Ronaldo Martins

Tradução Automática:
faça você mesmo

Roteiro

- Podem as máquinas falar?
- Podem as máquinas traduzir?
- Estudo de caso: UNL
- Faça você mesmo

Podem as máquinas falar?

O que é "falar"?

- Quando falar é fazer:
  - ler textos em voz alta
  - transcrever textos orais
  - revisar textos
  - localizar informações
  - resumir textos
  - traduzir textos
  - etc.

O lugar da Lingüística Computacional

- Uma disputa onomasiológica:
  - Lingüística Computacional
  - Lingüística Matemática
  - Lingüística Informática
  - Processamento de Linguagem Natural
  - Engenharia da Linguagem
- Uma disputa territorial:
  - Isotopia x Heterotopia

O lugar da Lingüística Computacional

- O objeto da Lingüística Computacional:
  - *ergon* x *energeia*
  - imitação x emulação
- Os objetivos da Lingüística Computacional:
  - Acelerar
  - Ampliar
  - Substituir

Podem as máquinas traduzir?

Podem as máquinas traduzir?

Q: Can a kangaroo jump higher than the Empire State Building?

A: Yes, because the Empire State Building can't jump!

Q: Pode um canguru saltar mais alto do que o Empire State Building?

R: Sim, porque o Empire State Building não pode saltar!

Podem as máquinas traduzir?

A: Doctor, will I be able to play the piano after the operation?

B: Yes, of course.

A: Great! I never could before!

A: Doutor, eu vou ser capaz de tocar piano após a operação?

B: Sim, claro.

A: Ótimo! Eu nunca poderia antes!

Podem as máquinas traduzir?

When I was young I didn't like going to weddings. My grandmother would tell me, "You're next". However, she stopped doing that after I started saying the same thing to her at funerals.

Quando eu era jovem não gostava de ir a casamentos. A minha avó ia me dizer, "Você está próximo". No entanto, ela parou de fazer isso depois que eu comecei dizendo a mesma coisa com ela nos funerais.

Podem as máquinas traduzir?

Teacher: Maria, please point to America on the map.

Maria: This is it.

Teacher: Well done. Now class, who found America?

Class: Maria did.

Professora: Maria, por favor aponte para a América no mapa.

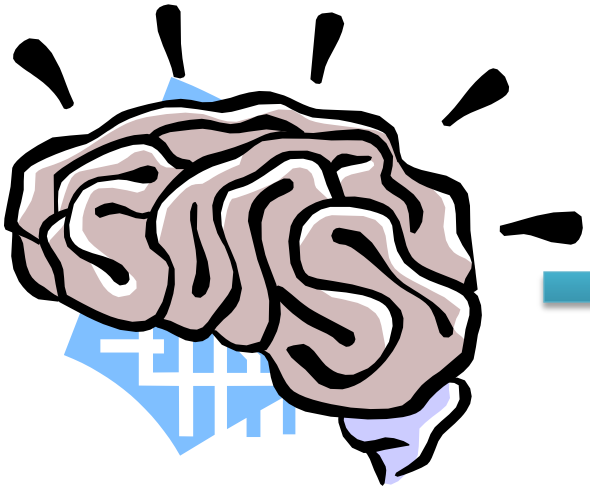
Maria: Isso é que é.

Docente: Bem feito. Agora classe, que descobriu América?

Classe: Maria fez.

O problema

tradução automática



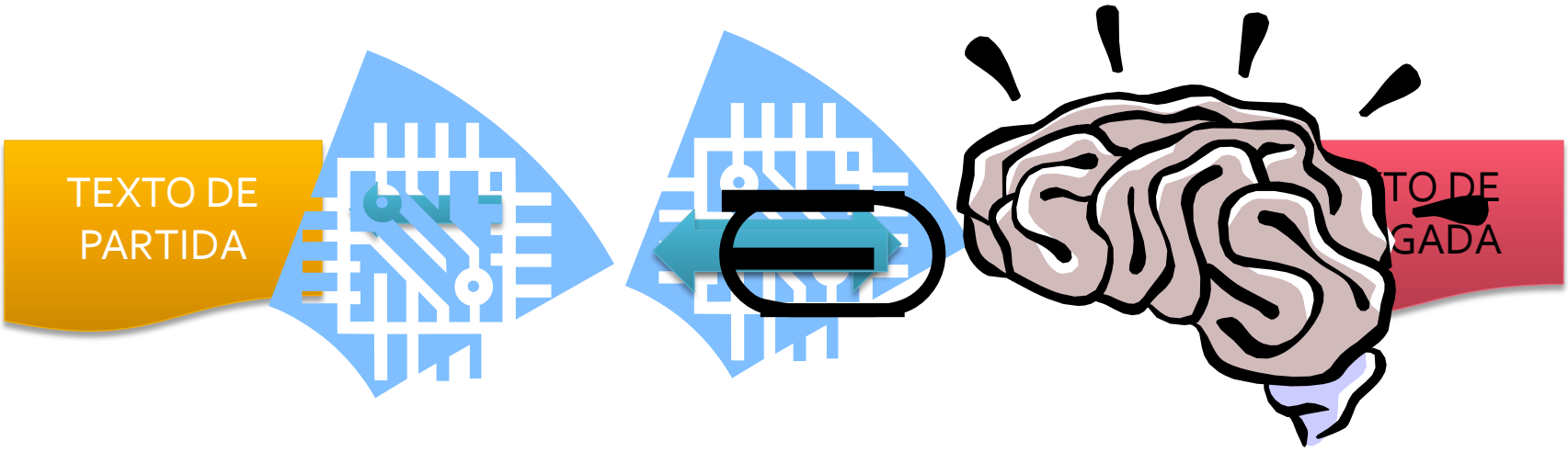
TEXTO DE PARTIDA

TEXTO DE CHEGADA

Como emular a habilidade lingüística de traduzir textos?

O problema do problema

tradução automática



Como emular a habilidade lingüística de traduzir textos?

História da TA

1949 1960 1970 1980 1990 2000

MICROSOFT TRANSLATOR

995)

setting

Caterpillar document workflow (mid-90s)

Knowledge-based system

Designed for translation of technical documents written in Caterpillar Technical English (CTE) to French, Spanish, and German

Controlled English – no pronouns, conjunctions,…

reader to learn this patois in order to understand what the Russian actually wrote. Learning Russian would not be much more difficult.

in some not too

remote future.

TA hoje

Metodologia

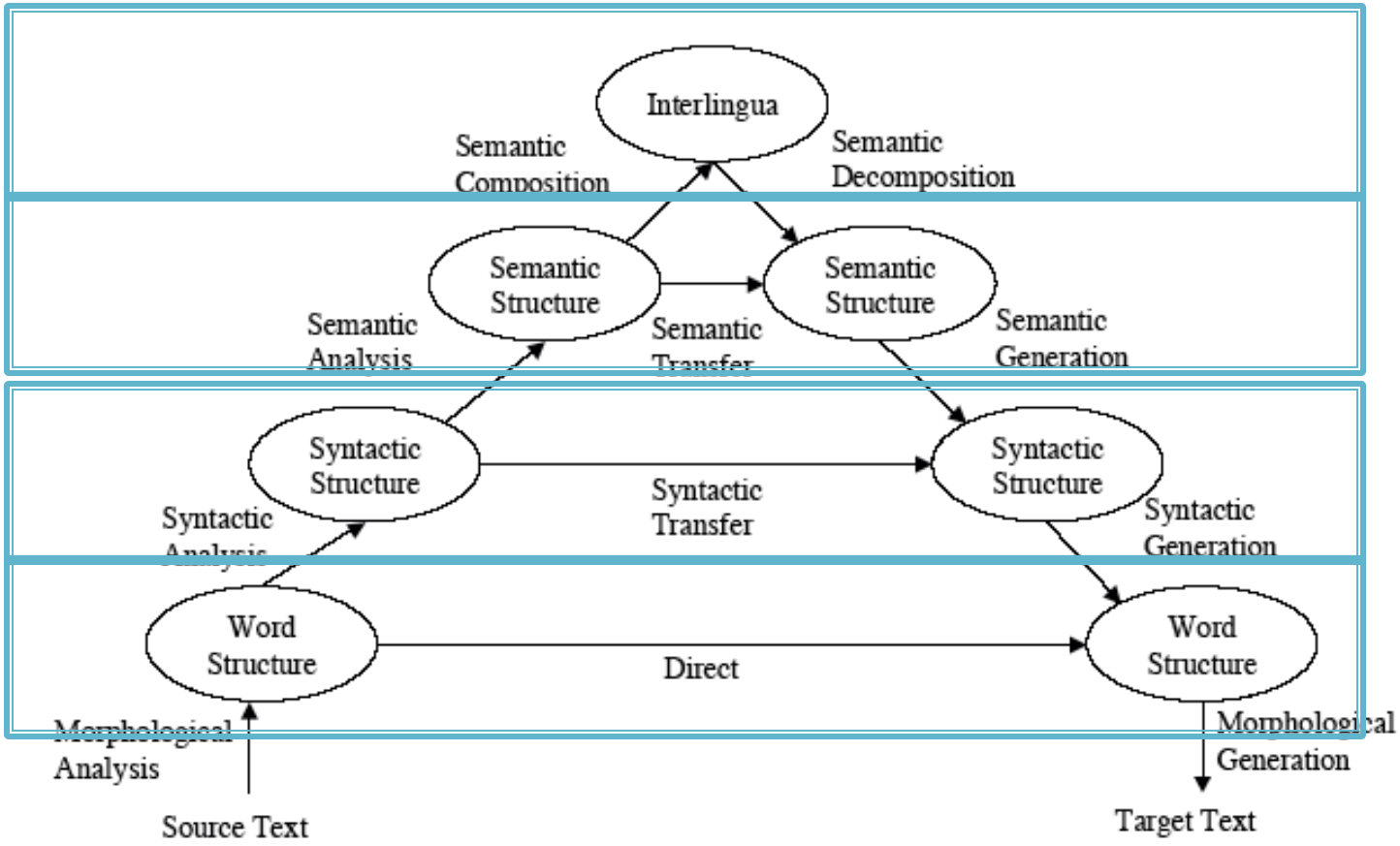


Figure 1: The Vauquois Triangle for MT

TA hoje

MACHINE TRANSLATION

Objetivos

- Substituir o desempenho humano
  - FAHQMT – Fully-Automatic High-Quality Machine Translation
- Estender o desempenho humano
  - HAMT – Human-Aided Machine Translation
- Acelerar o desempenho humano
  - MAHT – Machine-Aided Human Translation

