

Macau, March 11th

XI UNL School

Day #1



Day #1

- Welcome
- Context
 - The UNL
 - The UNL System
 - The UNDL Foundation Road Map
 - The UNL^{versity}
- Corpus
- Dictionaries
- X-bar

Welcome

Participants

- Ana Luísa Varani Leal
- Angela Manukian
- Chittaphone Chanyililath
- Chunlei Yang
- Dang Quang Vinh
- Jeanette Tan Yi Wen
- Kim Sokphyrum
- Lilit Paremuzyan
- Muhammad Zulhelmy Bin Mohd Rosman
- Ronaldo Martins
- Somdev Kar
- Suos Samak
- Teng Wei Min
- Violeta Liu
- Viraj Ramyanath Karunananda
- Viraj Welgama



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

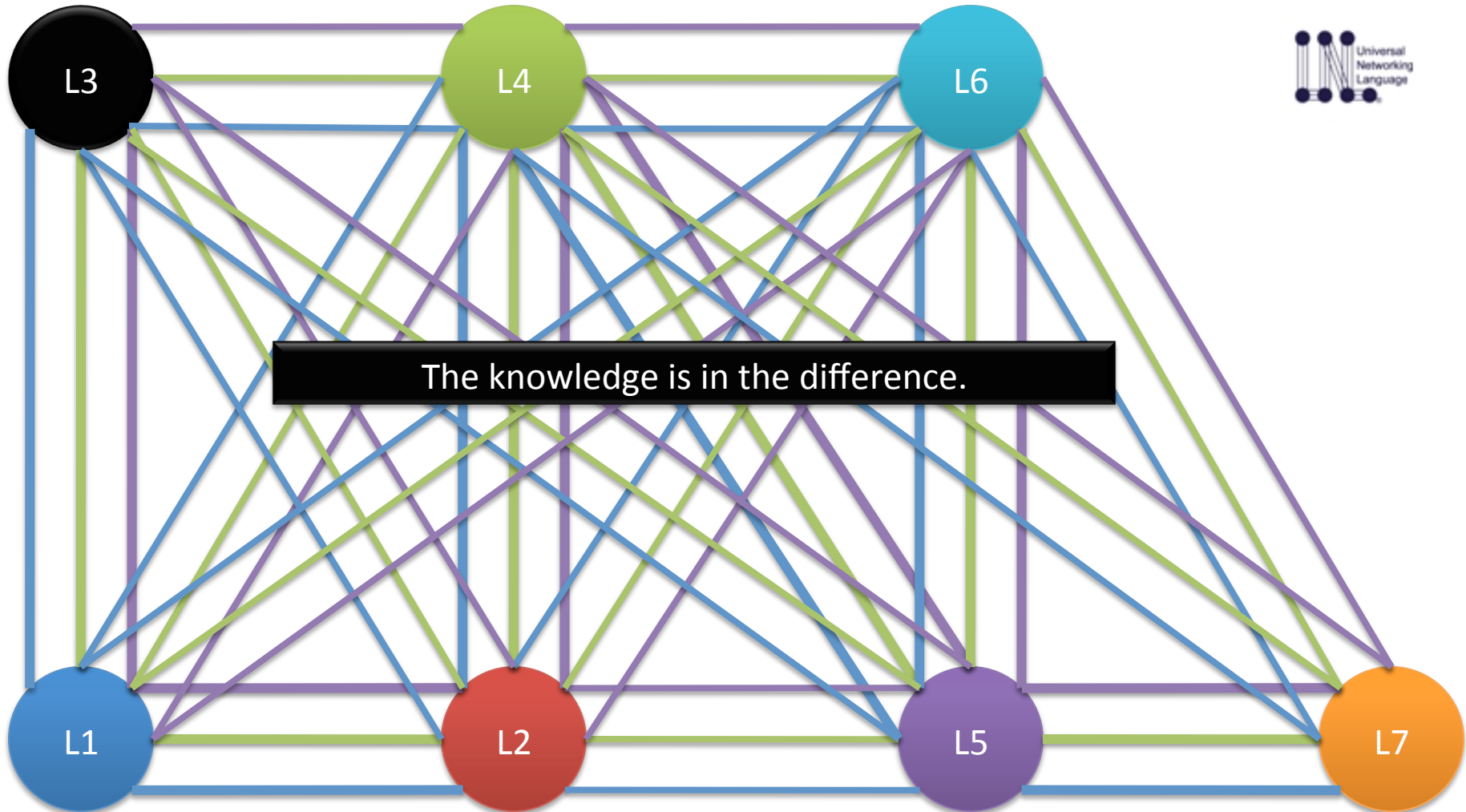
= Universal Attributes

RELATIONS BETWEEN
CONCEPTS

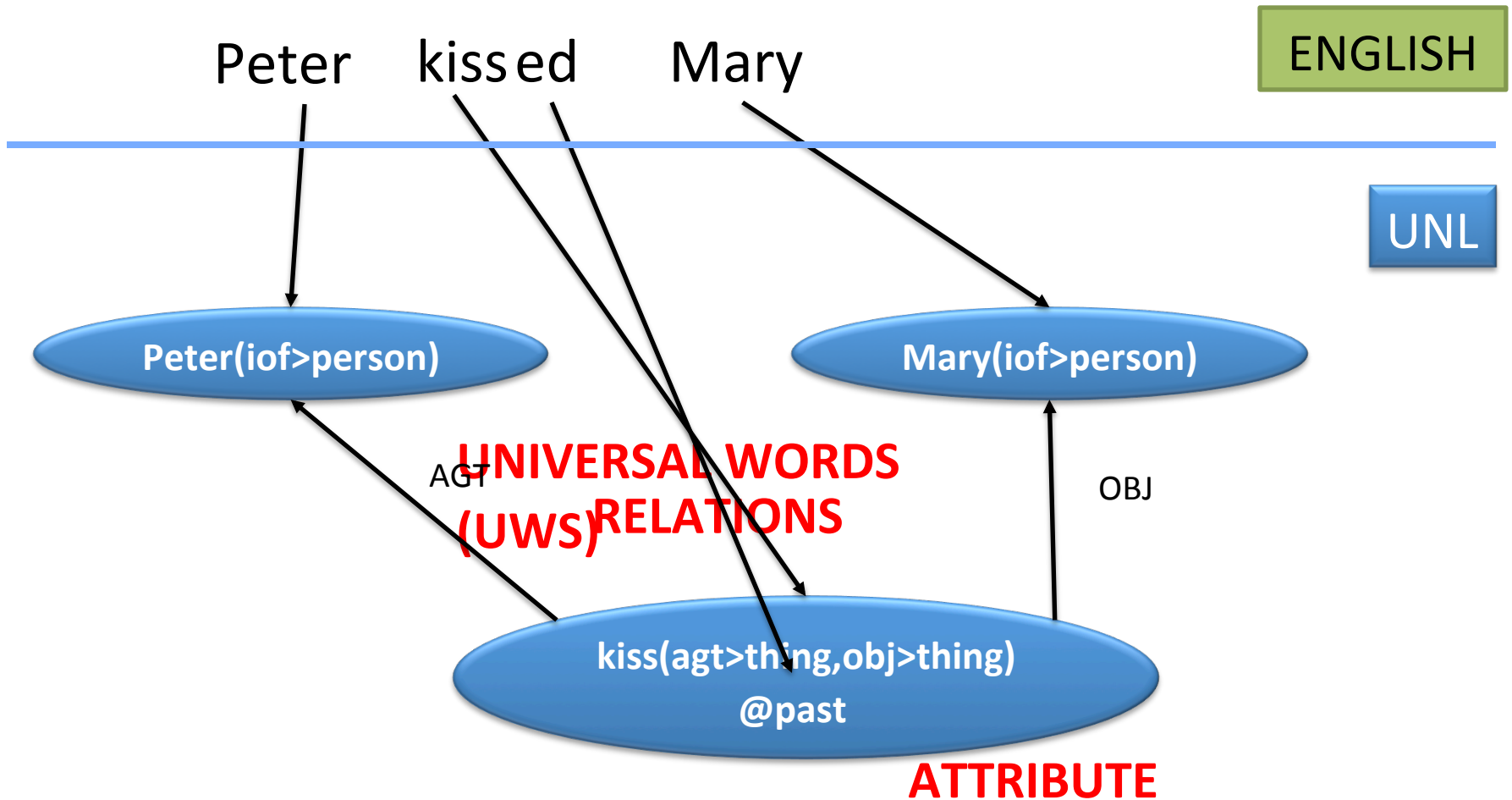
= Universal Relations



The Universal NETWORKING Language



Natural Language-to-UNL (UNL-ization)



