

XIV UNL School

Patras, March 10-14, 2014



Welcome

Participants

- Aikaterini Tsiouma (Modern Greek)
- Anastasia Glaraki (Ancient Greek)
- Anna Loukopoulou (Modern Greek)
- Athina Papachrysostomou (Ancient Greek)
- Daniel Maslovsky (Czech)
- Efi Zacharopoulou (Modern Greek)
- Eirini Kouvara (Ancient Greek)
- Eleni Metheniti (Modern Greek)
- Evangelos Katsarelis (Ancient Greek)
- Ewa Czajkowska (Polish)
- Filippos-Panagiotis Xyrafas (Latin)
- Foteini Papadimitriou (Latin)
- Georgia-Rengina Loukatou (Modern Greek)

- Katerina Ntotsika (Ancient Greek)
- Maria Antonopoulou (Ancient Greek)
- Maria Kotopouli (Ancient Greek)
- Meritxell Fernandez (Catalan)
- Natalia Pavlou (Modern Greek)
- Nikolaos Tsoukalas (Ancient Greek)
- Nikolitsa Glaraki (Ancient Greek)
- Panagiota Taktikou (Latin)
- Ronaldo Martins (UNL)
- Sameh Alansary (Arabic)
- Sigita Dereskeviciute (Lithuanian)
- Stamatia Androna (Latin)
- Vasiliki Saloustrou (Latin)





XIV UNL School

Goals

 To teach you how to build the basic infrastructure for UNLization and NLization



Program

- March 10th
 - Introduction
 - Corpus
- March 11th
 - Grammar: Morphology
- March 12th
 - Dictionaries
- March 13th
 - Grammar: Syntax
- March 14th
 - Evaluation & Discussion



Timetable

10th	11th	12th	1 3th	14th
10-12	9-12	9-12	9-12	9-12
14-17	14-17	14-17	14-17	

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/XIV_UNL_School



Day #1

Morning

- Introduction
- The UNL
- The UNL System
- The UNDL Foundation Road Map

Lunch break

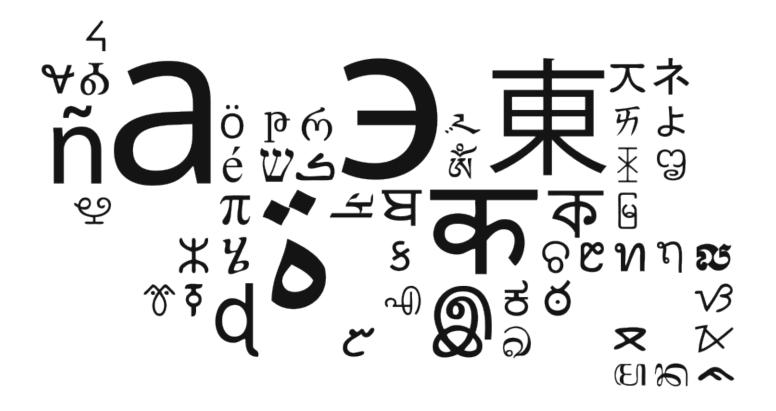
Afternoon

- Corpus
 - UNLization
 - NLization

Introduction

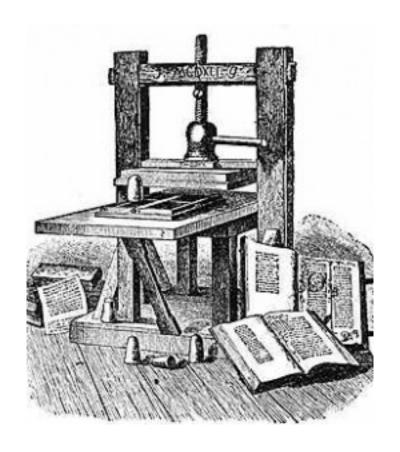


Language was the first information technology



the technology of **Writing** released information from space and time constraints.





the invention of the **printing** press made the information less expensive





With **broadcasting**, information became instantaneously available worldwide.





With the **HTML** information became extremely cheap, easily storable, readily distributable





The Information society is not the knowledge society: information is often

immoderate, unreliable, biased.