COMPUTER-AIDED TRANSLATION

TA hoje

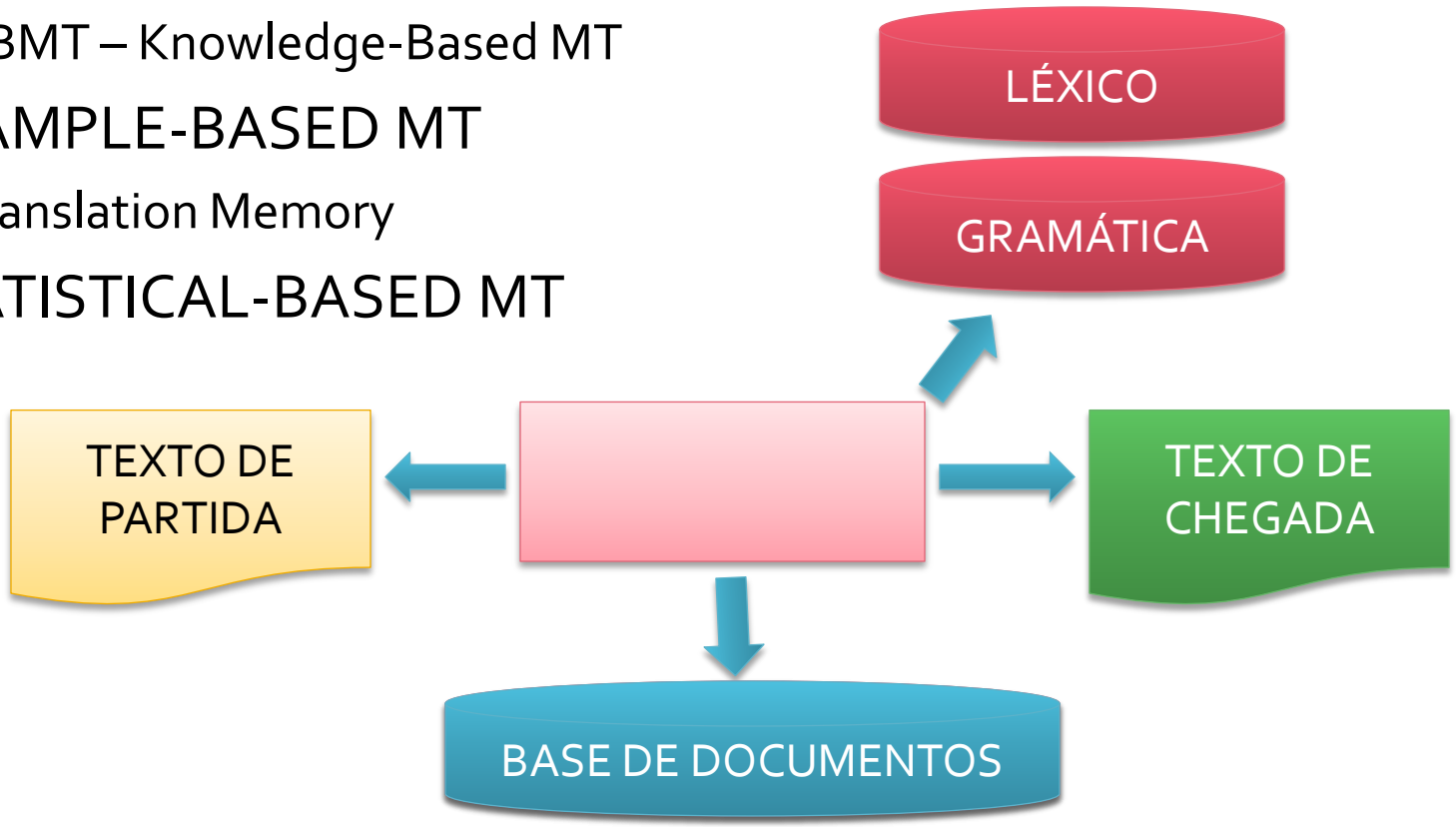
Computer-Aided Translation

- Rough Translation (pre-translation)
- Pre-Editing
  - Sublanguage
    - universo do discurso
    - tipo de texto
  - Controlled languages
  - Tagging
- Post-Editing
  - raw translation

TA hoje

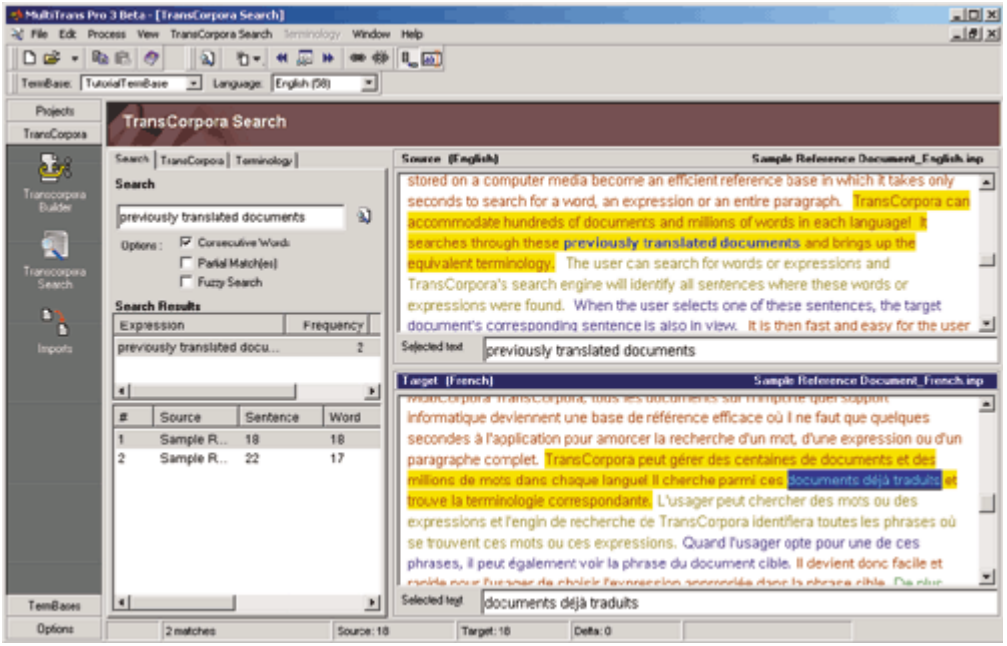
Paradigmas

- RULE-BASED MT
  - LBMT – Language-Based MT
  - KBMT – Knowledge-Based MT
- EXAMPLE-BASED MT
  - Translation Memory
- STATISTICAL-BASED MT



Statistical MT (SMT)

EuroMatrix
Statistical and Hybrid Machine Translation
Between All European Languages



The spirit is willing but the flesh is weak

- O espírito é disposto mas a carne é fraca
- The spirit is made use but the meat is weak

- O espírito está pronto, mas a carne é fraca
- The spirit is ready but the flesh is weak

