

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 Iliaoaia (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

Universal
Networking
Language



UNITED NATIONS
UNIVERSITY

☀ 1996



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 \cong Mary was killed by Peter \cong 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

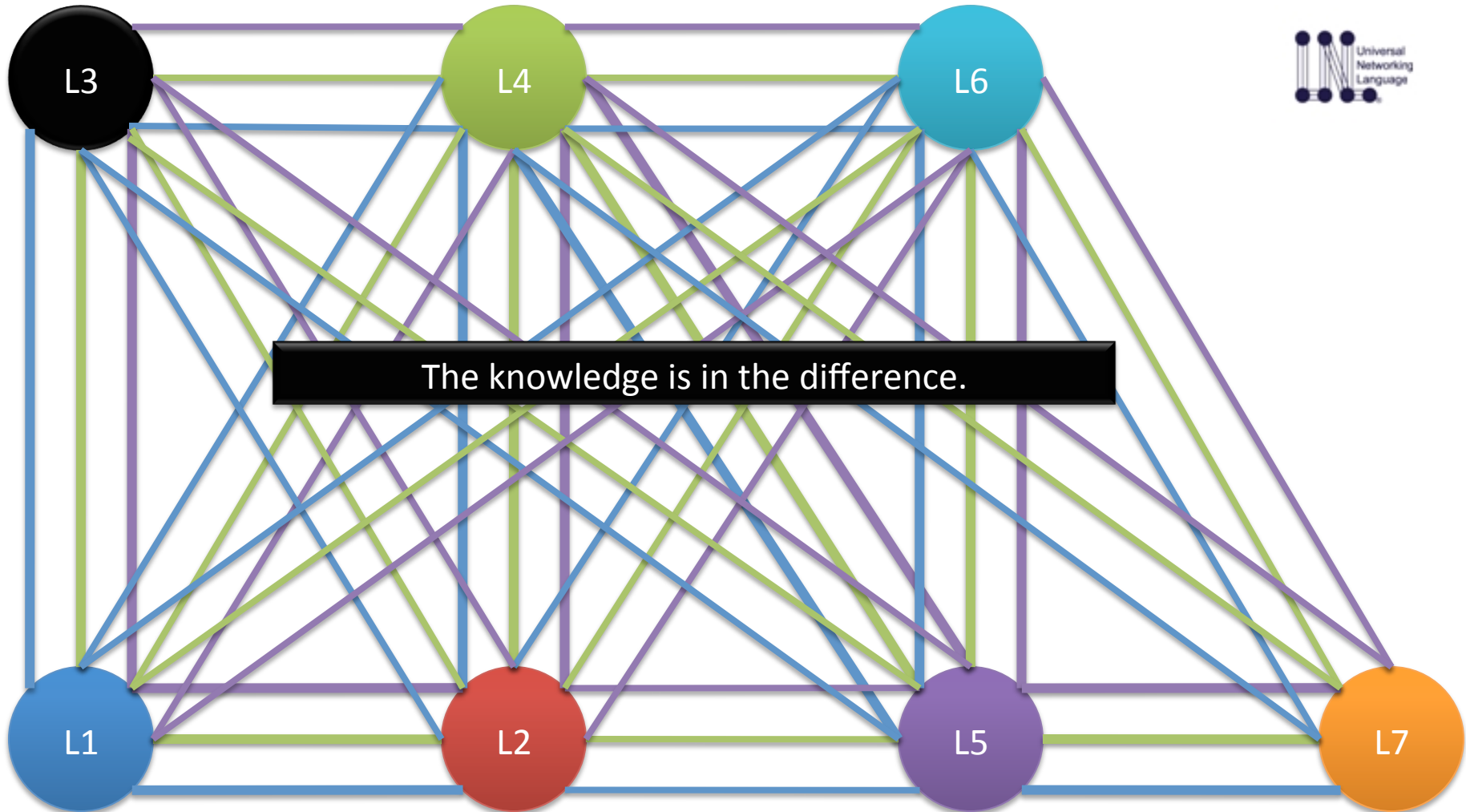