A new revolution is needed

ANALOGIC LANGUAGE

C2 cotidiano TERCA-FEIRA, 22 DE JANEIRO DE 2007 FOLHA DE SPAULO

E RA UMA war um jowem casal que estava muito feliz. Ela es-tava gràxida, e eles esperavam com grande ansiedade o filho que iria

Intracormotos os nove mises de gravides, ele nosce. Ela des à lugur parheiros reclamovars, disiam que la focorquistades Que felicidade tra apadas colas que lha ser ame restra-um comprotador como filhat Eno de serio faciam amitico. Sua initialho que desejarum test Per isos els- glantam cuestram-ne agrander. El-hastin resudo misió durante todas i ravam motes a ruins. Picavam de para uma boa educação gravidez, chegando mesmo a fager recuperação,

para passar no vestibular.

professores ensinavam. Mas tudo trangeiras. Sabia de cor todas as in-mesmo. E não reclamava. Seus com-formações sobre o mundo cultural. bular. O cursinho que ele frequenta-

ra publicou sua fotografia em outdoes. Apareceu na televisão como complo a ser seguido por todos os des?" Que mais prazer the des?" iovens. Na universidade, foi a mesma coisa. Só tirava dez. Chegou, finalmente, o dia tio omerado a formetua. Memorioso foi o grande herdi, elogiado pelos professores. Ganhou medalhas e mesmo uma boisa para doutoramento no MIT.

Depois-da cerimônia acadômica foi asta. E estreuen todos feliosa po-viu um chiado estranho dentro de Depos du cerrimonia academica lete a fosta. E categori de cerrimon de menora se aproseimou da Mermuriaso e se aproserioru man da Mermuriaso e se aproserioru con servicio de consultare porgunta?" "Pode fixer", disse Me-morioso confiante. Sua memória continha todas as respostas.

Ai ela falou: "De tudo o que você ram que seu disco rigido estava inte-

RUBEM ALVES

O aluno perfeito

subiu. E, de repente, seus olhos fica-

a estimalos. Depois apagou, entrou

em coma. Levado às pressas para o hospital de computadores, verifica

gravace, cheganos meams a faser recuperação.

Ses nola occutacia com Memorios por poramensas.

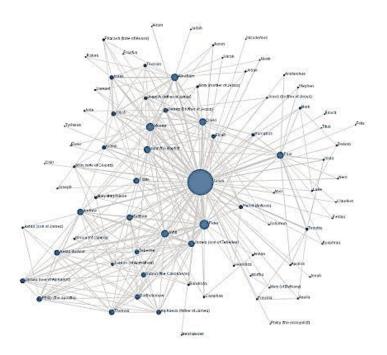
O batinado foi uma festuma, Derum-lho o nome de Memorioso, porcididade a maneira de exturir mia uma
equipidagement qua uma memério peresta é o essencia plaza uma bos endesta é o essencia plaza uma bos endesta é o essencia plaza uma bos endesta é a festa, a chaterias pognificos, escação. Educação é memoriosojo,
populações de palese longitupos, deproduce de la composição porte de porte.

Para o este pais. E or outros casis,
para o este pais. E or outros casis,
para o este pais. Es or outros casis,
para o este para de la casis para de la Crianças com memória perfeita vilo tas de eventos históricos, nomes de so, morriam de inueja. Quando filhos bem na escola e não têm problemas reis, imperadores, revolucionários, chegavam em casa trazendo bolettas santos, escritores, descobridores, com notas em vermelho eles grit-E foi isso mesmo que acontecea, cientístas, pelavras novas, regras de varn: "por que vocá não é camo o Me Memorioso memorizava tudo que os gramática, Divos Inteiros, linguas es-mociosos"

Memorioso foi o primeiro no vest

memorizou qual foi aquilo que você-Há penguntas para as quais a memória não tem respostas . É que tais respostas não se encontram na momória. Encontram-se no ceração Memorioso ficou mudo. Os circuitos de sua memória funcionavam code moro aemocão. com a velocidade da luz procurando a resposta. Mas aquilo não lhe fora ensiredo. Seu rosto ficou vermelho Começou a suar. Sua temperatura

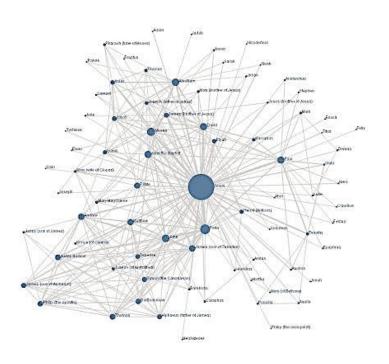
DIGITAL LANGUAGE



SEMANTIC NETWORK

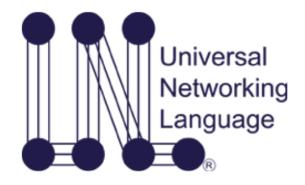


Uses of a semantic network



- Organize information
- Search for information
- Extract information
- Summarize information
- Simplify information
- Translate information





The Universal Networking Language (UNL)

The distinctiveness of UNL

COMMON TO ALL MEMBERS OF A GROUP



ADAPTED TO VARIOUS PURPOSES, SIZES, FORMS, OPERATIONS



to represent what was meant, i.e.,
the meaning,
and not what was said.