0.3954

Arabic <-> English

0.4281

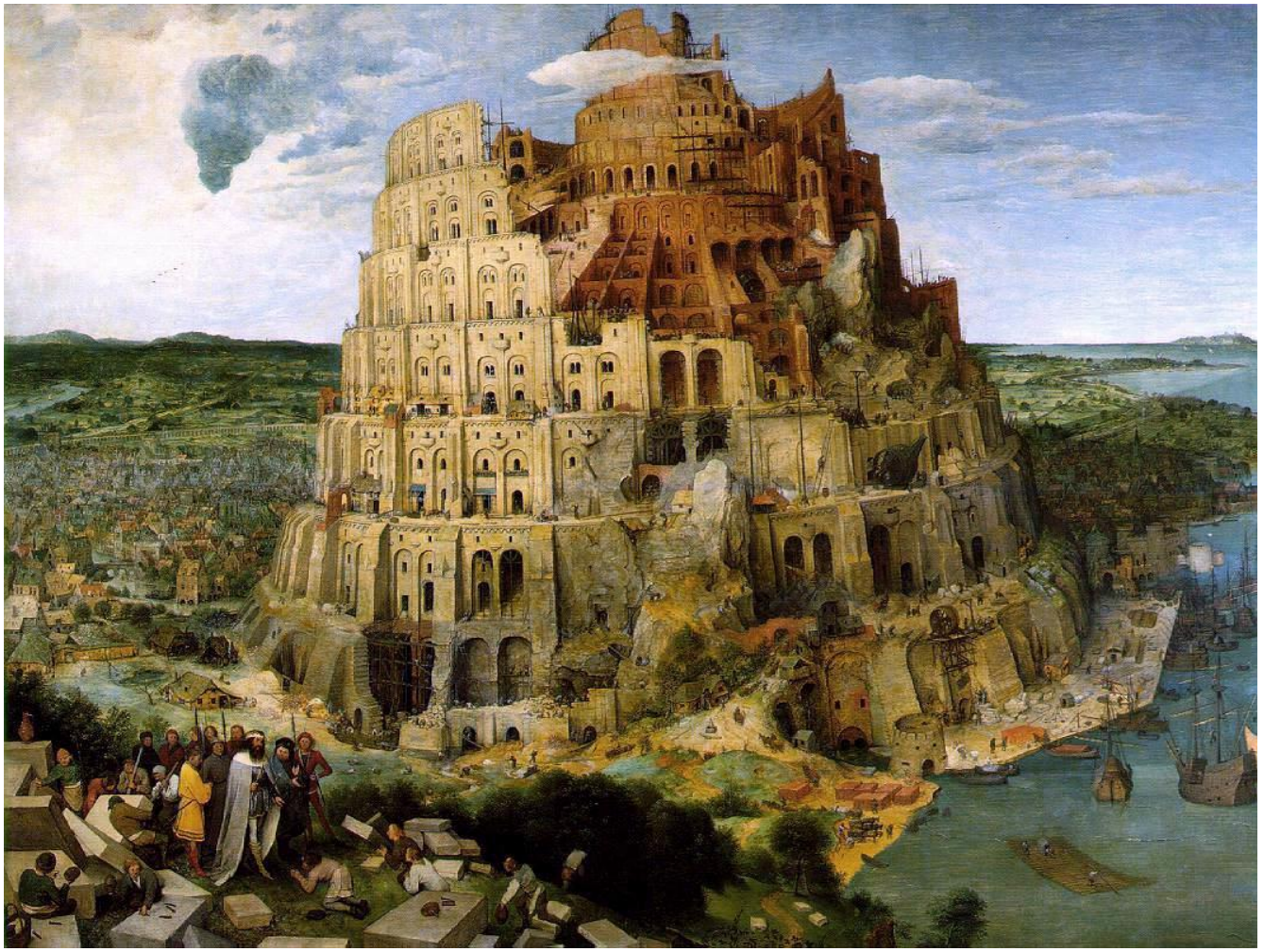
SYSTRAN

GOOGLE TRANSLATOR

Ronaldo Martins
UNDL Foundation

Estudo de caso: UNL

The Idea of an Universal Language

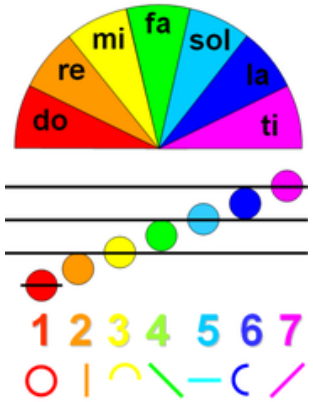


The Idea of an Universal Language

Ars Magna (1275)
Ramon Llull

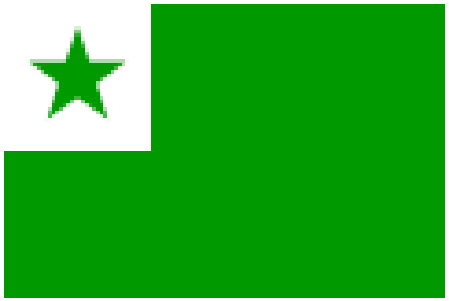
De Arte Combinatoria (1666)
Gottfried Leibniz

An Essay towards a Real Character and a Philosophical Language (1668)
John Wilkins



Solresol (1827)
François Sudre

Volapük (1879)
Johann Martin Schleyer



Esperanto (1887)
L. L. Zamenhof

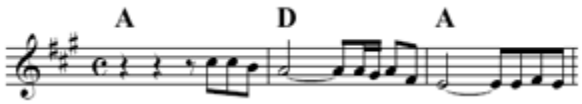
Interlingua (1937-1951)
International Auxiliary Language Association

The Idea of an Universal Language

LANGUAGE INDEPENDENCE

$[Co(NH_3)_6]^{3+}$

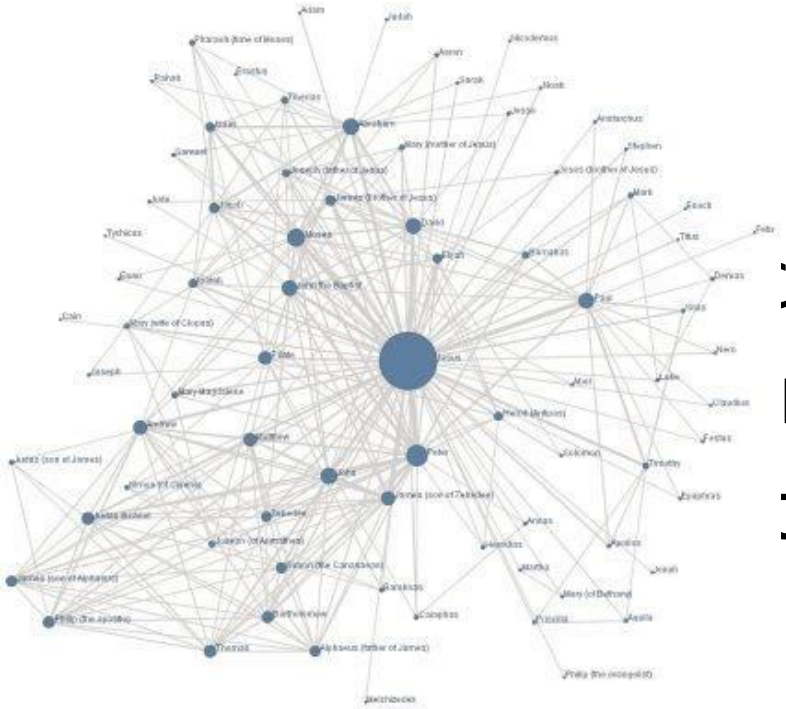
$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$



DOMAIN INDEPENDENCE?

Foundations (I)

Informa natural
langua :an be
represe nguage-
indepe :ructure.



Foundations (II)

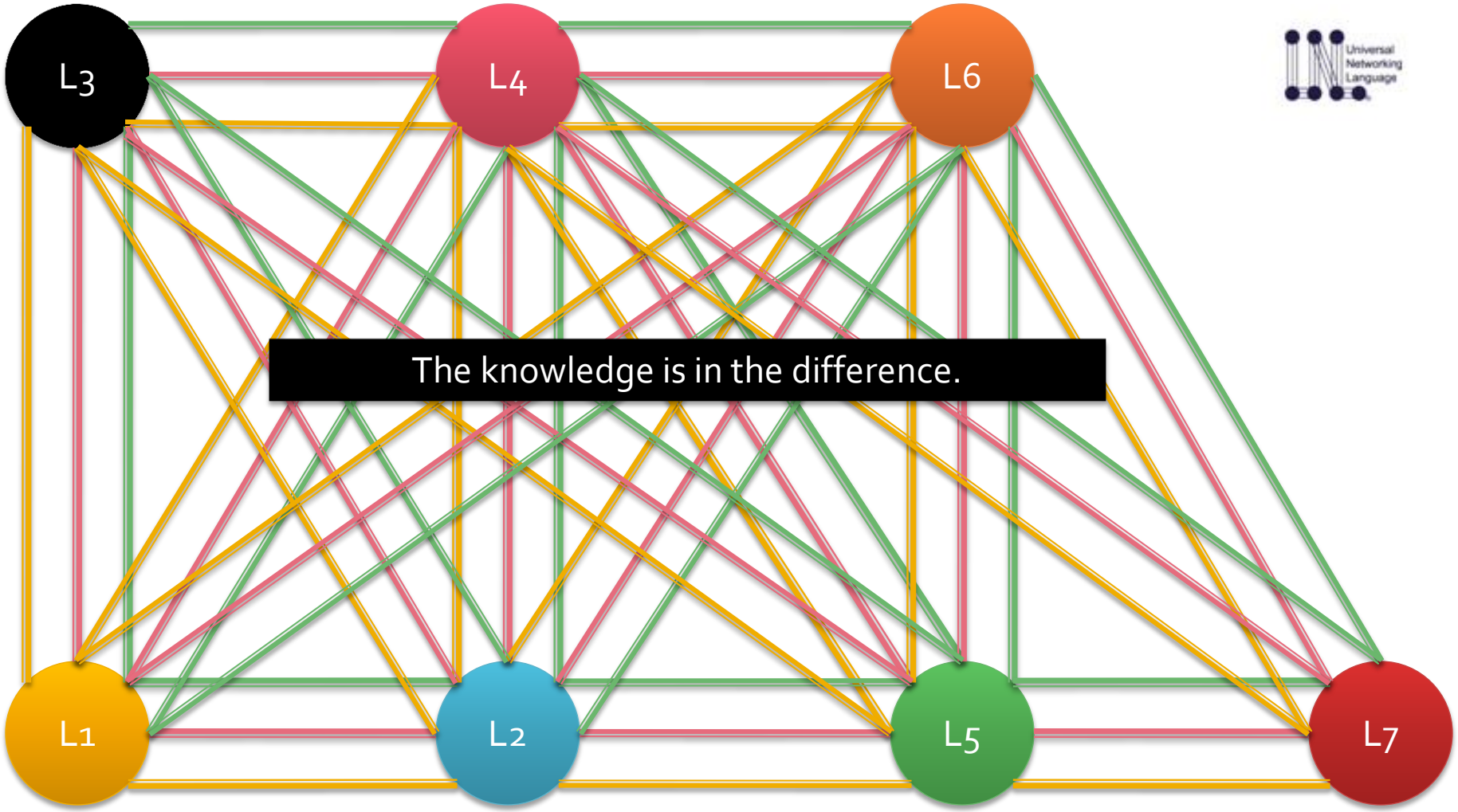
- Arguments
  - Universal Words (UWs)
- One-place predicates
  - Attributes
- Two-place predicates
  - Relations

