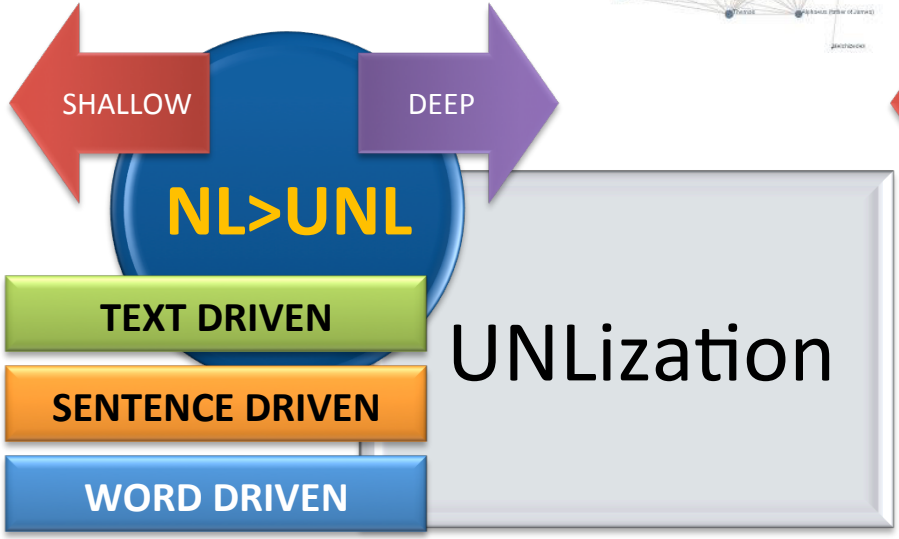


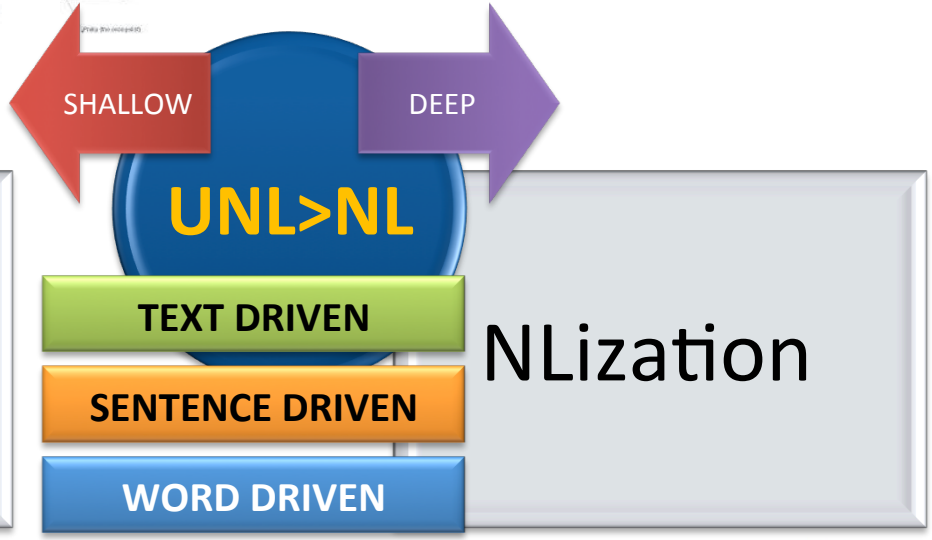

C2 cotidiano TERÇA-FEIRA, 23 DE JANEIRO DE 2007 FOLHA DE S.PAULO

RUBEM ALVES

O aluno perfeito

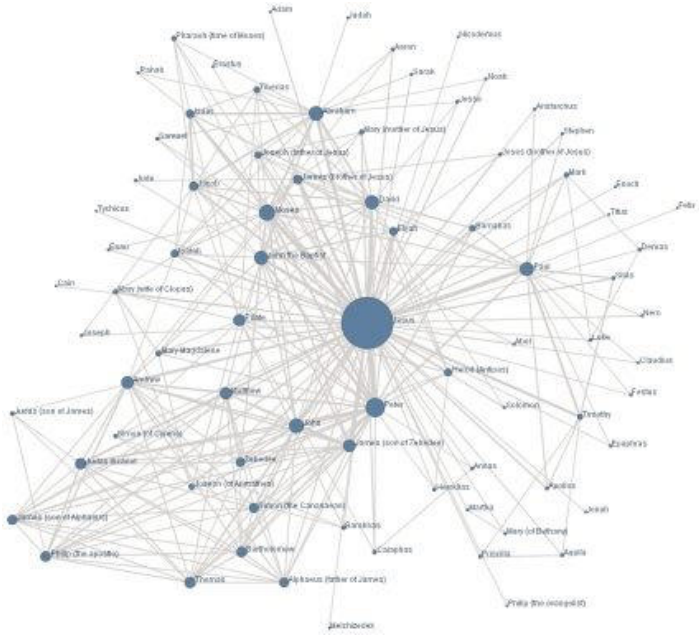
Ele se chamava Memorioso, pois seus pais julgavam que a memória perfeita é essencial para uma boa educação