= Universal Attributes

RELATIONS BETWEEN
CONCEPTS

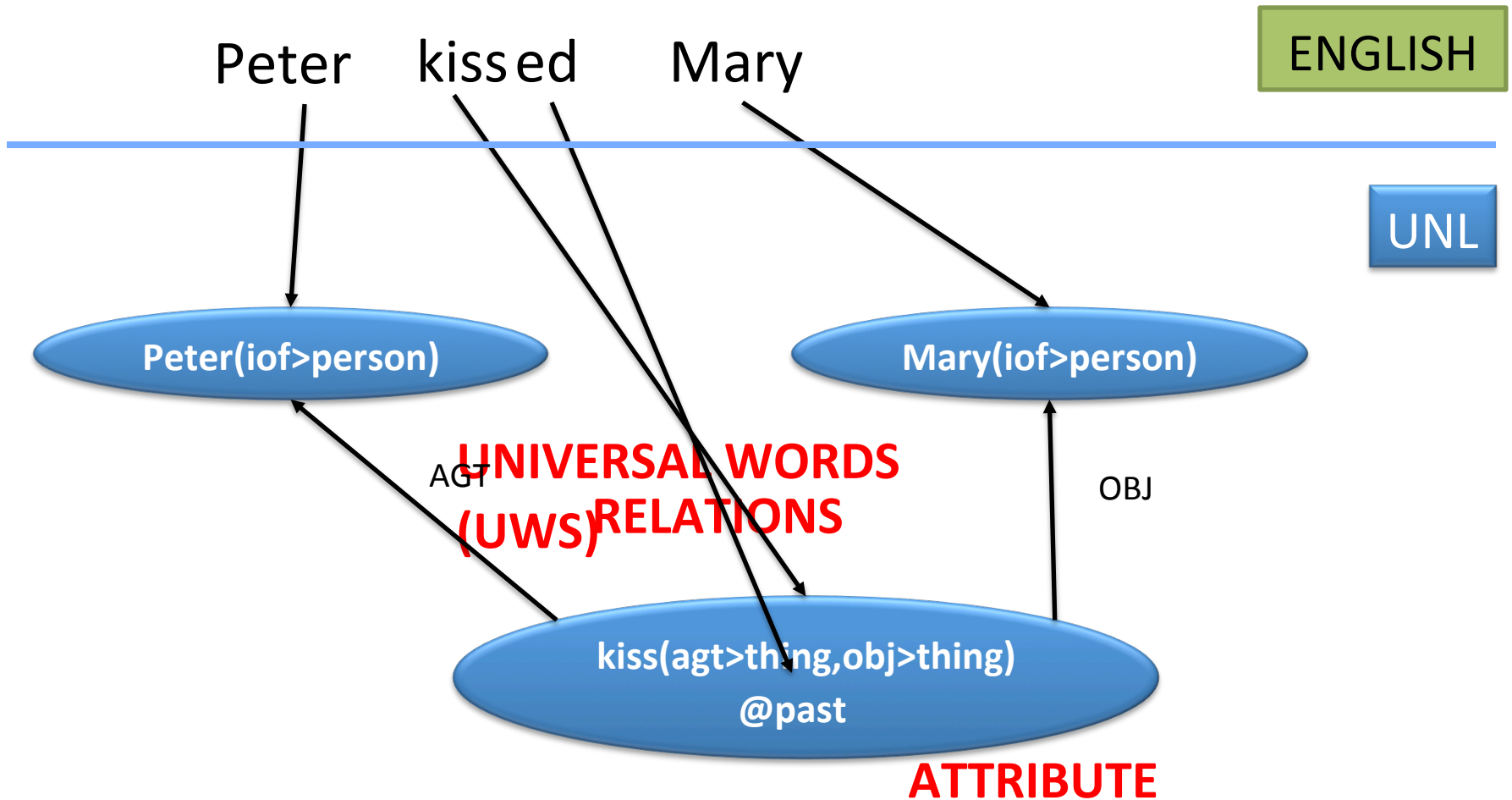
= Universal Relations



The Universal NETWORKING Language



Natural Language-to-UNL (UNL-ization)



The UNL System

The UNL System

RUBEM ALVES
Aluno perfeito

...a vida, se um jovem está...
...que estava sendo feita...
...um grande, com o momento...
...um grande sentido e que se...
...mente.