CONCEPTS

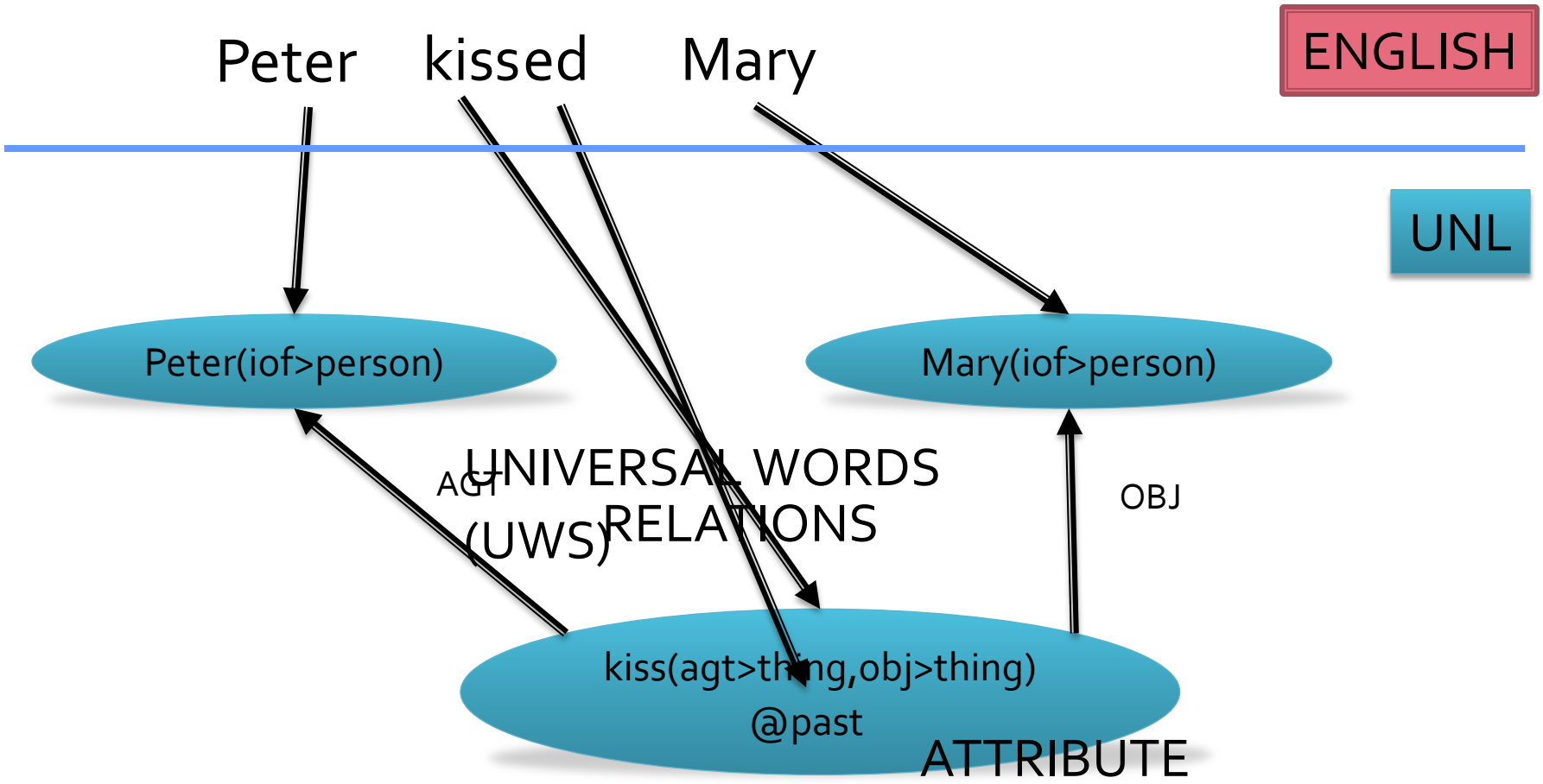
CONCEPT MODIFIERS

RELATIONS BETWEEN CONCEPTS

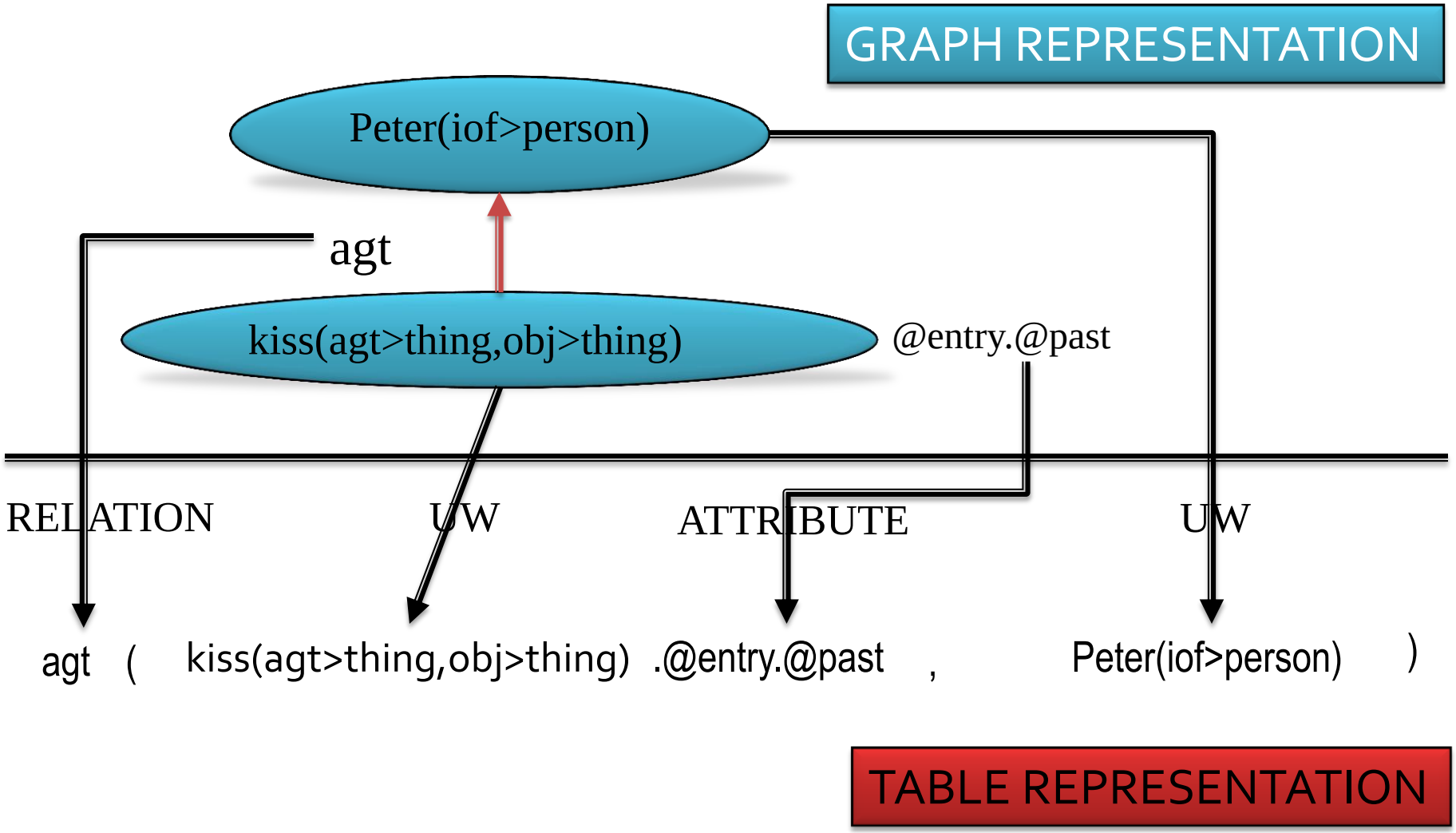
The Universal NETWORKING Language



NL-to-UNL (Enconversion)



Syntax of UNL



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)

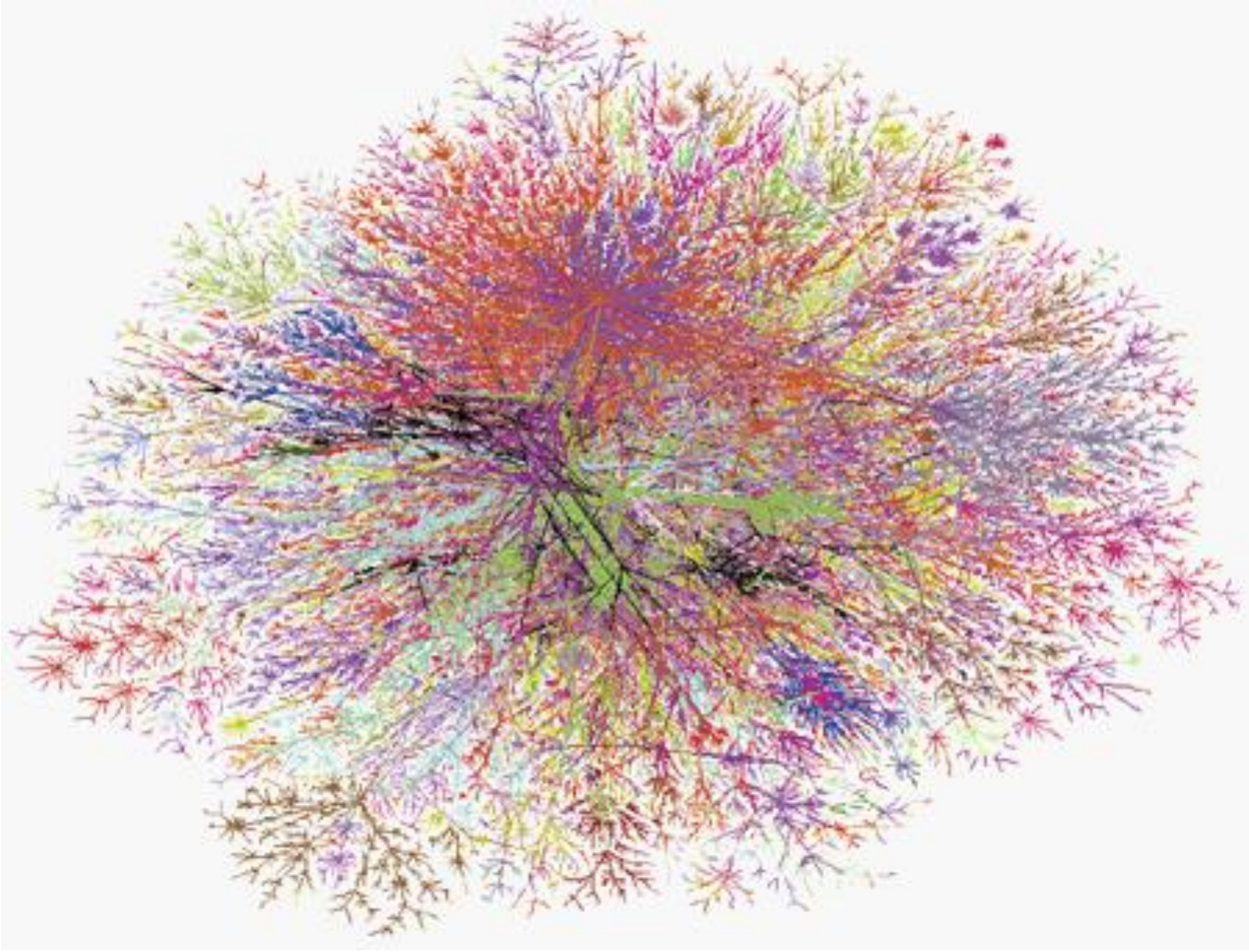
<unl>
agt(kiss(agt>thing,obj>thing).@entry.@past, Peter(iof>person))
obj(kiss(agt>thing,obj>thing).@entry.@past, Mary(iof>person))
</unl>

Peter kissed Mary.