Syntax of UNL

GRAPH REPRESENTATION

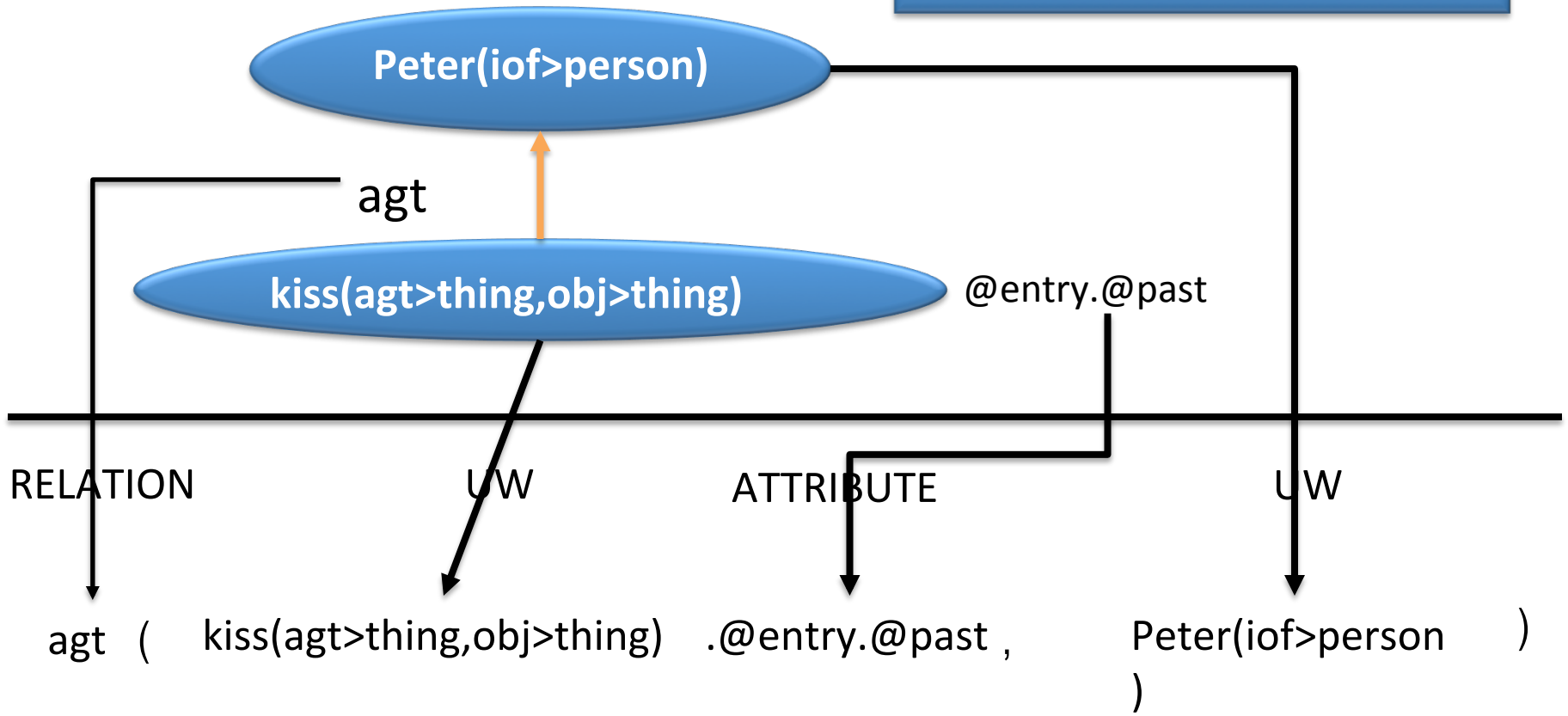


TABLE REPRESENTATION

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]

UNL (<http://anydomain/anydocument.unl>)

<unl>

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

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

</unl>



Peter kissed Mary.



Pierre a embrassé Marie.



Pedro besó a María.

The UNL System

The UNL System

ROBEM ALVES
Aluno perfeito

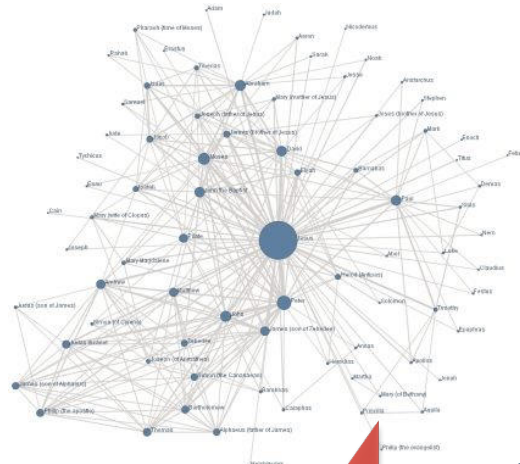
Temperado ao que os seus colegas chamam de "o aluno perfeito", ele não tem nada de excepcional. É apenas um jovem comum, com uma vida normal, que vive no bairro de Ipiranga, em São Paulo. Mas, apesar de ser um estudante comum, ele tem uma característica que o torna especial: ele é um aluno perfeito.