...Temporária, se quer mesmo...
...fornecer, em comum, à vida, há...
...compartilhado entre eles. Ela é...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...O trabalho é um bem, não...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

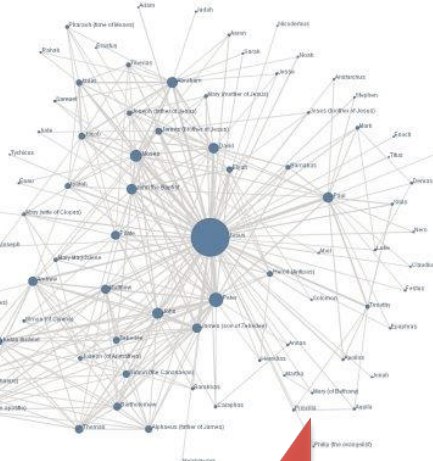
...E há um sentido que...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...poderia incluímos, além...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...Ese chamava Memotomio...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...A história de Memotomio...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...E há um sentido que...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.



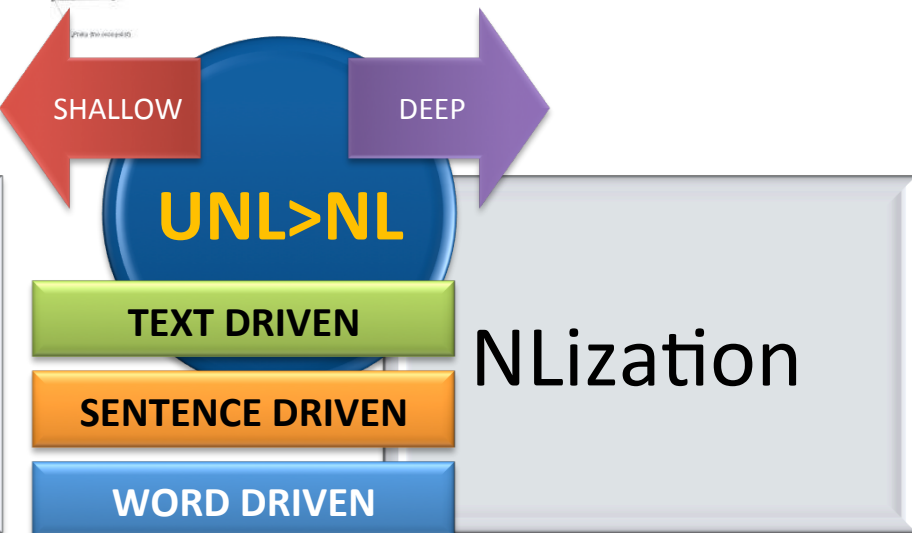
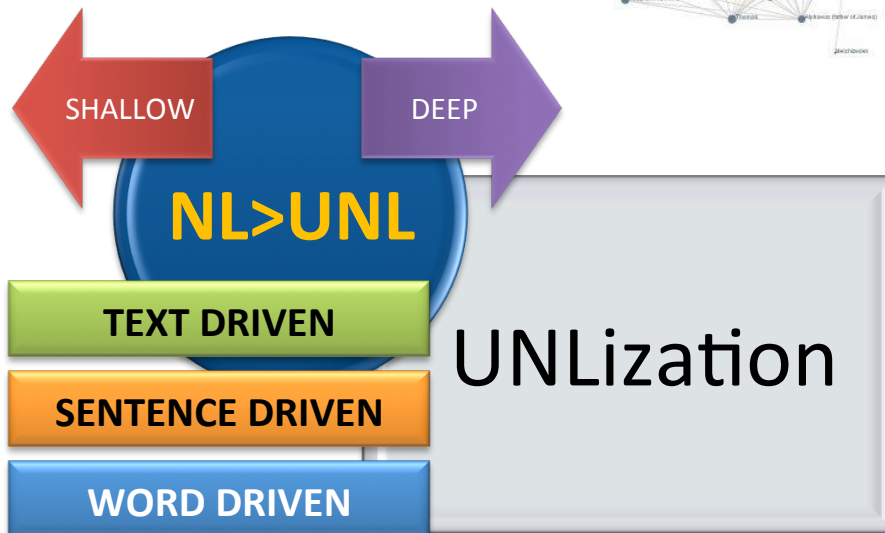
RUBEM ALVES
Aluno perfeito

...a vida, se um jovem está...
...que estava sendo feita...
...um grande, com o momento...
...um grande sentido e que se...
...mente.