Pierre a embrassé Marie.

Pedro besó a María.

The UNL Web



Ronaldo Martins
UNDL Foundation

The structure of UNL

Structure of UNL

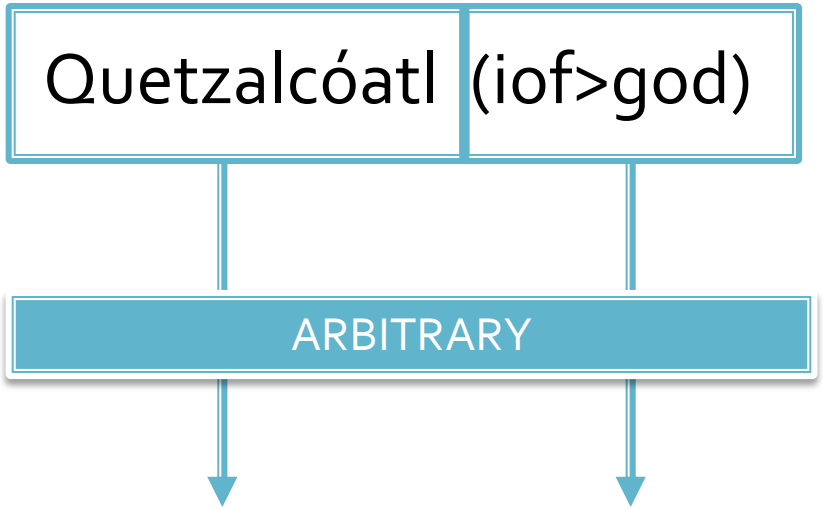
- Arguments
  - Universal Words (UWs)
- One-place predicates
  - Attributes
- Two-place predicates
  - Relations

CONCEPTS

CONCEPT MODIFIERS

RELATIONS BETWEEN CONCEPTS

Universal Words



IDENTIFICATION

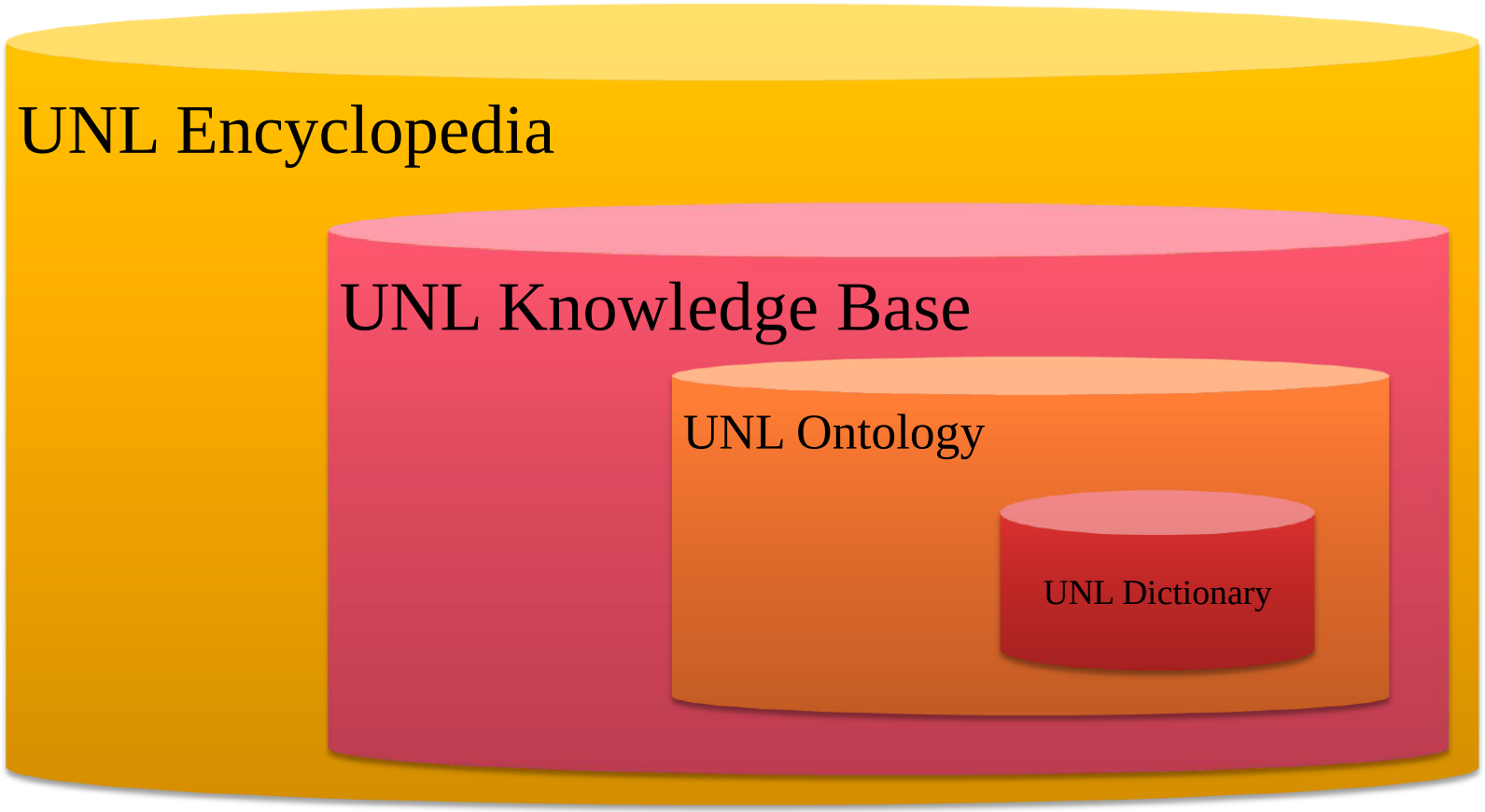
DISAMBIGUATION

READABILITY

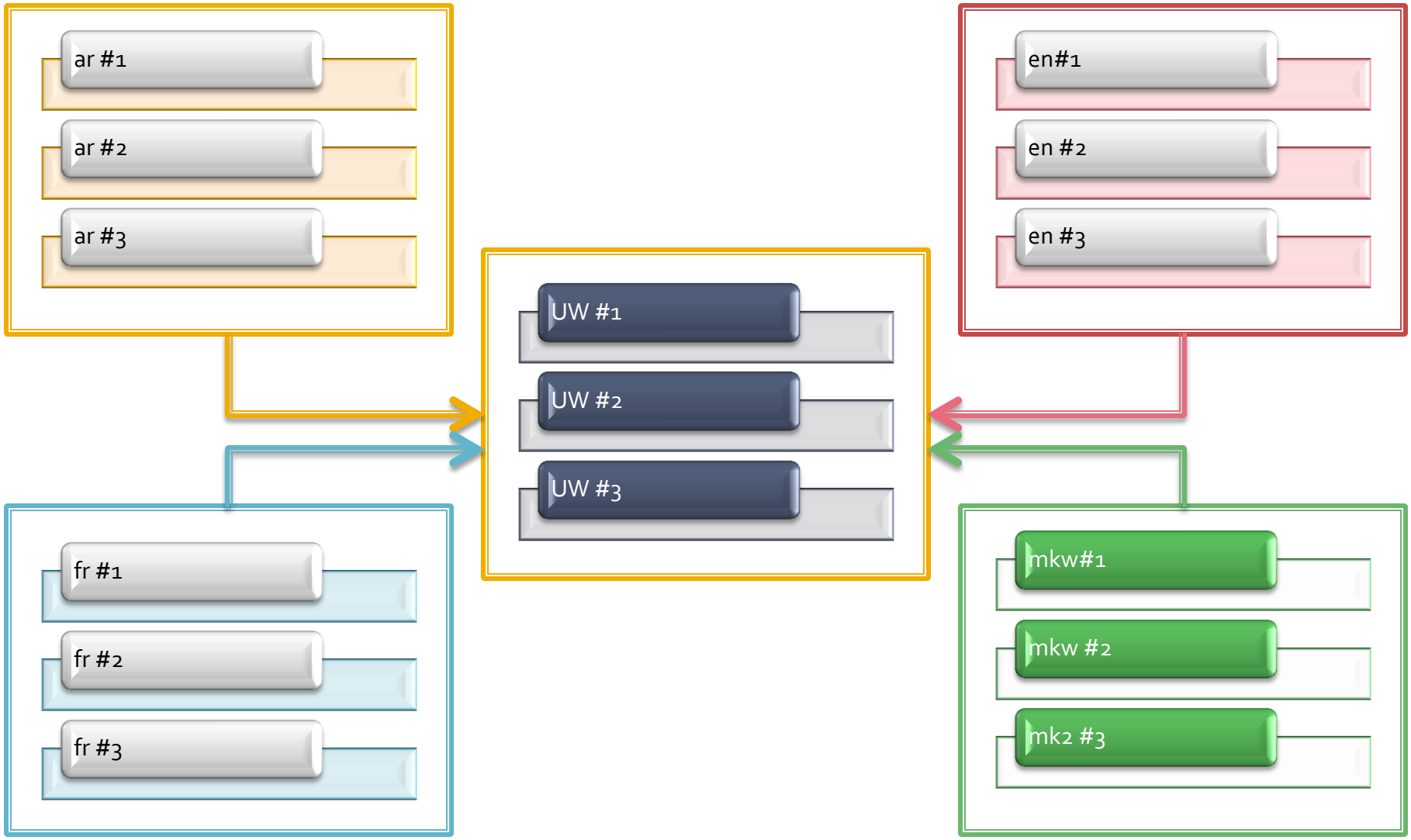
Examples

'Universal Word'
'uw'
'nominal concept'
'thing'
'abstract thing'
'activity(icl>abstract thing)'
'broadcasting(icl>activity)'
'tale(icl>information)'
'above(icl>direction)'
'month(icl>date)'
'April'
'do(agt>thing)'
'dance(agt>person)'
'bark(agt>dog)'
'explain(icl>express(agt>thing,gol>person,obj>thing))'