UNL

translation knowledge representation

⇔1996







Commitments

- The UNL must represent information
 The UNL must represent "what was meant" (and not "what was said").
- The UNL must be a language for computers
 The UNL must be computable.
- 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.
- 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_i and the monkey_i ate the banana_i

Structure

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

CONCEPTS

= Universal Words (UWs)

CONCEPT SPECIFIERS

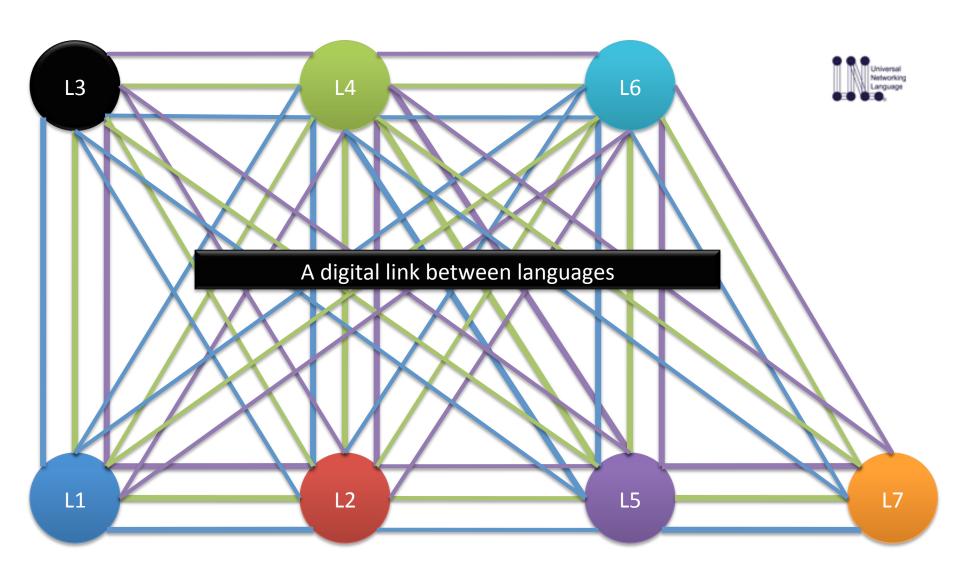
= Universal Attributes

RELATIONS BETWEEN
CONCEPTS

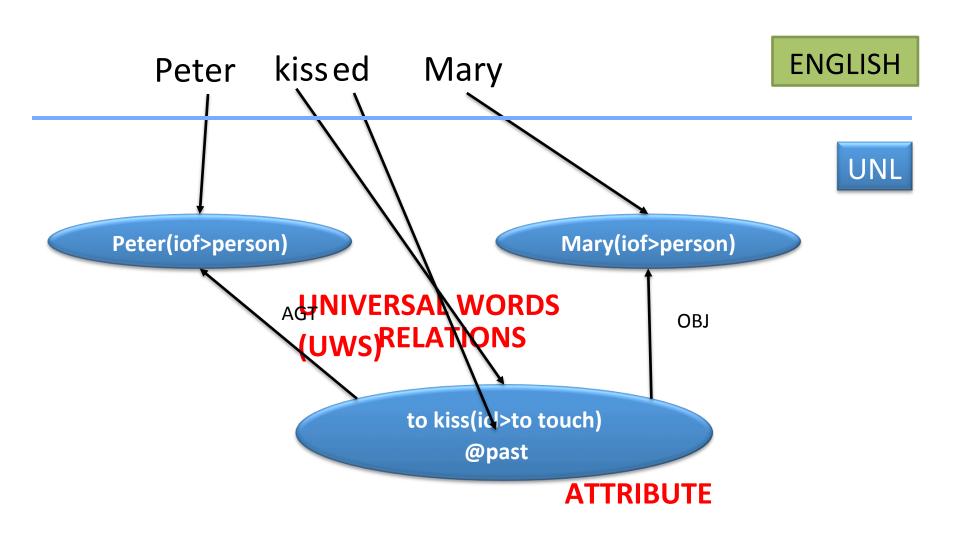
= Universal Relations



The Universal NETWORKING Language

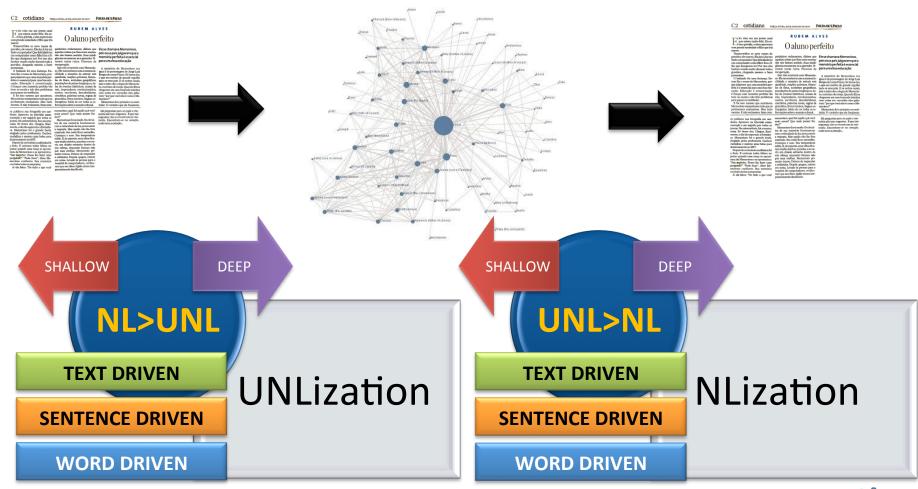


Natural Language-to-UNL (UNL-ization)



