

Geneva, July 1st

XII UNL School

Day #1



Day #1

- Welcome
- Context
 - The UNL
 - The UNL System
 - The UNDL Foundation Road Map
- Normalization Grammar

Welcome

Participants

- Kim Sokphyrum
- Marwa Saber
- Muhammad Zulhelmy Bin Mohd Rosman
- Ofelia Hovhannisyan
- Parameswarappa S
- Parteek Kumar
- Ronaldo Martins
- Sameh Alansary
- Serhii Prots
- Suos Samak
- Teng Wei Min
- Yordanka Stancheva



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

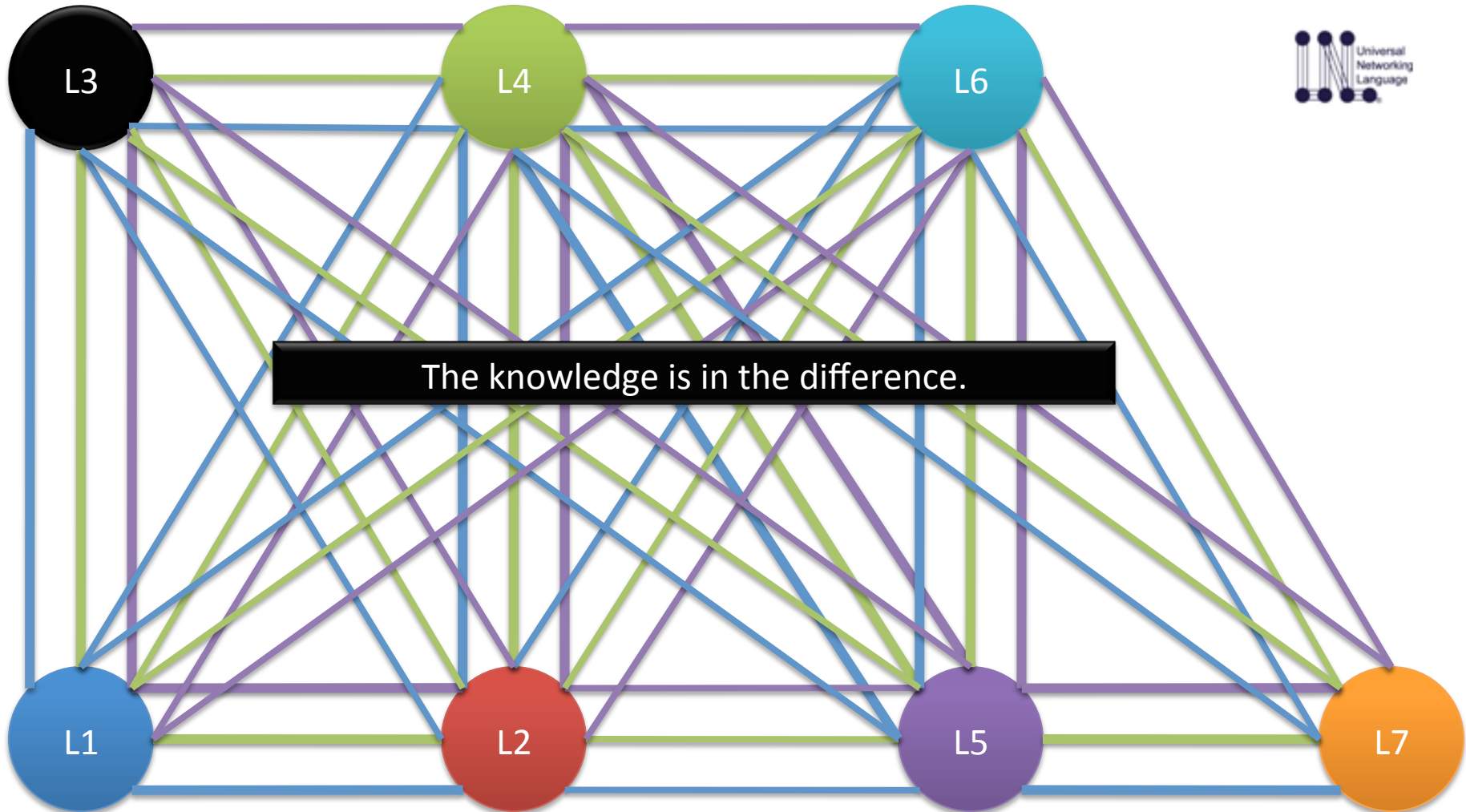
= Universal Attributes

RELATIONS BETWEEN
CONCEPTS

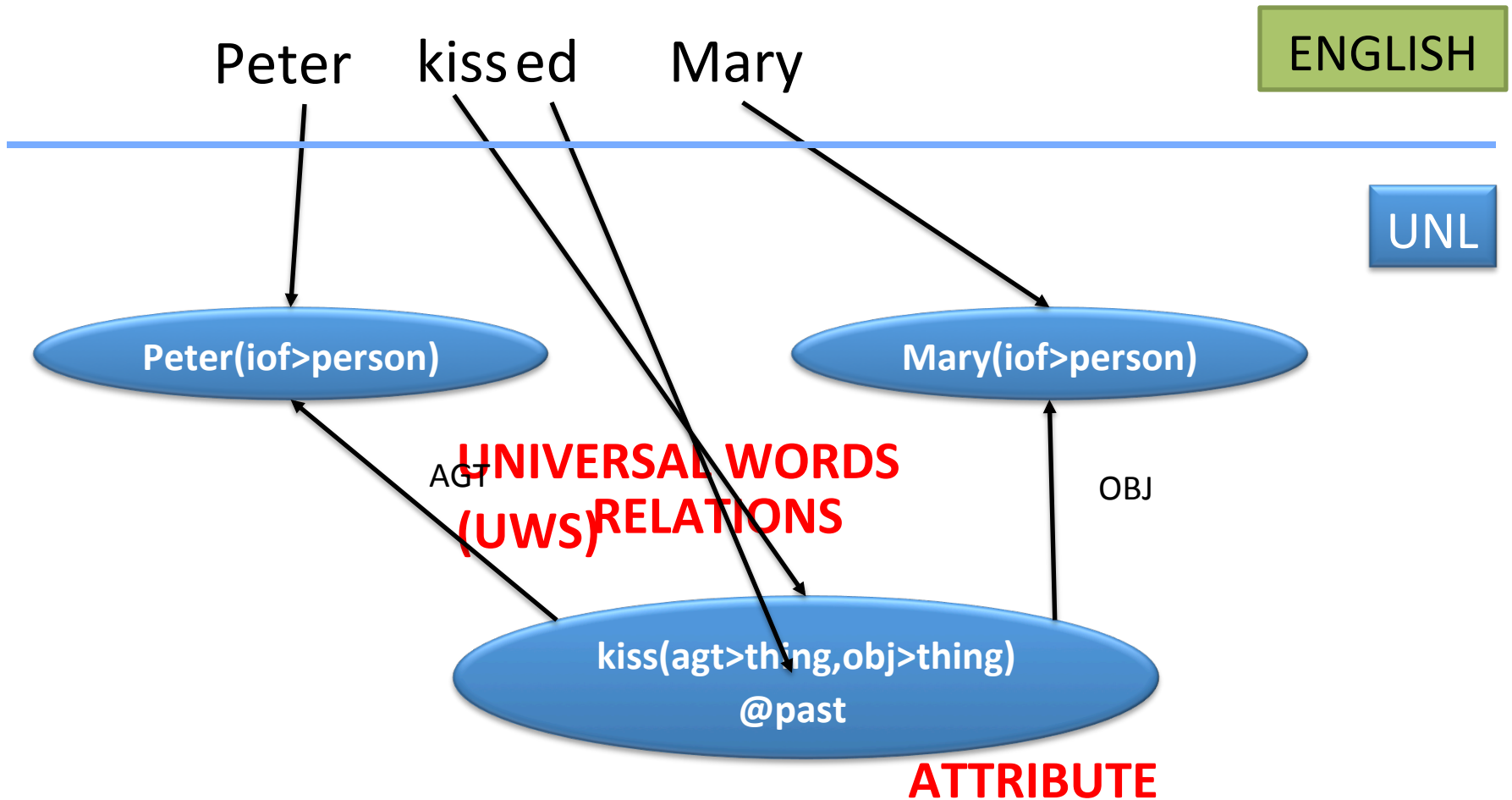
= Universal Relations



The Universal NETWORKING Language



Natural Language-to-UNL (UNL-ization)