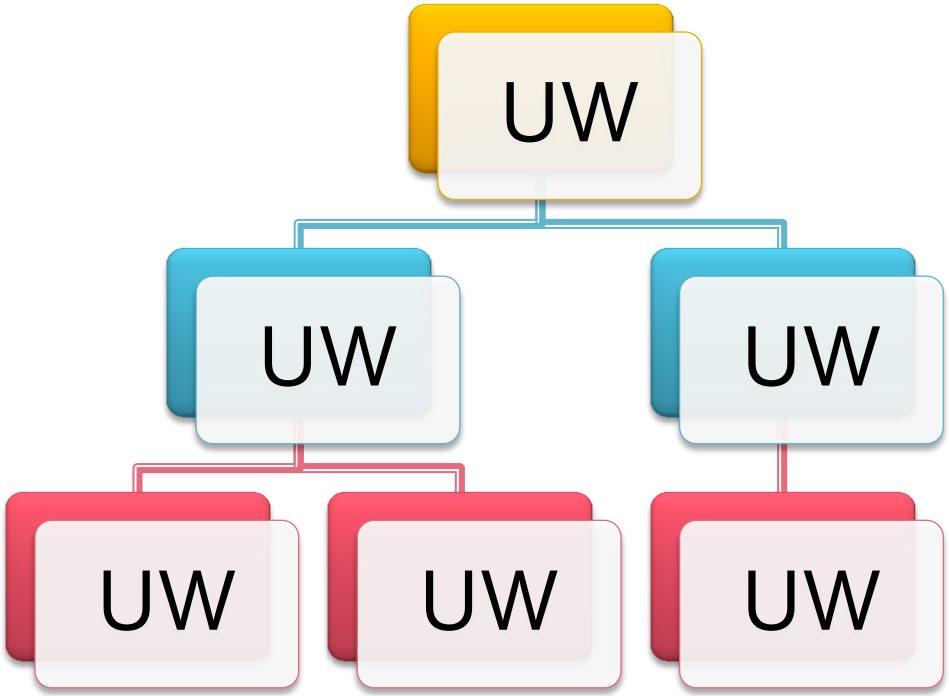
Lexical Databases



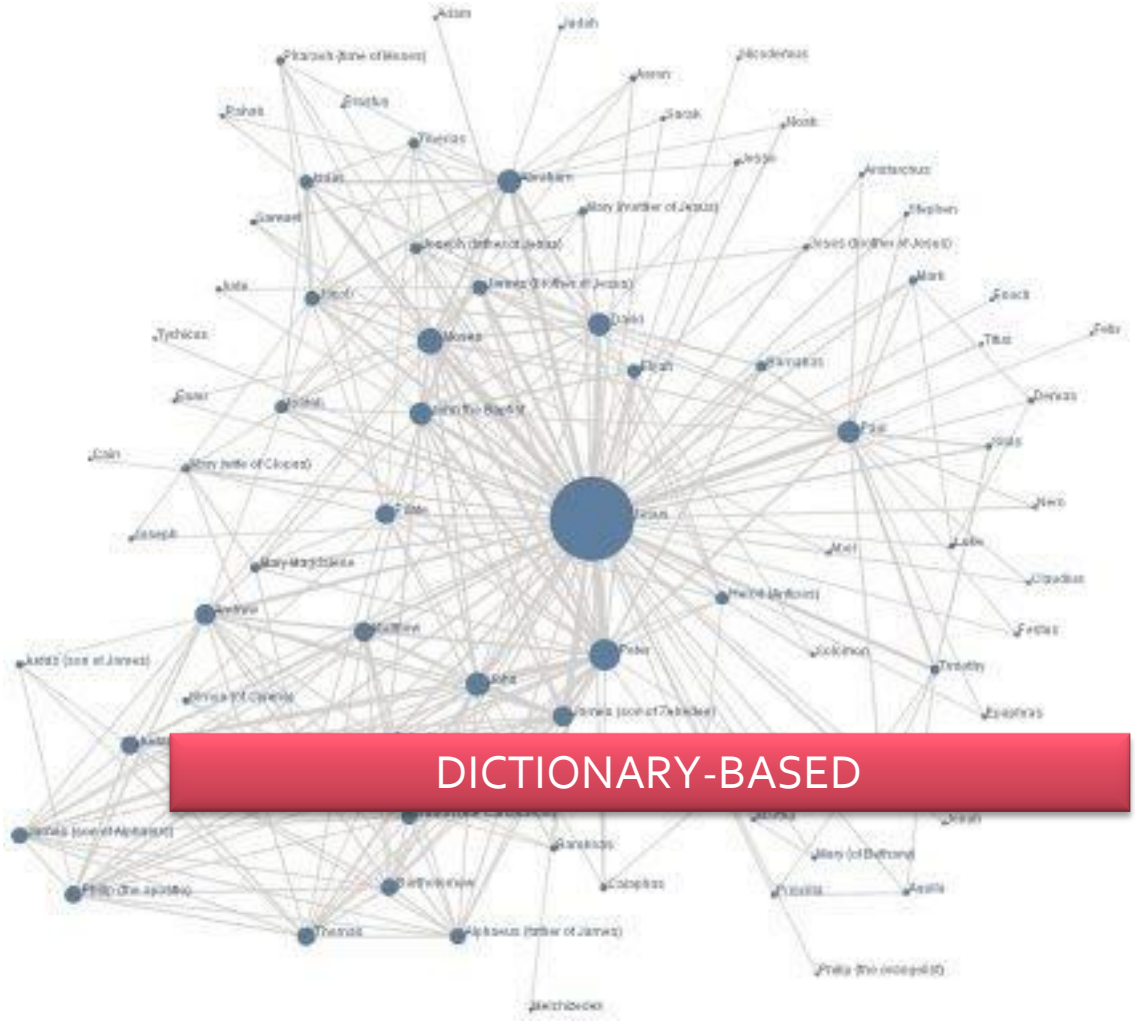
UNL Dictionary



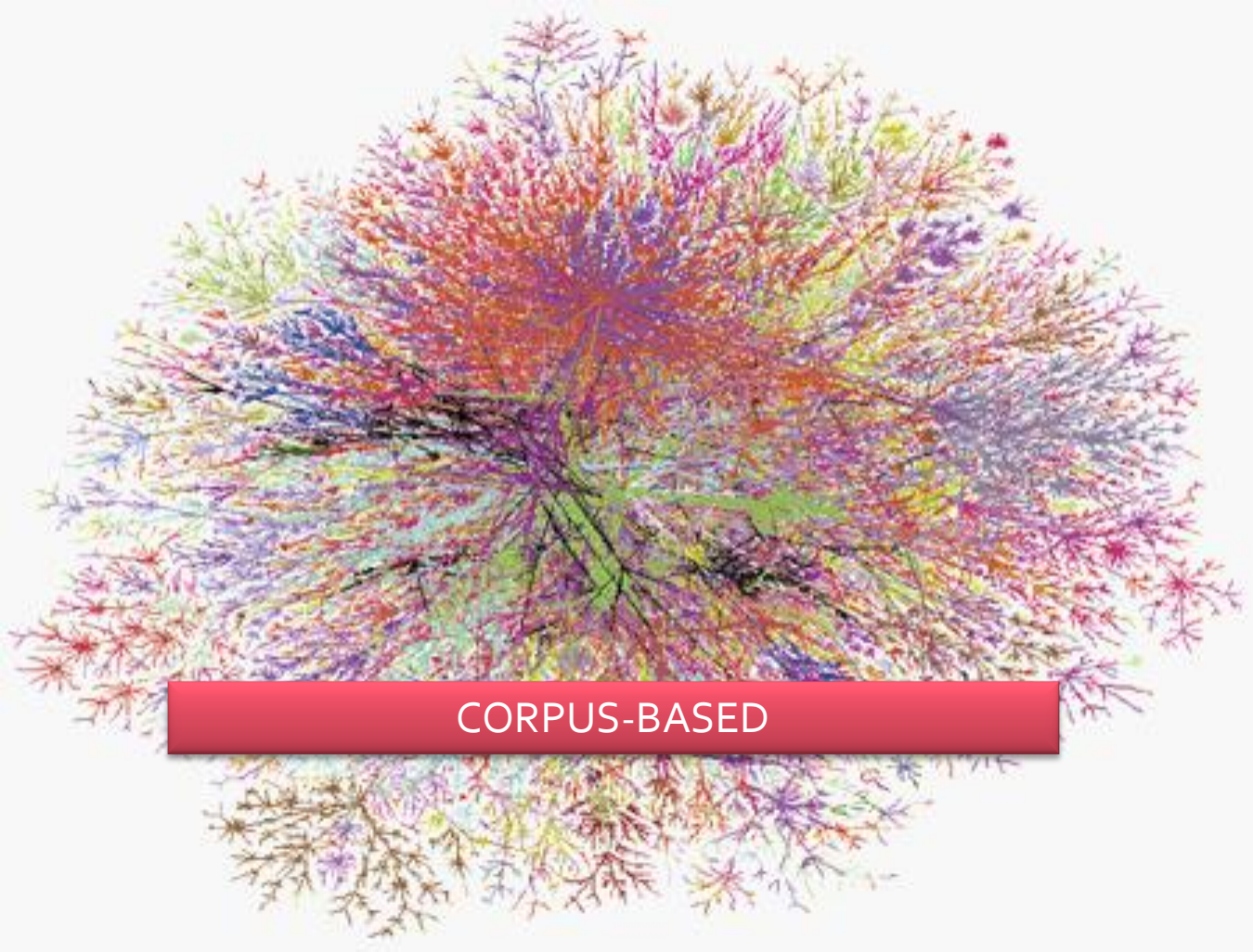
UNL Ontology



UNL KB



UNL Encyclopedia



Attributes

- Tense
- Aspect
- Speaker's view of reference
- Speaker's emphasis, focus, topic, etc.
- Convention
- Speaker's attitudes
- Speaker's feelings and viewpoints