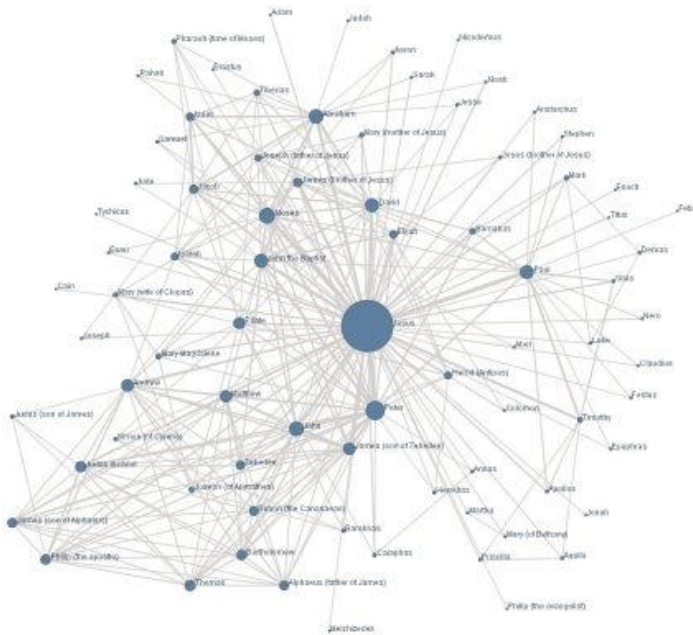
...Temporária, se quer mesmo...
...fornecer, em comum, à vida, há...
...compartilhado entre eles. Ela é...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...O trabalho é um bem, não...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.

...E há um sentido que...
...de que, enquanto vive, há...
...há um sentido, mesmo a...
...tempo.



Uses of UNL



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

The UNDL Foundation Road Map

FRONT-END APPLICATIONS

BACK-END APPLICATIONS

RESOURCES



multilingual dictionary



information extraction



simplification and summarization



translation

UNL^{dev}



UNL editor



IAN



SEAN



EUGENE



NORMA



EDGES



UNLarium

TRAINING & RESEARCH



VALERIE



UNLiversity



UNL forum

www.unlweb.net



Welcome to the UNLWEB

I UNL Olympiad



The UNDL Foundation has launched the I UNL Olympiad, which is devoted to the development of grammars for the corpus UC-A1, comprising 100 sentences. The competition is open to any participant, and the deadline is February 15th, 2013.

[READ MORE...](#)

I UNL Panel



December 15, 2012 - Mumbai, India (associated event to COLING 2012)

The main purpose of the UNL Panel is to collect the opinion of specialists, from inside and outside the UNL Community, about technical issues of UNL, in order to prepare the ground for an in-depth revision of the current UNL specifications. The I UNL Panel, an associated event to [COLING 2012](#), is devoted to the nature and role of Universal Words (UW's), the nodes in the UNL semantic graph.

[READ MORE...](#)

Internship Program in Computational Linguistics

The Internship Program in Computational Linguistics is devoted to training students for the creation and development of linguistic resources (dictionaries, grammars, knowledge bases, translation memories and corpora) for UNL-driven projects. The participation is open to students enrolled in any undergraduate or graduate program in the field of Languages and Linguistics, in Switzerland and abroad, and may be pursued from distance (online) or at the UNDL Foundation office in Geneva.

[READ MORE...](#)

Call for Participation in the UNL Programme

The UNDL Foundation searches language specialists to provide dictionary entries and grammar rules for the UNL program in their native language. Tasks are distributed upon availability and are carried out in a distance-working environment through a specific web interface. Entries are paid through PayPal according to the expertise. Candidates are not required to have any previous experience in natural language processing but are expected to have some acquaintance with Linguistics and a good knowledge of English. Undergraduate and graduate students of Linguistics, Language Studies and Translation Studies from all over the world are especially welcome.