The UNL System

The UNL System

RUBEM ALVES
Aluno perfeito

Três vezes em um mesmo ano ele ganhou o prêmio de melhor aluno de um colégio de elite. Ele conseguiu isso porque tem um método muito peculiar de estudar. Ele chama de "UNL".

Ele lê muito devagar, mas lê tudo. Não deixa nada passar. E, quando encontra algo que não entende, para ali mesmo e procura entender. Ele não se dá conta de que está estudando muito mais do que os outros. Mas ele não se importa. Ele quer aprender de verdade. E para isso, ele precisa entender tudo.

Ele não se importa com a velocidade. Ele quer a compreensão. Ele quer saber o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.

Ele não se importa com a quantidade. Ele quer a qualidade. Ele quer entender o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.

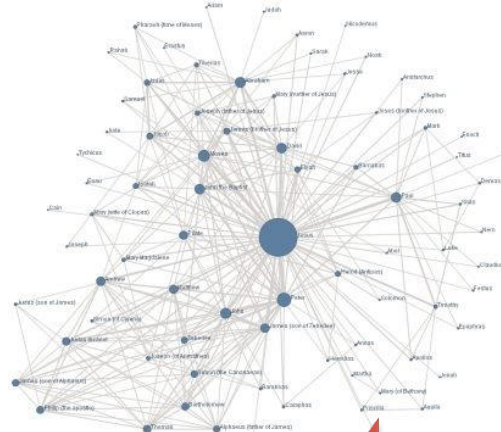
RUBEM ALVES
Aluno perfeito

Três vezes em um mesmo ano ele ganhou o prêmio de melhor aluno de um colégio de elite. Ele conseguiu isso porque tem um método muito peculiar de estudar. Ele chama de "UNL".

Ele lê muito devagar, mas lê tudo. Não deixa nada passar. E, quando encontra algo que não entende, para ali mesmo e procura entender. Ele não se dá conta de que está estudando muito mais do que os outros. Mas ele não se importa. Ele quer aprender de verdade. E para isso, ele precisa entender tudo.

Ele não se importa com a velocidade. Ele quer a compreensão. Ele quer saber o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.

Ele não se importa com a quantidade. Ele quer a qualidade. Ele quer entender o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.



RUBEM ALVES
Aluno perfeito

Três vezes em um mesmo ano ele ganhou o prêmio de melhor aluno de um colégio de elite. Ele conseguiu isso porque tem um método muito peculiar de estudar. Ele chama de "UNL".

Ele lê muito devagar, mas lê tudo. Não deixa nada passar. E, quando encontra algo que não entende, para ali mesmo e procura entender. Ele não se dá conta de que está estudando muito mais do que os outros. Mas ele não se importa. Ele quer aprender de verdade. E para isso, ele precisa entender tudo.

Ele não se importa com a velocidade. Ele quer a compreensão. Ele quer saber o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.

Ele não se importa com a quantidade. Ele quer a qualidade. Ele quer entender o que cada palavra significa, o que cada frase quer dizer. Ele quer entender o contexto, a intenção do autor. Ele quer entender a vida que há por trás das palavras.