Tense

@past	happened in the past	ex) It was snowing yesterday
@present	happening at present	ex) It is raining hard.
@future	will happen in future	ex) He will arrive tomorrow

Aspect

@begin	beginning of an event or a state	Ex) It began to work again.
@complete	finishing/completion of a (whole) event.	Ex) I've looked through the script. look.@entry.@complete
@continue	continuation of an event	Ex) He went on talking. talk.@continue.@past
@custom	customary or repetitious action	Ex) I used to visit [I would often go] there when I was a boy. visit.@custom.@past
@end	end/termination of an event or a state	Ex) I have done it. do.@end.@present
@experience	experience	Ex) Have you ever visited Japan? visit.@experience.@interrogation Ex) I have been there. visit.@exterience
@progress	an event is in progress	Ex) I am working now. work.@progress.@present
@repeat	repetition of an event	Ex) It is so windy that the tree branches are knocking against the roof. knock.@entry.@present.@repeat
@state	final state or the existence of the object on which an action has been taken	Ex) It is broken. break.@state

Reference

@generic	generic concept	Ex) The dog is a faithful animal.
@def	already referred	Ex) the book you lost
@indef	non-specific class	Ex) There is a book on the desk.
@not	complement set	Ex) Don't be late!
@ordinal	ordinal number	Ex) the 2nd door

Focus

@contrast	Contrasted UW	For instance, "but" in the examples below is used to introduce a word or phrase that contrasts with what was said before. Ex) It wasn't the red one but the blue one. Ex) He's poor but happy.
@emphasis	Emphasized UW	Ex) I do like it.
@entry	Entry or main UW of a sentence or a scope	Ex) He promised (entry of the sentence) that he would come (entry of the scope)
@qfocus	Focused UW of a question	Ex) Are you painting the bathroom blue? To this question, the answer will be "No, I'm painting the LIVING-ROOM blue"
@title	Title	Ex) UNL Summer School
@topic	Topic	Ex) He(@topic) was killed by her. Ex) The girl(@topic) was given a doll. Ex) This doll(@topic) was given to the girl.

Attitude

@affirmative	affirmation	Ex) His name is Peter.
@confirmation	confirmation	Ex) You won't say that, will you? Ex) It's red, isn't it? Ex) Then you won't come, right?
@exclamation	feeling of exclamation	Ex) kirei na! ("How beautiful (it is)!" in Japanese) Ex) Oh, look out!
@imperative	imperative	Ex) Get up! Ex) You will please leave the room.
@interrogative	interrogation	Ex) Who is it?
@invitation	inducement to do something	Ex) Will / Won't you have some tea? Ex) Let's go, shall we?
@polite	polite feeling. Puts emphasis on a way of talking.	Ex) Could you (please)... Ex) If you could … I would …
@request	request	Ex) Please don't forget…
@respect	respectful feeling. In many cases, some special words are used.	Ex) o taku ("(your) house" in Japanese) Ex) Good morning, sir.
@vocative	vocative	Ex) Boys, be ambitious!

Point of View

@ability	Ability, capability of doing something	Ex) The child can 't walk yet. Ex) He can speak English but he can't write it very well.
@admire	Admiring feeling of the speaker about something	Ex) What a wonderful world!
@conclusion	Logical conclusion due to a certain condition	Ex) He is her husband; she is his wife.
@consequence	Logical consequence	Ex) He was angry, wherefore I left him alone.
@blame	Blameful feeling of the speaker about something	Ex) A sailor, and afraid of the sea!
@dissent	Dissenting feeling of the speaker about something	Ex) But that's not true.
@grant	To give/get consent/permission to do something	Ex) Can I smoke in here? Ex) You may borrow my car if you like.
@grant-not	Not to give consent to do something	Ex) You {mustn't/are not allowed to/may not} borrow my car.
@although	Something follows against [contrary to] or beyond expectation	Ex) Although he didn't speak, I felt a certain warmth in his manner.

Relations

- Set theory
- Thematic theory
- Others

Set theory

CONCEPTUAL RELATION	RL	DEFINITION	EXAMPLE (UNL)	EXAMPLE (ENGLISH)
disjunction	or	$X \vee Y$	or(Peter, Mary)	Mary or Peter
conjunction	and	$X \wedge Y$	and(Peter, Mary)	Mary and Peter
condition	con	$X \rightarrow Y$	con(get wet, rain)	If it rains, it will get wet.
description	cnt	$X \equiv Y$	cnt(language, UNL)	A language: UNL
intersection	Int	$X \cap Y$	int(tableware(icl>tool), cookware(icl>tool))	an intersection of tableware and cookware
Hyponymy	icl	$X \subset Y$	icl(table(icl>furniture), furniture(icl>functional thing))	table
Synonymy	equ	$X \equiv Y$	equ(UNL(equ>Universal Networking Language), Universal Networking Language)	UNL
Instance	iof	$X \in Y$	iof(New York(iof>city), city(icl>region))	New York

Thematic Theory

Actors

RELATION	RL	EXAMPLE (UNL)	EXAMPLE (ENGLISH)
Agent	agt	agt(kill, John)	John killed (Peter)
Co-agent	cag	cag(kill, Mary)	(John) killed (Peter) with (the help of) Mary
Beneficiary	ben	ben(kill, Mary)	(John) killed (Peter) for Mary
Instrument	ins	ins(kill, knife)	(John) killed (Peter) with a knife
Object	obj	obj(kill, Peter)	(John) killed Peter
Co-object	cob	cob(kill, Mary)	(John) killed (Peter) with Mary
Partner	ptn	ptn(fight, Peter)	(John) fought Peter

Thematic Theory

Space

CONCEPTUAL RELATION	RL	EXAMPLE (UNL)	EXAMPLE (ENGLISH)
Origin	frm	frm(train, NY)	Train from New York
Place	plc	plc(kill, church)	(John) killed (Peter) in the church
Initial place	plf	plf(go, NY)	To go from New York
Final place	plt	plt(go, NY)	To go to New York
Destination	to	to(train, NY)	Train to New York
Intermediary place	via	via(go, NY)	To go through New York
Affected place	opl	opl(hit, back)	(John) hit (Peter) in the back
Scene	scn	scn(kill, dream)	(John) killed (Peter) in dreams

Thematic Theory

Time

CONCEPTUAL RELATION	RL	EXAMPLE (UNL)	EXAMPLE (ENGLISH)
Duration	dur	dur(kill, mess)	(John) killed (Peter) during the mess.
Time	tim	tim(kill, night)	(John) killed (Peter) at night.
Initial time	tmf	tmf(walk, morning)	(John) have been walking since morning.
Final time	Tmt	tmt(walk, noon)	(John) will walk until noon.
co-occurrence	coo	coo(kill, sing)	(John) killed (Peter) while (Mary) was singing.
Sequence	seq	seq(kill, think)	(John) thought (about Mary) before killing (Peter)

Others

Nominal relations

CONCEPTUAL RELATION	RL	EXAMPLE (UNL)	EXAMPLE (ENGLISH)
Attribute	aoj	aoj(jealous, John)	John (is) jealous.
co-attribute	cao	cao(jealous, Peter)	(John killed) Peter, the jealous one.
comparison	bas	bas(like, professional)	(John killed Peter) as a professional.
Modifier	mod	mod(story, whole)	The whole story
Name	Nam	nam(city, NY)	City of New York
Part-of	Pof	pof(head, Peter)	The head of Peter
Possessor	Pos	pos(knife, John)	John`s knife
Quantity	Qua	qua(stab, 3)	Three stabs
Proportion	Per	per(2, day)	Two per day

Ronaldo Martins
UNDL Foundation

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

ERA UMA VEZ um jovem casal que estava muito feliz. Ela estava grávida, e eles esperavam com grande ansiedade o filho que iria nascer.

Transcorridos os nove meses de gravidez, ele nasceu. Ela deu à luz um lindo computador! Que felicidade ter um computador como filho! Era o filho que desejavam ter! Por isso eles haviam rezado muito durante toda a gravidez, chegando mesmo a fazer promessas.

O batizado foi uma festança. Deram-lhe o nome de Memorioso, porque julgavam que uma memória perfeita é o essencial para uma boa educação. Educação é memorização. Crianças com memória perfeita vão bem na escola e não têm problemas para passar no vestibular.

E foi isso mesmo que aconteceu. Memorioso memorizava tudo que os professores ensinavam. Mas tudo mesmo. E não reclamava. Seus companheiros reclamavam, diziam que aquelas coisas que lhes eram ensinadas não faziam sentido. Suas inteligências recusavam-se a aprender. Tiravam notas ruins. Ficavam de recuperação.

Isso não acontecia com Memorioso. Ele memorizava com a mesma facilidade a maneira de extrair raiz quadrada, reações químicas, fórmulas de física, acidentes geográficos, populações de países longínquos, datas de eventos históricos, nomes de reis, imperadores, revolucionários, santos, escritores, descobridores, cientistas, palavras novas, regras de gramática, livros inteiros, línguas estrangeiras. Sabia de cor todas as informações sobre o mundo cultural.

A memória de Memorioso era igual à do personagem do Jorge Luis Borges de nome Funes. Só tirava dez, o que era motivo de grande orgulho para os seus pais. E os outros casais, pais e mães dos colegas de Memorioso, morriam de inveja. Quando filhos chegavam em casa trazendo boletins com notas em vermelho eles gritavam: "por que você não é como o Memorioso?"

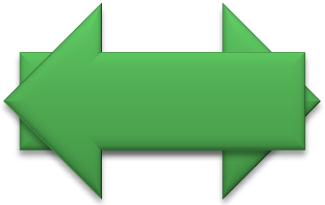
Memorioso foi o primeiro no vestibular. O cursinho que ele freqüentara publicou sua fotografia em outdoors. Apareceu na televisão como exemplo a ser seguido por todos os jovens. Na universidade, foi a mesma coisa. Só tirava dez. Chegou, finalmente, o dia tão esperado: a formatura. Memorioso foi o grande herói, elogiado pelos professores. Ganhou medalhas e mesmo uma bolsa para doutoramento no MIT.