Desde a infância, ele demonstra uma facilidade excepcional para aprender idiomas. Aos sete anos, já falava inglês fluentemente. Aos dez, aprendeu francês. Aos quinze, começou a estudar espanhol. Hoje, além do português, fala fluente em inglês, francês e espanhol. Sua capacidade de aprender idiomas é considerada uma verdadeira maravilha.

Além de ser um aluno perfeito, ele também é um excelente estudante. Seus professores o elogiam constantemente por sua dedicação e desempenho. Ele sempre obtém notas altas em todas as disciplinas. Sua organização e disciplina são exemplos para todos os colegas.

Apesar de ser tão estudioso, ele não é um aluno sem vida social. Ele gosta de jogar futebol e tocar violão. Ele também gosta de ler livros e assistir filmes. Ele é um jovem equilibrado e feliz.

Al de lábio "O tal do que você...



ROBEM ALVES
Aluno perfeito

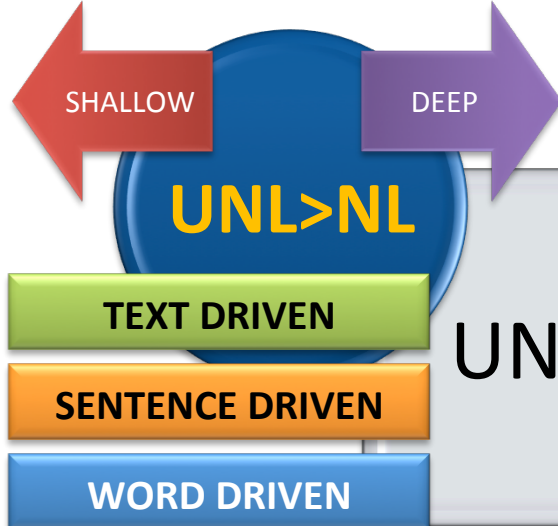
Temperado ao que os seus colegas chamam de "o aluno perfeito", ele não tem nada de excepcional. É apenas um jovem comum, com uma vida normal, que vive no bairro de Ipiranga, em São Paulo. Mas, apesar de ser um estudante comum, ele tem uma característica que o torna especial: ele é um aluno perfeito.

Desde a infância, ele demonstra uma facilidade excepcional para aprender idiomas. Aos sete anos, já falava inglês fluentemente. Aos dez, aprendeu francês. Aos quinze, começou a estudar espanhol. Hoje, além do português, fala fluente em inglês, francês e espanhol. Sua capacidade de aprender idiomas é considerada uma verdadeira maravilha.

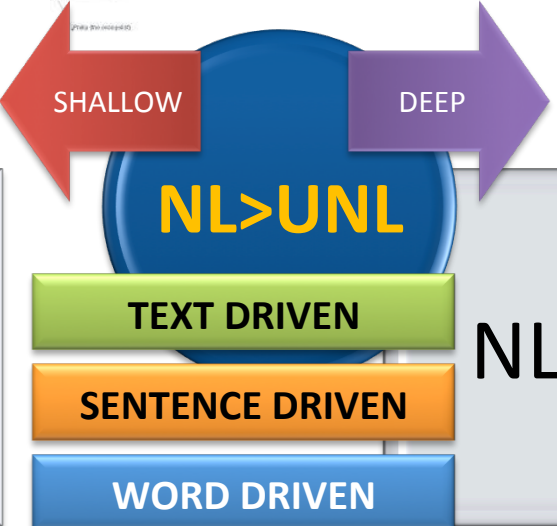
Além de ser um aluno perfeito, ele também é um excelente estudante. Seus professores o elogiam constantemente por sua dedicação e desempenho. Ele sempre obtém notas altas em todas as disciplinas. Sua organização e disciplina são exemplos para todos os colegas.

Apesar de ser tão estudioso, ele não é um aluno sem vida social. Ele gosta de jogar futebol e tocar violão. Ele também gosta de ler livros e assistir filmes. Ele é um jovem equilibrado e feliz.

Al de lábio "O tal do que você...



UNLization



NLization

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

The UNL^{versity}

The UNL^{versity}

- Training and Research Program
 - Internship Program in Computational Linguistics
 - UNL Panel
 - I UNL Panel (Mumbai, COLING 2012)
 - II UNL Panel (Sofia, ACL 2013)
 - UNL School
 - 2012
 - VIII UNL School (Geneva, February 2012)
 - IX UNL School (Mumbai, June 2012)
 - X UNL School (Alexandria, October, 2012)
 - 2013
 - XI UNL School (Macau, March 2013)
 - XII UNL School (Geneva, June 2013)
 - XIII UNL School (Yerevan, September 2013)
 - XIV UNL School (Kuwait, November 2013)

XI UNL School

Goals

- To build the basic modules of a NL-UNL (analysis) grammar for the corpus UCA₁
- To build the basic modules of a UNL-UNL (generation) grammar for the corpus UCA₁



Schedule

- March 11th
 - Introduction
 - Corpus
 - Dictionaries
- March 12th
 - Grammars
- March 13th
 - UNLization (with IAN)
- March 14th
 - NLization (with EUGENE)
- March 15th
 - 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/XI_UNL_School
- You may also use the forum (macau2013@unlweb.net) and the UNLCommunity to discuss problems and issues



Corpus

Corpus

- NC (NL>UNL)
- UC (UNL>NL)
 - A1 (100 sentences)
 - A2 (+300 sentences)
 - B1 (+ 5 texts)

Exercise #1 (10 min)

1. Translate the file `uca1_eng.txt` to your locale and name the resulting file `uca1_<your locale>.txt`. (pre workshop task)
2. Upload the file `uca1_<your locale>.txt` to:
 1. the UNL Community (Macau Group)
 2. to IAN (UNLWEB>UNLDEV>IAN>NL Files)

Exercise #2 (60 min)

1. Download the corpus `uca1_unl_exp.txt` from www.unlweb.net/resources/corpus/UCA1/uca1_unl_exp.txt.
2. Generate the corpus `uca1_unl_exp.txt` to your locale by hand (in a separate `.txt` file, one sentence per line, only the results). Name it `uca1_<your locale>_exp.txt`.
3. Upload the file `uca1_<your locale>_exp.txt` to the UNL Community (Macau group).