UNDL FOUNDATION

Uses of UNL



- Search
- Sentiment analysis
- Information extraction
- Generation
- Normalization
- Summarization
- Simplification

The UNDL Foundation Road Map

FRONT-END APPLICATIONS

muhit

multilingual dictionary

keys

information extraction

tut

cadmos

simplification and summarizaiton

lily

translation

BACK-END APPLICATIONS

UNL dev

UNL editor

IAN

SEAN

EUGENE

NORMA

EDGES

RESOURCES

UNL arium

TRAINING & RESEARCH

VALERIE

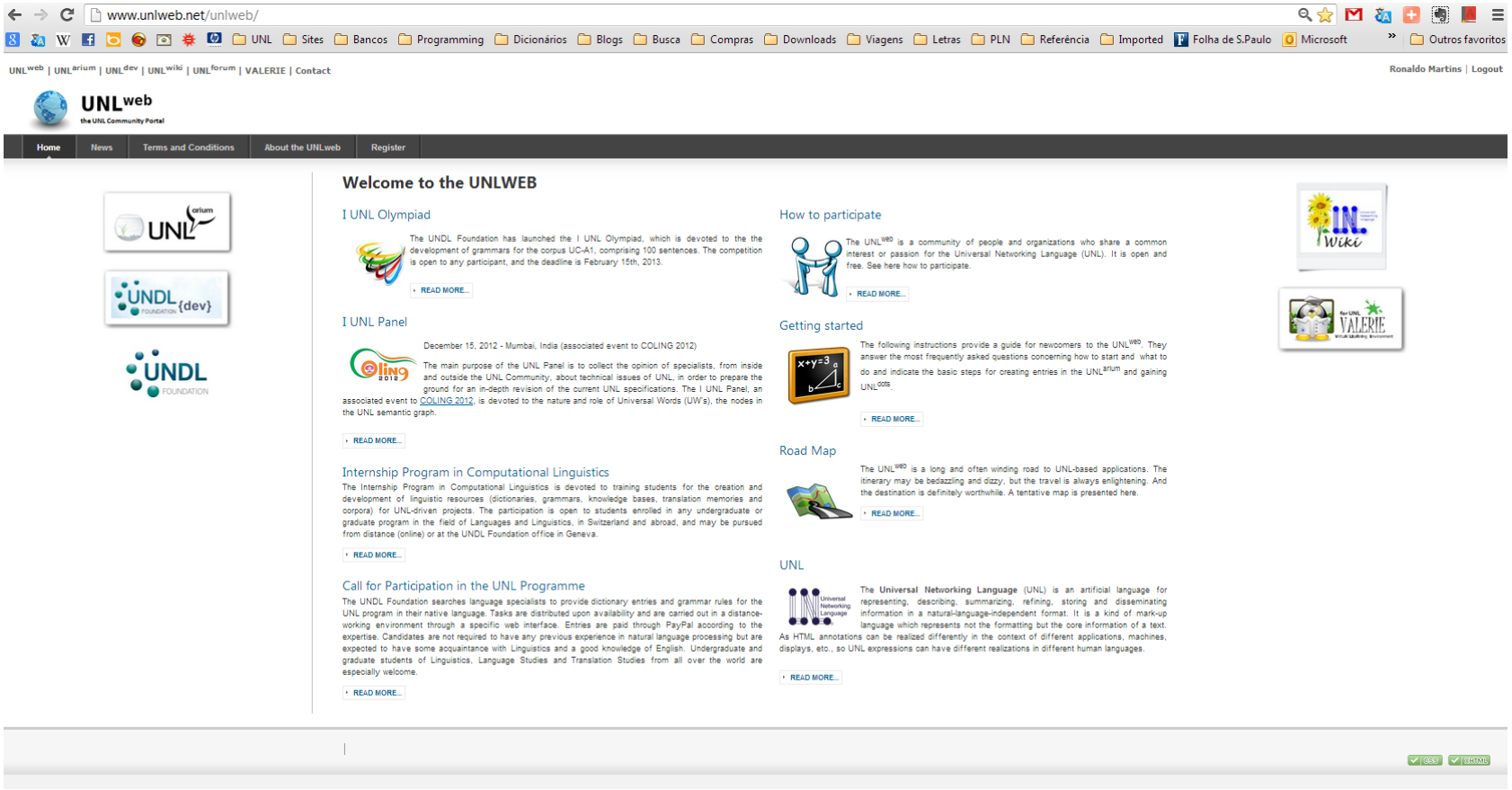
UNLversity

UNL Wiki

UNL forum

UNDL FOUNDATION

www.unlweb.net



FoR-UNL

LEVEL	DICTIONARIES			GRAMMARS	
	GD (MIR)	ND (NADIA)	AD (BRUNO)	UC (UGO)	NC (CORNELIA)
A1	**3,000**	**3,000**	**2,000**	250	250
A2	**7,000**	**7,000**	**3,000**	250	250
B1	**10,000**	**10,000**	**5,000**	250	250
B2	**10,000**	**10,000**	**5,000**	250	250
C1	**40,000**	**40,000**	**5,000**	250	250
C2	**40,000**	**40,000**	**5,000**	250	250

UNDL FOUNDATION

XIII UNL School

Goals

- To build the basic infrastructure for the following projects:
  - MIR
  - NADIA
  - BRUNO
  - UC
  - CORNELIA

Program

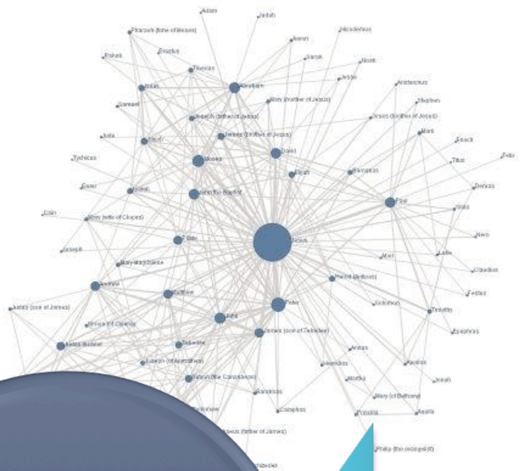
- October 7th
  - Introduction
  - Generation (UGO)
- October 8th
  - Analysis (CORNELIA)
- October 9th
  - Morphology
    - Inflectional Paradigms
    - Subcategorization Frames
- October 10th
  - Dictionary
    - MIR, NADIA and BRUNO
- October 11th
  - Evaluation & Discussion

Warnings

- Doubts are allowed: don't be afraid or shy.
- This is an ongoing initiative: we don't have all the answers yet.
- This is not a competition.
- All the material will be available at www.unlweb.net/wiki/XIII_UNL_School

NLIZATION

NLization




C2 cotidiano FOLHA DE S.PAULO

RUBEM ALVES

O aluno perfeito

SHALLOW

DEEP

UNL>NL

NLization

TEXT DRIVEN

SENTENCE DRIVEN

WORD DRIVEN

Nlization

attributes

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

Nlization

relations

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

Nlization

attributes and relations

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