[READ MORE...](#)

How to participate



The UNL^{web} is a community of people and organizations who share a common interest or passion for the Universal Networking Language (UNL). It is open and free. See here how to participate.

[READ MORE...](#)

Getting started



The following instructions provide a guide for newcomers to the UNL^{web}. They answer the most frequently asked questions concerning how to start and what to do and indicate the basic steps for creating entries in the UNL^{web} and gaining UNL^{web}.

[READ MORE...](#)

Road Map



The UNL^{web} is a long and often winding road to UNL-based applications. The itinerary may be bedazzling and dizzy, but the travel is always enlightening. And the destination is definitely worthwhile. A tentative map is presented here.

[READ MORE...](#)

UNL



The Universal Networking Language (UNL) is an artificial language for representing, describing, summarizing, refining, storing and disseminating information in a natural-language-independent format. It is a kind of mark-up language which represents not the formatting but the core information of a text. As HTML annotations can be realized differently in the context of different applications, machines, displays, etc., so UNL expressions can have different realizations in different human languages.

[READ MORE...](#)



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

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



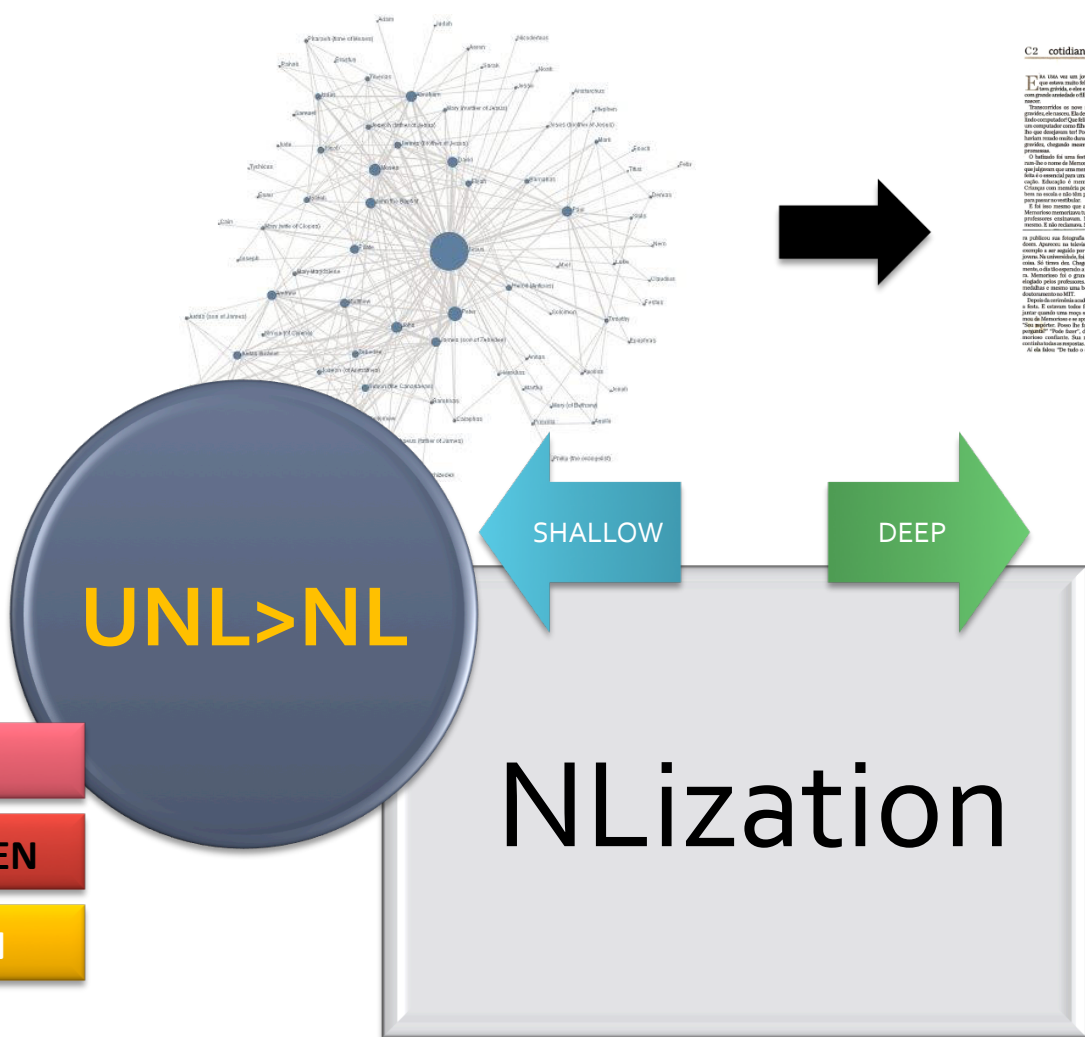
Warnings

- 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 VERSÃO DIÁRIA DO DIÁRIO DE SÃO PAULO **FORUMS SUD**

ROBEM ALVES

O aluno perfeito

Eu não sou um jovem idealista, sou apenas um jovem. Eu sou um jovem comum, com um grande sonho e uma grande vontade de fazer algo de bom.

Trabalho em uma empresa de tecnologia e sou apaixonado por isso. Quero aprender mais e crescer profissionalmente. Estou sempre buscando novas oportunidades e desafios. Não quero ficar parado, quero estar sempre evoluindo.

Quero ser um profissional de sucesso, alguém que seja respeitado e valorizado. Quero ter um bom salário e uma boa qualidade de vida. Quero ser feliz e realizar meus sonhos.

Além disso, sou muito curioso e gosto de aprender coisas novas. Quero ler muito e viajar. Quero conhecer o mundo e as diferentes culturas. Quero ser uma pessoa bem-humorada e agradável.