Exercise #3

1. Compare the results of:
 1. uca1_<your locale>.txt
 2. uca1_<your locale>_exp.txt
2. Present and explain the differences to the group

Dictionaryes

Universal Words

- nodes in the UNL graph
- temporary (not to be included in the UNL Dictionary)
 - 3.14159
 - www.undlfoundation.org
 - H₂O
- permanent (to be included in the UNL Dictionary)

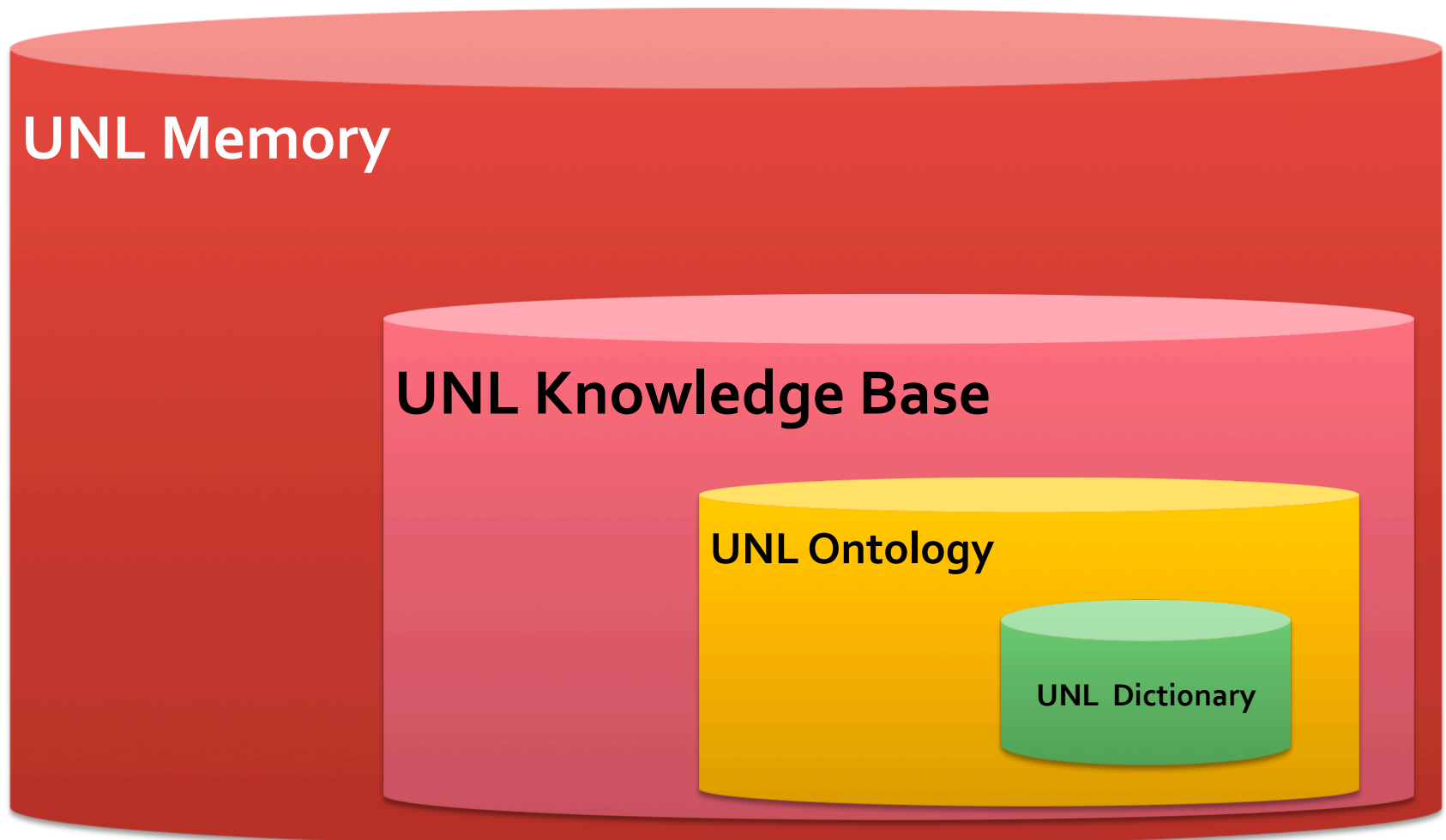
TYPE	GRANULARITY	UNL (simplified)	ENGLISH
SIMPLE	NODE	big	big
COMPOUND	NODE + ATTRIBUTE	big.@more	bigger
COMPLEX	HYPER-NODE	obj(affix,stamp)	to stamp

Semantic Accessibility

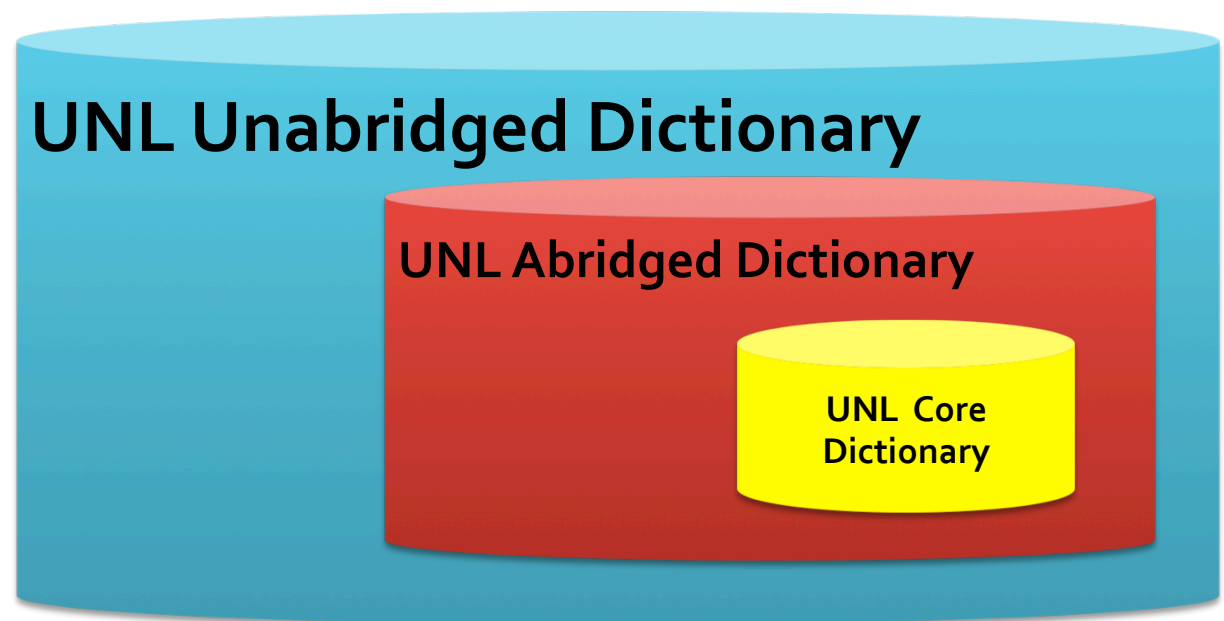
UCI (Uniform Concept Identifier)