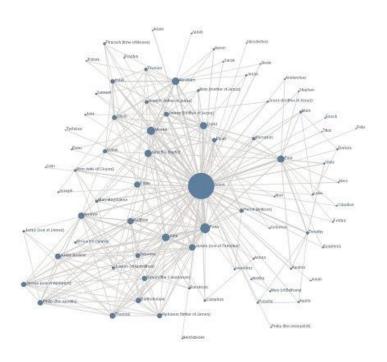
The UNL System

The UNL System





Uses of UNL



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

The UNDL Foundation Road Map

BACK-END APPLICATIONS

RESOURCES



multilingual dictionary





information extraction



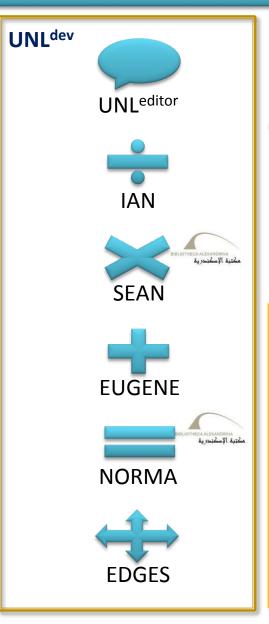


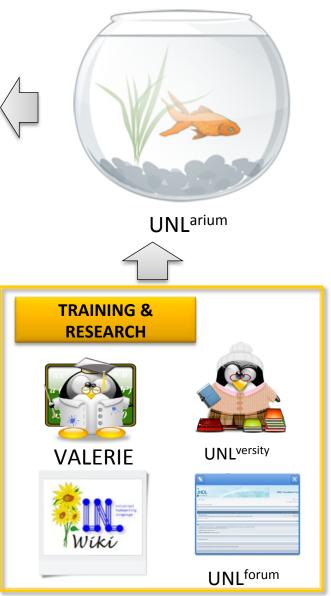
simplification and summarizaiton





translation

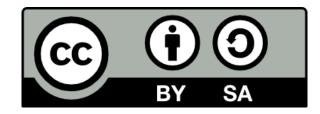




Language Resources (lingware)

Crowdsourcing





25 projects52 languages

>10 M dictionary entries >50 K grammar rules >150 K entries in the knowledge base >300 K corpus tokens



FoR-UNL

LEVEL	DICTIONARY	GRAMMAR
A1	5,000	Morphology: N
A2	10,000	Morphology: J, A, V
B1	20,000	Syntax: NP
B2	40,000	Syntax: VP
C1	70,000	Syntax: IP
C ₂	100,000	Syntax: CP