NL>UNL

UNL>NL

UNLization

NLization

TEXT DRIVEN

TEXT DRIVEN

SENTENCE DRIVEN

SENTENCE DRIVEN

WORD DRIVEN

WORD DRIVEN

The UNDL Foundation Road Map

FRONT-END APPLICATIONS

BACK-END APPLICATIONS

RESOURCES



keys

information retrieval
and extraction



tut

summarization and
simplification



lily

translation

UNL^{dev}



UNL^{editor}



IAN



SEAN



EUGENE



NORMA



EDGES



UNL^{arium}

TRAINING &
RESEARCH



VALERIE



UNL^{iversity}



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 (UWs), 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	UNL-NL DIC	NL-UNL DIC	UNL-NL Grammar	NL-UNL Grammar
A1	MIR-A1	BRUNO-A1	UC-A1	NC-A1
A2	MIR-A2	BRUNO-A2	UC-A2	NC-A2
B1	MIR-B1	BRUNO-B1	UC-B1	NC-B1
B2	MIR-B2	BRUNO-B2	UC-B2	NC-B2
C1	MIR-C1	BRUNO-C1	UC-C1	NC-C1
C2	MIR-C2	BRUNO-C2	UC-C2	NC-C2

XII UNL School

Goals

- To build the basic modules of a NL-UNL (analysis) grammar for the corpus NCA₁
- To build the basic modules of a UNL-UNL (generation) grammar for the corpus NCA₁



Program

- July 1st
 - Introduction
 - Normalization Grammar
- July 2nd
 - Closed-Class Dictionary
- July 3rd
 - Corpus
- July 4th
 - Grammar
- July 5th
 - 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/XII_UNL_School



Normalization Grammar

Normalization (I)

Why is this necessary?

ORIGINAL

- Dr. Peter H. Smith isn't coming on July 1st. He'll be in another meeting in N.Y. I'll check with him another date asap. Would u be available next week, say, around 2 PM?

NORMALIZED

- Doctor Peter H Smith is not coming on 01/07.
- He will be in other meeting in New York.
- I will check with him other date as soon as possible.
- You would be available in next week around 14:00:00?

Normalization (II)

What does it mean?

- Replacing contractions
 - don't > do not, he'll > he will (eng)
 - du > de le, aux > à les (fra)
- Replacing abbreviations
 - Dr. > doctor, N.Y. > New York, asap > as soon as possible
- Replacing variants and non-standard language
 - u > you, an > a
- Reordering
 - Would you > you would
- Filling gaps and ellipses
 - next week > in the next week
- Removing extra content
 - , say, > ∅
- Segmenting
 - He is not coming. He will be elsewhere > He is not coming.//He will be elsewhere.

Normalization (III)

How is this done?

- N-rules
 - $(\%a)(\%b)\dots(\%n):=(\%a)(\%b)\dots(\%n);$
 - Where:
 - left side (condition): % is a string or a regular expression
 - right side (action): % is coindexed to the left side
 - Examples:
 - ("don't"):=("do not");
 - ("dr."):=("doctor");
 - ("an "):=("a ");

Normalization (IV)

Segmentation

- Segmentation is done by assigning the features:
 - SHEAD (to the beginning of the new sentence) or
 - STAIL (to the end of the sentence)
 - There is no need to assign SHEAD and STAIL simultaneously
 - SHEAD and STAIL are automatically assigned to new line or line breaks
- Examples:
 - ("?",%a):=(%a)(%b,+STAIL);
 - (".",%a)(" ",%b)(/[A-Z]/,%c):=(%a)(%d,+SHEAD)(%c);

Exercise #1

- Create a normalization grammar to go from ORIGINAL to NORMALIZED below:
- Upload the ORIGINAL and the N-GRAMMAR to IAN at devtest.undlfoundation.org and check the results.

ORIGINAL

- Dr. Peter H. Smith isn't coming on July 1st. He'll be in another meeting in N.Y. I'll check with him another date asap. Would u be available next week, say, around 2 PM?

NORMALIZED

- Doctor Peter H Smith is not coming on 01/07.
- He will be in other meeting in New York.
- I will check with him other date as soon as possible.
- You would be available in next week around 14:00:00?

Exercise #2

- Create a normalization grammar as comprehensive as possible for your native language. Include:
 - Contractions
 - Abbreviations
 - Time expressions
 - Segmentation
- Upload your normalization grammar to the UNLarium (UNLARIUM>GRAMMAR>IMPORT)