- ❑ UCL (Uniform Concept Locator)
 - ❑ ucl://<AUTHORITY>/<ID>
 - ❑ ucl://unlkb.unlweb.net/104379964
- ❑ UCN (Uniform Concept Name)
 - ❑ ucn:<LID>:<NSS>
 - ❑ ucn:eng:table(icl>furniture)
 - ❑ ucn:fra:table(icl>mobilier)
 - ❑ ucn:esp:mesa(icl>mobiliario)
 - ❑ ucn:deu:Tisch(icl>Möbel)
 - ❑ ucn:rus:стол(icl>мебель)

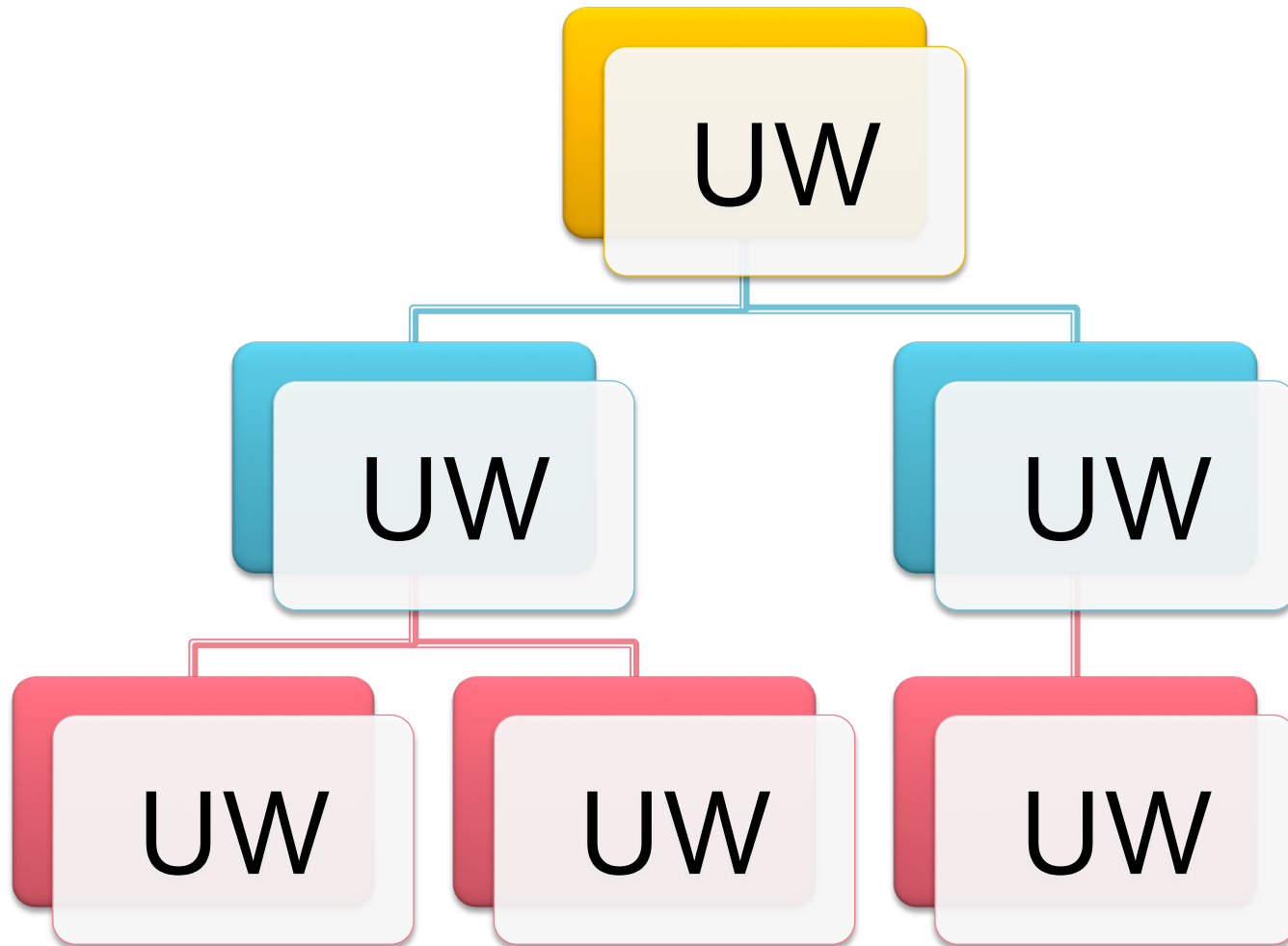
UNL Lexical Resources



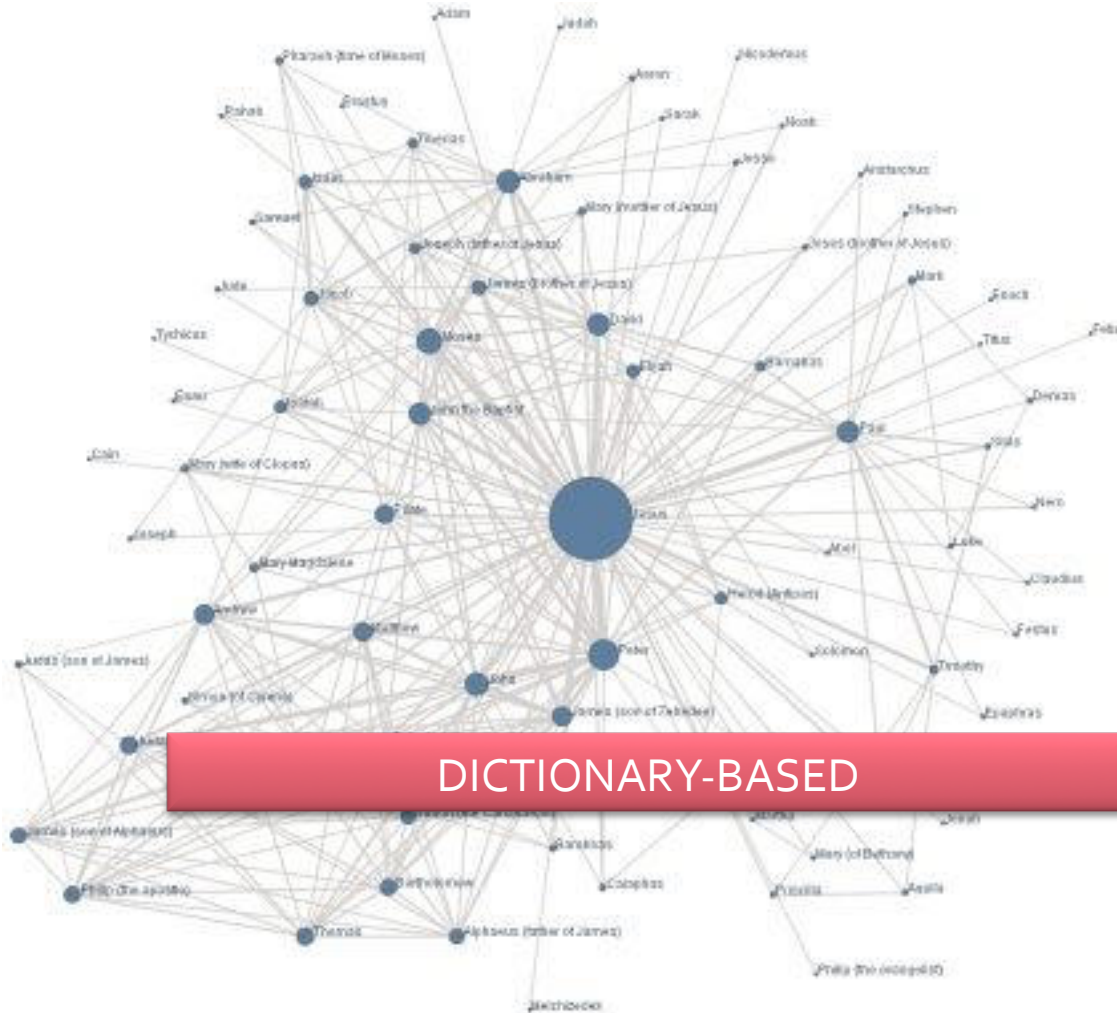
UNL Dictionary



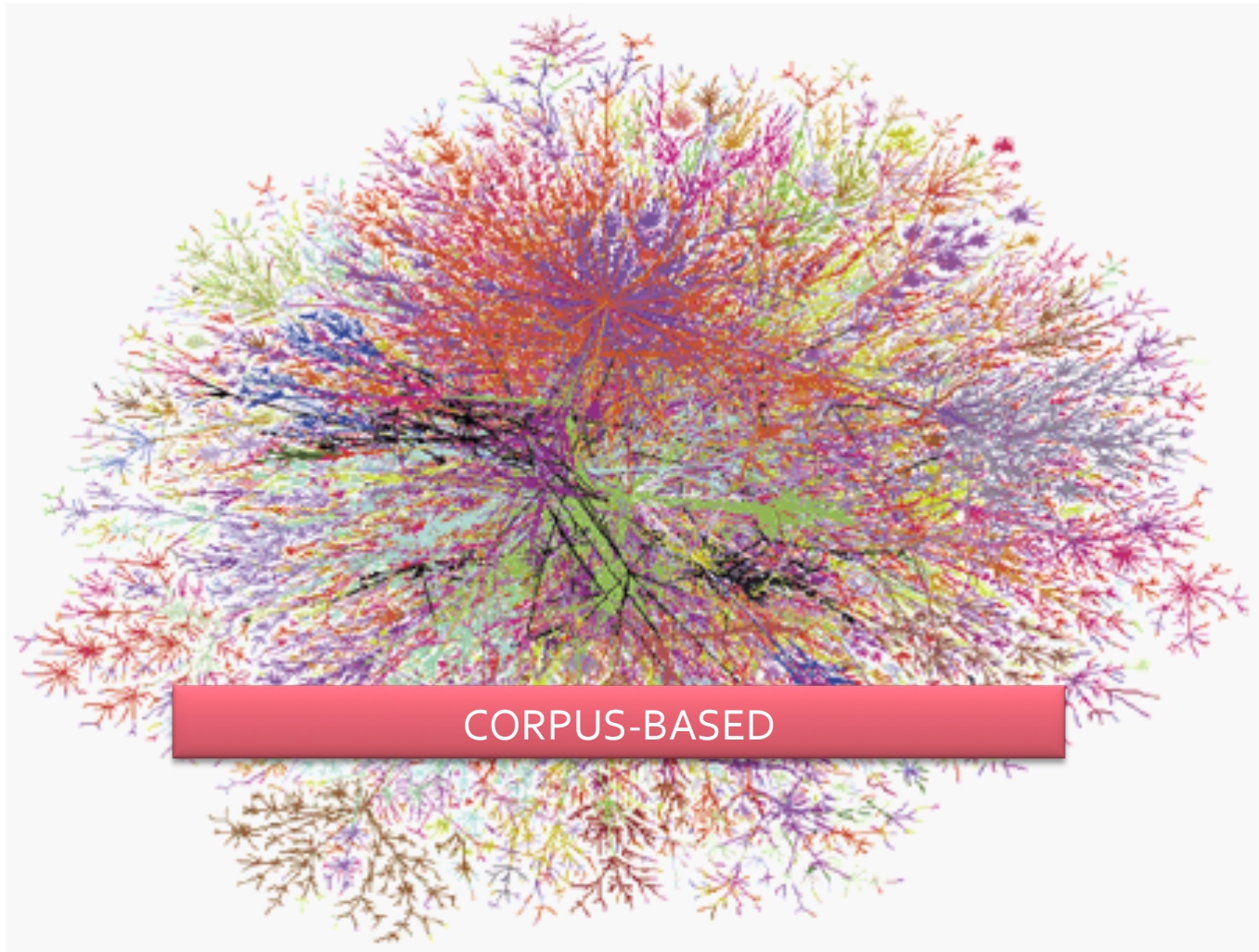
UNL Ontology



UNL Knowledge Base



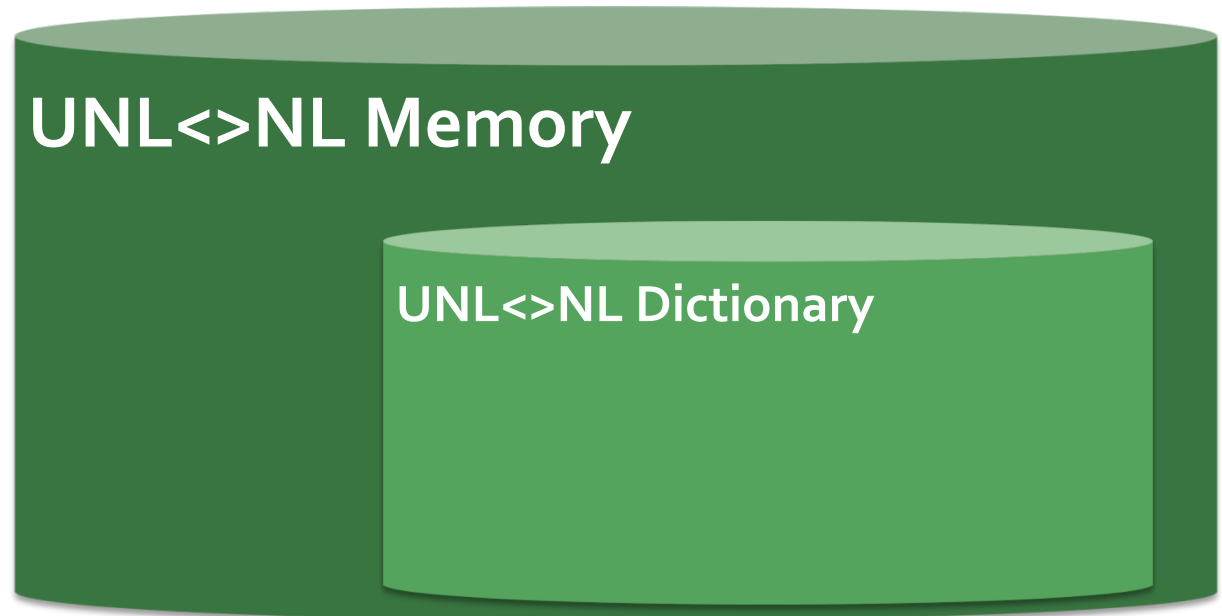
UNL Memory



NL Lexical Resources

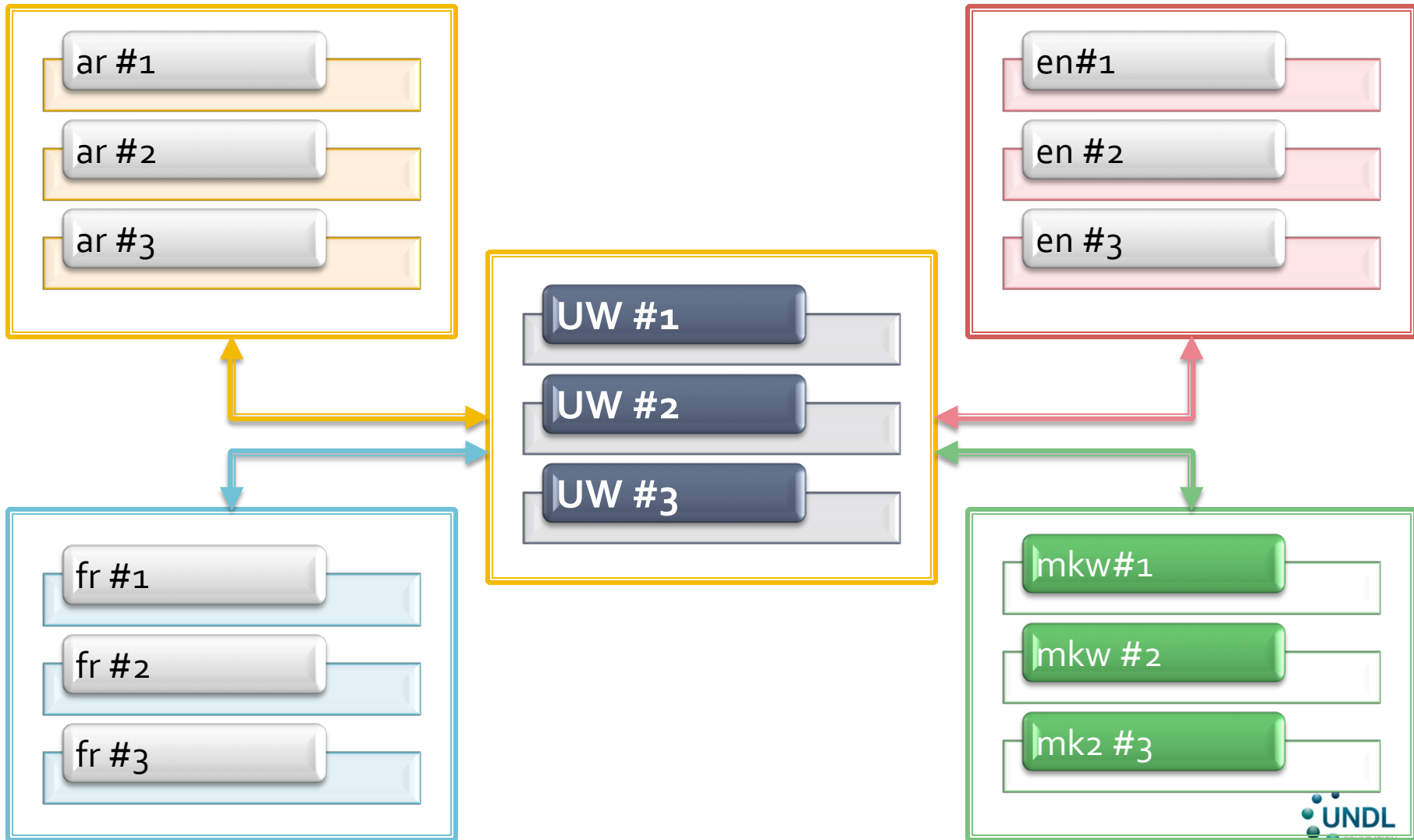