RESOURCES



 UNL^{arium}

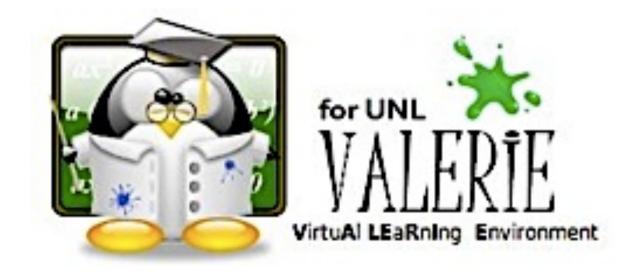






Distance Education

VALERIE (VirtuAl LEaRnIng Environment)



6 programs >800 users certified



Documentation open and public





Open discussion

UNLforum

UNLweb | UNLarium | UNLcommunity | UNLdev | UNLforum | UNLlab | UNLwiki | VALERIE | Contact

In total there are 2 users online :: 1 Registered, 0 Hidden and 1 Guest [Administrator] [Moderator]

Our users have posted a total of 2197 articles

Most users ever online was 30 on Sat Nov 16, 2013 1:39 pm

We have **1117** registered users The newest registered user is **olepar**

Registered Users: martins

This data is based on users active over the past five minutes

Universal Welcome to the UNL Forum Networking FAQ Search Memberlist Usergroups Profile You last visited on Tue Nov 26, 2013 6:40 pm View posts since last visit The time now is Tue Nov 26, 2013 6:40 pm View your posts Forum Index View unanswered posts **Last Post** Forum **Forums** Newsletter Tue Nov 26, 2013 3:57 pm 34 34 martins →D Moderator martins UNL +3 Tue Nov 26, 2013 5:58 pm Forum of the managers of the project UNL +3 martins →□ **UNL Open Forum** Wed May 15, 2013 3:09 pm 73 The UNL Open Forum is a place for discussing problems and suggestions concerning UNL and UNLweb-related matters martins → UNLization Sat Sep 14, 2013 4:04 pm 144 303 chengyong →□ Forum for discussing linguistic issues related to mapping natural language into UNL Wed Jul 17, 2013 9:17 am 30 Forum to test new features. Kamal →D Moderator Kamal Mark all forums read All times are GMT + 1 Hour Who is Online



Ronaldo Martins | Logout

Interaction and information exchange UNL community

UNLweb | UNLarium | UNLcommunity | UNLdev | UNLforum | UNLlab | UNLwiki | VALERIE | Contact Ronaldo Martins | Logout **UNL**web UNL Community <u>*</u> [4] 💌 Search Ronaldo Martins Home Members Groups Administration **News Feed** News Feed 🤼 Friends Share: 🧼 Status 📾 Bookmark 🚆 Blog 📗 File A Yerevan 2013 140 characters remaining UNL2015 🧘 test Post Create Group... See all Somdev Kar is now a friend with supeeti kulchan Photos m Bookmarks Blogs 4 days ago · 🦀 Pages Files Alfonso Javier Canosa Rodriguez is now a friend with sumaira nabi The Wire Bookmark this page 5 days ago · 📥 & Report this





the UNL training and research program



In-site Education



- ✓ VIII UNL School
 6-10 Feb 2012, Geneva
- ✓ IX UNL School

 11-15 Jun 2012, Mumbai
- ✓ X UNL School

 7-11 Oct 2012, Alexandria
- ✓ XI UNL School
 11-15 Mar 2013, Macau
- ✓ XII UNL School
 1-5 Jul 2013, Geneva
- ✓ XIII UNL School 23-27 Sep 2013, Yerevan
- ☐ XIV UNL School
 10-14 Mar 2014, Patras



Competition & Evaluation

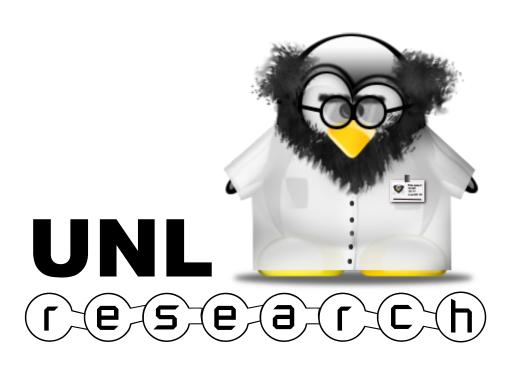


- ✓ I UNL Olympiad: 38 participants
- ✓ II UNL Olympiad: 44 participants
- ✓ III UNL Olympiad (open)