Quero ser um profissional que faça a diferença. Quero ajudar as pessoas e contribuir para o bem da sociedade. Quero ser um exemplo para os outros e inspirá-los a fazerem coisas boas.

Além disso, sou muito organizado e gosto de trabalhar em equipe. Quero ser um profissional que seja eficiente e produtivo. Quero sempre cumprir prazos e entregar resultados de qualidade.

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
<code>dur(trip.@def,long.@plus)</code>	the very long trip
<code>dur(trip.@def,long.@plus)</code> <code>agt(trip.@def,Ronaldo)</code>	the very long trip of Ronaldo
<code>dur(trip.@def,long.@plus)</code> <code>agt(trip.@def,Ronaldo)</code> <code>src(trip.@def,Geneva)</code> <code>via(trip.@def,Warsaw)</code> <code>gol(trip.@def,Yerevan)</code>	the very long trip of Ronaldo from Geneva to Yerevan through Warsaw
<code>dur(trip.@def,long.@plus)</code> <code>agt(trip.@def,Ronaldo)</code> <code>src(trip.@def,Geneva)</code> <code>via(trip.@def,Warsaw)</code> <code>lpl(trip.@def,flight.@def)</code> <code>cnt(flight.@def,"LOL416")</code> <code>agt(flight.@def,Polish Airlines)</code>	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
 - $\text{pos}(\text{book}, \text{John}) = \text{book of John, John's book}$
- Relations constitute a hierarchy, where lower relations may be replaced by upper relations + attribute
 - $\text{src}(\text{book}, \text{library}) = \text{book from the library}$
 - $\text{plc}(\text{book}, \text{library. @from}) = \text{book from the library}$
- Relations are not commutative
 - $\text{cnt}(\text{evidence}, \text{absence}) = \text{evidence of absence}$
 - $\text{cnt}(\text{absence}, \text{evidente}) = \text{absence of evidence}$
- The relation is always defined by the target
 - $\text{relation}(\text{source}, \text{target}) = \text{target is relation of source}$
 - $\text{pos}(\text{book}, \text{John}) = \text{John is possessor of book (John's book)}$
 - $\text{ben}(\text{book}, \text{John}) = \text{John is beneficiary of book (book for John)}$
 - $\text{cnt}(\text{book}, \text{John}) = \text{John is content of book (book about John)}$
 - $\text{mat}(\text{book}, \text{stone}) = \text{stone is the material of book (book of stone)}$
 - $\text{and}(\text{John}, \text{Mary}) = \text{Mary is coordinated to John (John and Mary)}$

Questions?

Activity #1

UGO

UGO

- **Universal Generation cOrpus**
 - 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
 - CUP_{1,000}
 - 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