UNL<>NL Lexical Resources



UNL<>NL Dictionaries

UNL<->NL Dictionaries



UNL<->NL Dictionaries

- NL-UNL Dictionary (Analysis)

- Enumerative (word forms)

- [table] {2883} "table" (POS=NOU,NUM=SNG) <eng,0,0>;
 - [tables] {2883} "table" (POS=NOU,NUM=PLR) <eng,0,0>;

- [foot] {2883} "foot" (POS=NOU,NUM=SNG) <eng,0,0>;
 - [feet] {2883} "foot" (POS=NOU,NUM=PLR) <eng,0,0>;

- UNL-NL Dictionary (Generation)

- Generative (base forms)

- [table] {2883} "table" (POS=NOU,NUM=SNG,PAR=M2) <eng,0,0>;
 - [foot] {2883} "100284665" (POS=NOU,PAR=M1,FLX(PLR:="feet";)) <eng,0,0>;



Dictionary Specs

www.unlweb.net/wiki/Dictionary_Specs

- Dictionary Specs
 - Dictionary structure
 - a plain text file (.txt)
 - one entry per line
 - entries must have the following format:

```
[NLW] {ID} "UW" (ATTR , ... ) < LG , FRE , PRI > ; COMMENTS
```

[NLW]

[NLW] {ID} "UW" (ATTR, ...) < LG, FRE, PRI >; COMMENTS

- a multiword expression: [United States of America]
- a compound: [hot-dog]
- a simple word: [happiness]
- a simple morpheme: [happ]
- a complex structure: [[bring] [back]]
- a non-motivated linguistic entity: [g]

{ID}

[NLW] {ID} "UW" (ATTR, ...) < LG, FRE, PRI >; COMMENTS

- The unique identifier (primary-key) of the entry.