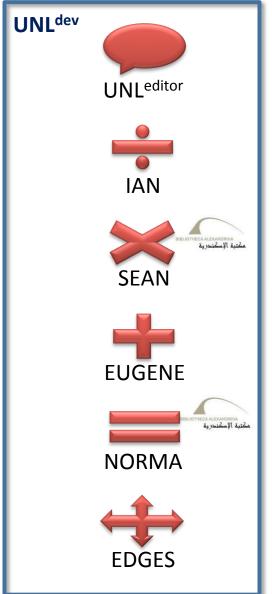


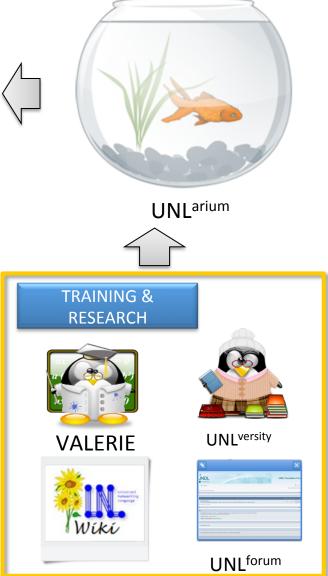




BACK-END APPLICATIONS

RESOURCES







Collaborative Software Development





UNL Applications (for the general public)





FRONT-END APPLICATIONS

BACK-END APPLICATIONS

RESOURCES



multilingual dictionary





information extraction





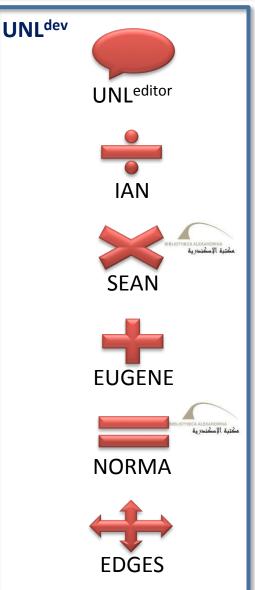


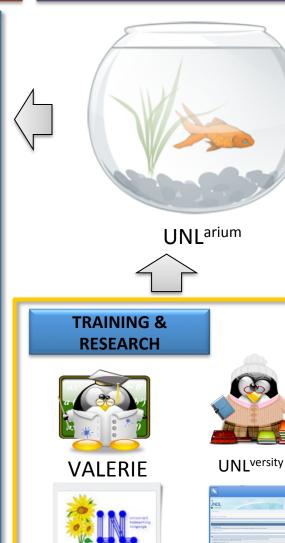
simplification and summarizaiton





translation





Wiki



UNLforum

MUltilingual Harmonized dlcTionary



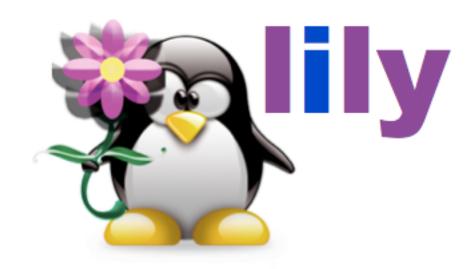


Knowledge Extraction sYStem





Translation



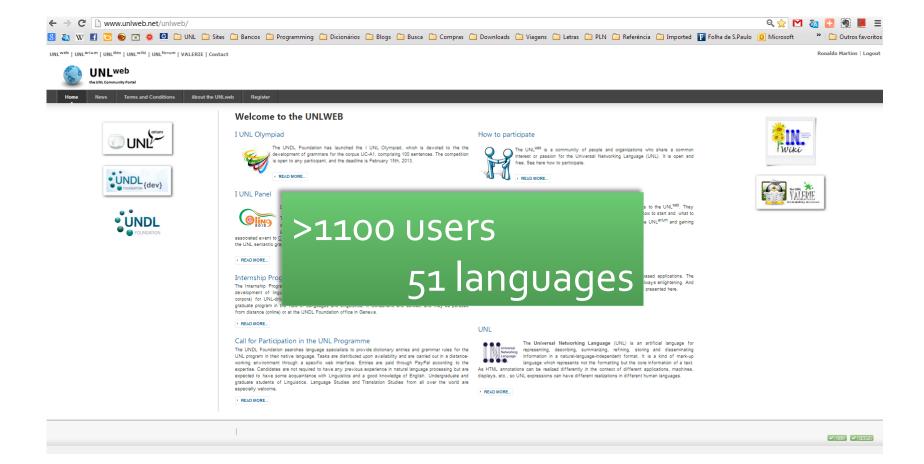


Summarization and Simplification





www.unlweb.net



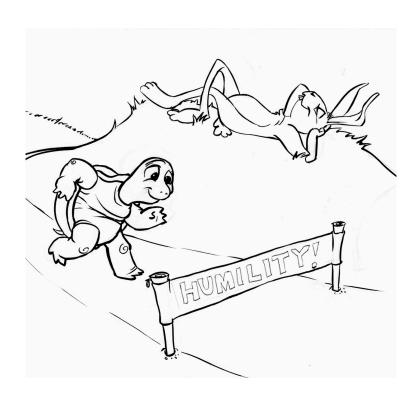


Questions?