Important!
attributes

- There can be more than one way of realizing an attribute
  - teacher.@multal = (many, several, plenty of) teachers
- Some attributes are not realized at all
  - teacher.@female = teacher
- Some attributes are cumulative
  - teacher.@all = all teachers (~~all teacher~~)
- Some attributes are reinterpreted by different categories
  - teacher.@trial = teachers
- Some attributes are context-sensitive
  - teacher.@multal = many teachers
  - water.@multal = much water
  - book.@most = most of the book
  - book.@most.@pl = most books

Important! (relations)

- There can be more than one way of realizing a relation
  - pos(book,John) = book of John, John's book
- Relations constitute a hierarchy, where lower relations may be replaced by upper relations + attribute
  - src(book,library) = book from the library
  - plc(book,library.@from) = book from the library
- 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)

Questions?

Activity #1

UGO

UGO

- **U**niversal **G**eneration c**O**rpus
  - UNL -> NL
  - Goals
    - Training
    - Setting the standards for NLization (Olympiads)
  - 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

Activity #1.1 (30 min)

- Register to the UNL^{sandbox} (at www.unlweb.org/sandbox)
  - Attention (for the Armenians):
    - In order to avoid superposition, do not select Armenian in the sandbox:
      - Anahit = Afar
      - Araksia = Akan
      - Lilit = Lao
      - Tatevik = Tajik
- Create an assignment for the project UGO-A1-SB with the 50 first entries (order by ORDER)
- Address the entries and verify problems, if any

Discussion #1.1

Activities #1.2-1.5

- #1.2
  - Create assignments for the next 50 entries (51-100)
  - Discuss the results
- #1.3
  - Create assignments for the next 50 entries (101-150)
  - Discuss the results
- #1.4
  - Create assignments for the next 50 entries (151-200)
  - Discuss the results
- #1.5
  - Create assignments for the next 50 entries (201-250)
  - Discuss the results