"UW"

[NLW] {ID} "UW" (ATTR, ...) < LG, FRE, PRI >; COMMENTS

- The Universal Word of UNL. This field can be empty if a word does not need a UW.

(ATTR, ...)

[NLW] {ID} "UW" (ATTR, ...) < LG , FRE , PRI >; COMMENTS

- The list of features of the NLW.
- Attributes should be separated by “,”.
- It can be:
 - a list of simple features: (NOU, MCL, SNG)
 - a list of attribute-value pairs: (pos=NOU, gen=MCL, num=SNG)
 - a list of transformation rules : (plural:="oo":"ee")
 - Replacement
 - <ATTRIBUTE>":="<SOURCE>":"<TARGET>
 - plural:="oo":"ee"
 - Left appending
 - <ATTRIBUTE>":="<LEFT DELETION>"<"<LEFT ADDITION>
 - not:="<"un"
 - Right appending
 - <ATTRIBUTE>":="<RIGHT ADDITION>">"<RIGHT DELETION>
 - plural:="y">ies"

<LG, FRE, PRI>

[NLW] {ID} "UW" (ATTR, ...) **< LG , FRE , PRI >;** COMMENTS

- **FLG**
 - The three-character language code according to ISO 639-2.
- **FRE**
 - The frequency of NLW in natural texts. Used for natural language analysis (NL-UNL). It can range from 0 (less frequent) to 255 (most frequent).
- **PRI**
 - The priority of the NLW. Used for natural language generation (UNL-NL). It can range from 0 to 255.

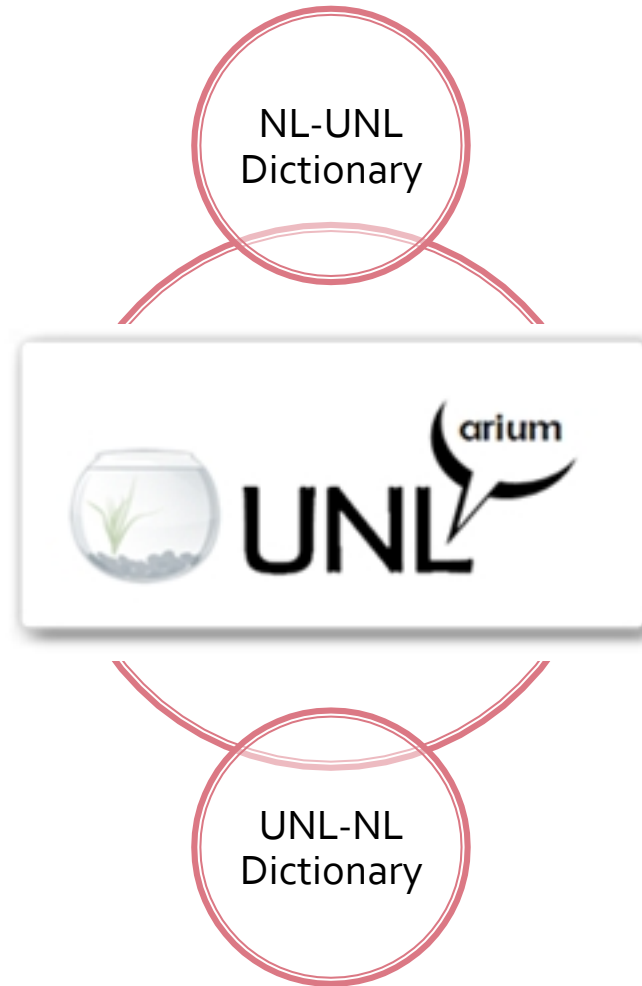
Examples

- [China]{24} "China(iof>Asian country)" (NOU, WRD, SNG, Po, Fo) <eng,0,0>;
- [choose]{106} "to choose(icl>to decide)" (POS=VER, LEX=WRD, PAR=M1, FRA=Y76, FLX(3PS&PRS&IND:=o>"s"; PAS:="chose"; PTP:="chosen"; GER:="choosing";)) <eng,0,0>;
- [clear-eyed]{25} "clear-eyed(icl>discerning)" (POS=ADJ, LEX=WRD, PAR=Mo, FRA=Yo) <en,0,0>;
- [Peter]{177} "Peter(iof>person)" (NOU) <eng,10,30>;
- [kill]{5987} "kill(icl>do)" (FLX(PAS:=o>"ed";)) <eng,70,80>;
- [[bring] [back]]{2345} "bring back" (POS=VER, VA(01>02), #01(POS=VER, FLX(PAS:=3>"ought";)), #02(POS=PRE)) <eng,50,34>;
- [/br(ing|ought)/] "bring(icl>do)" (POS=VER) <eng,0,0>;
- [[/br(ing|ought)/] [back]]{2345} "bring back(icl>do)" (POS=VER, #01(POS=VER), #02(POS=PRE)) <eng,50,34>;
- [/colo(u)?r/] "color" (POS=NOU) <eng,0,0>; (NLW = {color, colour})
- [/cit(y|ies)/] "city" (POS=NOU) <eng,0,0>; (NLW = {city, cities})
- [/\d{4}/] "" (ENT=YEAR) <eng,0,0>; (NLW = any sequence of four digits)
- [city] "/city(.)*/" (POS=NOU) <eng,0,0>; (UW = any UW that starts by the string "city")
- [city] "/(.)+\ (iof>city\)/" (POS=NOU) <eng,0,0>; (UW = any UW that ends by the string "(iof>city)")

Special issues (Null UW)

- [extremely]{}"extremely"(LEX=A,POS=SAV)<eng,0,0>;
 - an extremely beautiful car
- [very]{}"" (LEX=A,POS=SAV,att=@plus)<eng,255,0>;
 - very beautiful
- [of]{}"" (LEX=P,POS=PRE,rel=mod)<eng,255,0>;
 - the book of John
- [after]{}"" (LEX=P,POS=PRE,rel=tim,att=@after)<eng,255,0>;
 - few minutes after noon

Building dictionaries



Exercise #4 (10 min)

- Create the UNLization dictionary for the corpus UCA1_<your locale>.txt. (pre-workshop task)
- Upload the dictionary to IAN
 - (UNLWEB>UNLDEV>IAN>DICTIONARY)
- Run IAN for the dictionary and corpus (without any rule) and analyze the results

Exercise #5 (10 min)

- Create the NLization dictionary for the corpus UCA1_unl.txt (pre-workshop task)
- Upload the dictionary to EUGENE
 - (UNLWEB>UNLDEV>EUGENE>DICTIONARY)
- Run EUGENE for the dictionary and corpus (without any rule) and analyze the results