Corpus

The Hare and the Tortoise



The Hare one day ridiculed the short feet and slow pace of the Tortoise.

The Tortoise replied:

"Though you be swift as the wind, I will beat you in a race."

The Hare believed her assertion to be simply impossible and assented to the proposal.

They agreed that the Fox should choose the course and fix the goal.

On the day appointed for the race the two started together.

The Tortoise did not stop.

She went on with a slow but steady pace straight to the end of the course.

The Hare laid down by the wayside and took a nap under a tree.

At last, he woke up and ran as fast as he could, but it was too late.

The Tortoise had already won the race.

Slow but steady progress wins the race.

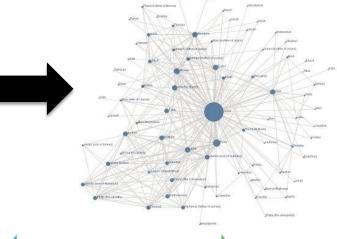
UNLization

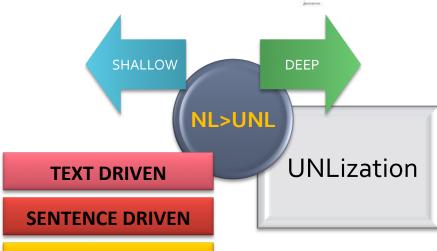
UNLization



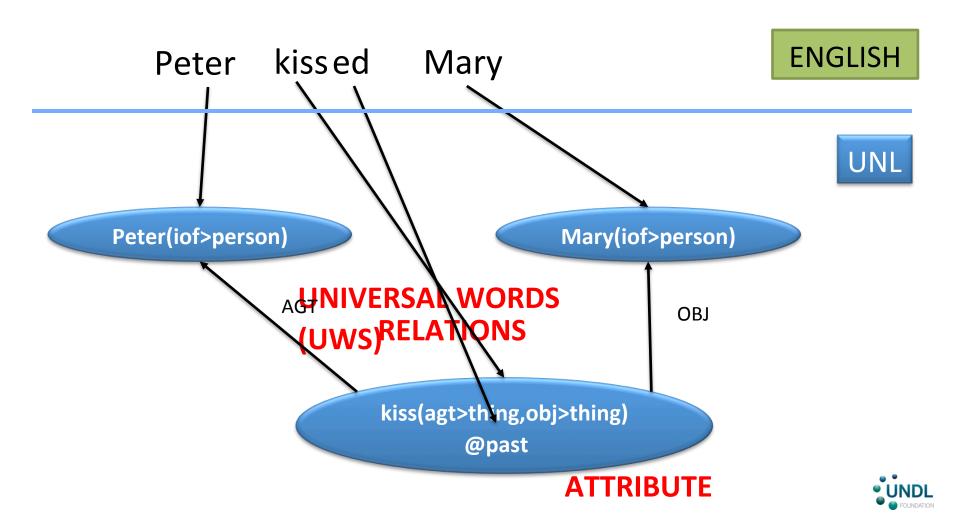
O aluno perfeito

WORD DRIVEN

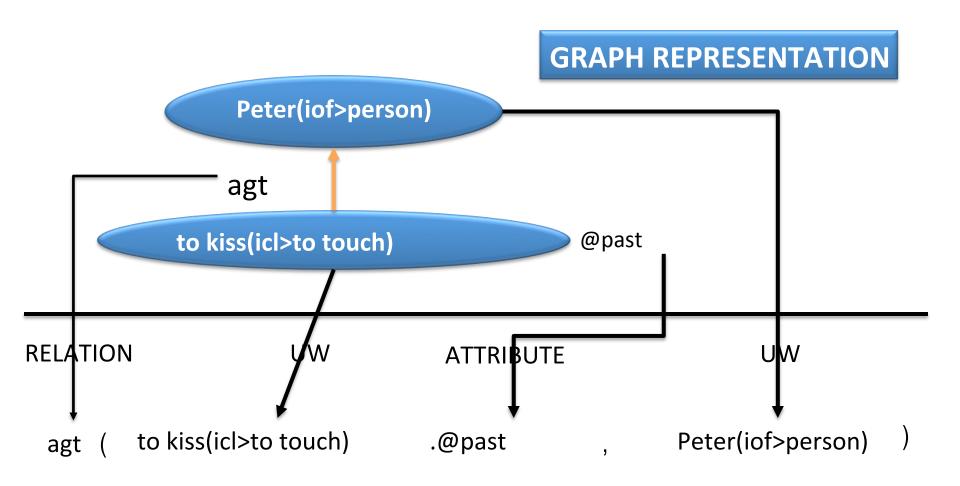




Natural Language-to-UNL (UNL-ization)



Syntax of UNL





UNL document

```
[D]
[S:01]
{org:en}
Peter kissed Mary.
{/org}
{unl}
agt(to kiss(icl>to touch).@past, Peter(iof>person))
obj(to kiss(icl>to touch).@past, Mary(iof>person))
{/unl}
[/S]
[/D]
```



UNLization

NL	UNL
the hare	hare.@def
a tree	tree.@indef
slow progress	mod(progress, slow)
the short feet	mod(foot.@def.@pl, short)
the end of the course	pof(course.@def, end.@def)
the short feet of the Tortoise	pof(tortoise.@def, foot.@def.@pl)
the Hare and the Tortoise	and(hare.@def, tortoise.@def)
slow but steady progress	mod(progress, :01) and:01(slow, steady.@contrast)
the Tortoise replied	agt(reply.@past, tortoise.@def)
the Tortoise went on	agt(go on.@past, tortoise.@def)

activity

Aesop-1

Aesop-1

 NLize the sentences of the project Aesop-1 available at the www.unlweb.org/sandbox.

Important!

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

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

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))

Learning Environment

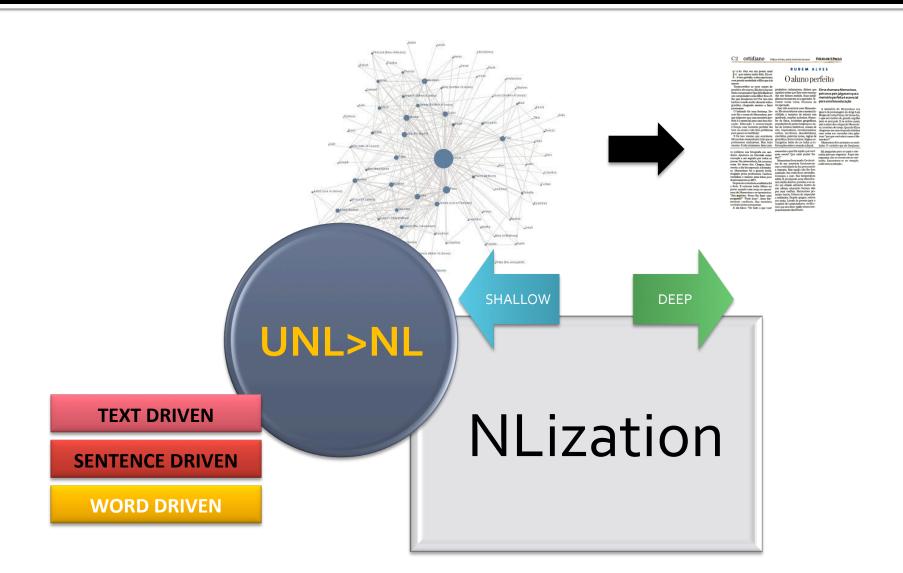


Discussion #1



NLization

NLization



NLization

UNL	NL
hare.@def	the hare
tree.@indef	a tree
mod(progress, slow)	slow progress
mod(foot.@def.@pl, short)	the short feet
pof(course.@def, end.@def)	the end of the course
pof(tortoise.@def, foot.@def.@pl)	the short feet of the Tortoise
and(hare.@def, tortoise.@def)	the Hare and the Tortoise
mod(progress, :01) and:01(slow, steady.@contrast)	slow but steady progress
agt(reply.@past, tortoise.@def)	the Tortoise replied
agt(go on.@past, tortoise.@def)	the Tortoise went on

Activity

Aesop-2

Aesop-2

 NLize the graphs from the project Aesop-2, available at the UNLsandbox

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. atrial = 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)

Discussion #2