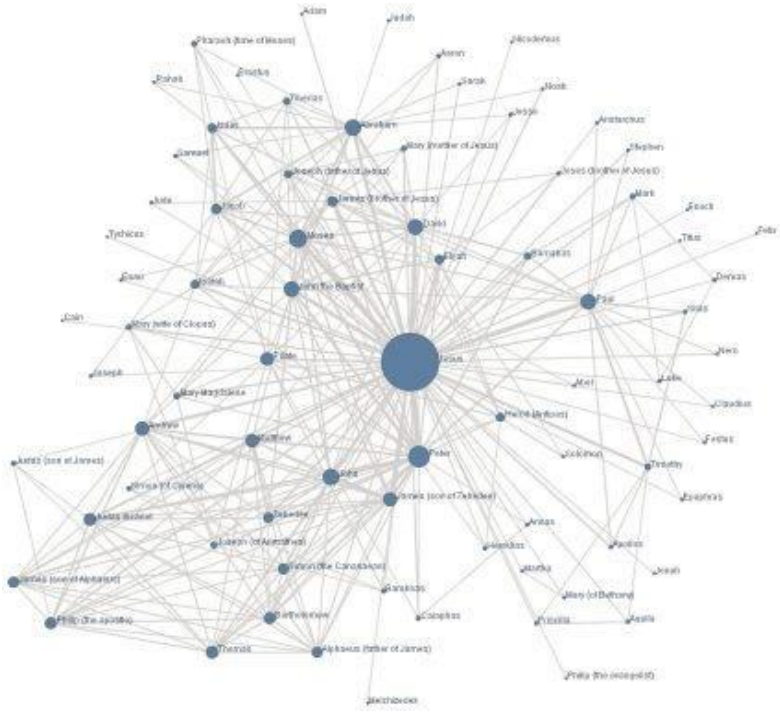
Depois da cerimônia acadêmica foi a festa. E estavam todos felizes no jantar quando uma moça se aproximou de Memorioso e se apresentou: "Sou repórter. Posso lhe fazer uma pergunta?" "Pode fazer", disse Memorioso confiante. Sua memória continha todas as respostas.

Aí ela falou: "De tudo o que você memorizou qual foi aquilo que você mais amou? Que mais prazer lhe deu?"

Memorioso ficou mudo. Os circuitos de sua memória funcionavam com a velocidade da luz procurando a resposta. Mas aquilo não lhe fora ensinado. Seu rosto ficou vermelho. Começou a suar. Sua temperatura subiu. E, de repente, seus olhos ficaram muito abertos, parados, e se ouviu um chiado estranho dentro de sua cabeça, enquanto fumaça saía por suas orelhas. Memorioso primeiro travou. Deixou de responder a estímulos. Depois apagou, entrou em coma. Levado às pressas para o hospital de computadores, verificaram que seu disco rígido estava irreparavelmente danificado.

Há perguntas para as quais a memória não tem respostas. É que tais respostas não se encontram na memória. Encontram-se no coração, onde mora a emoção...



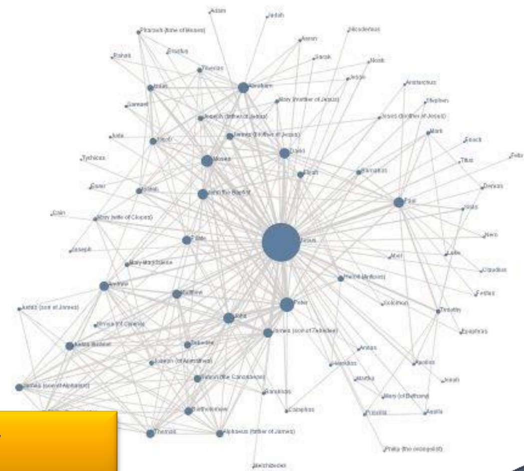


The UNL System


C2 cotidiano FOLHA DE S.PAULO

RUBEM ALVES

O aluno perfeito




C2 cotidiano FOLHA DE S.PAULO

RUBEM ALVES

O aluno perfeito

SHALLOW

DEEP

UNLization

TEXT DRIVEN

SENTENCE DRIVEN

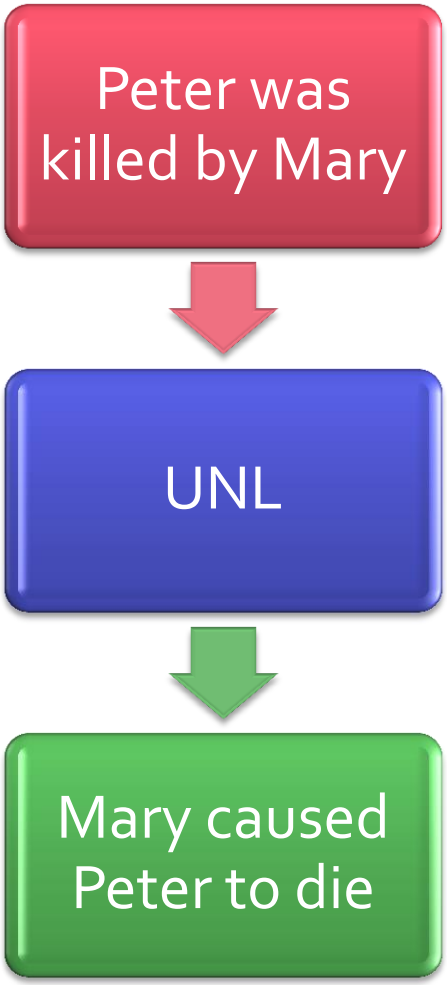
WORD DRIVEN

To provide a UNL version of language-restricted data

NLization

To provide resources to generate natural language data out of UNL

Is this translation?

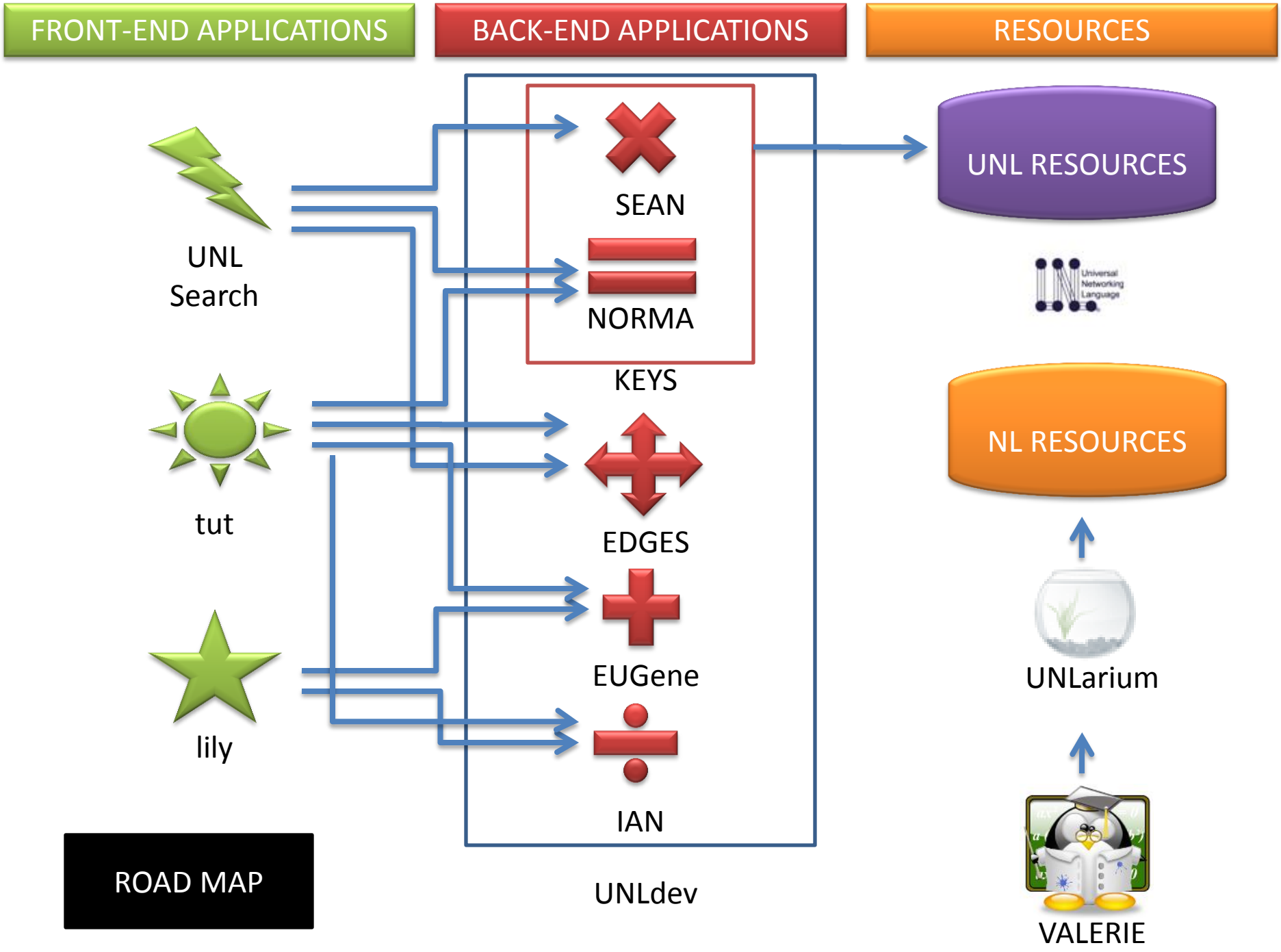


What for?

Roles of UNL

representing
organizing
retrieving
extracting
inferring
generating

monolingual
multilingual
information


FRONT-END APPLICATIONS
BACK-END APPLICATIONS
RESOURCES
UNL Search
tut
lily
SEAN
NORMA
KEYS
EDGES
EUGene
IAN
UNLdev
UNL RESOURCES
Universal Networking Language
NL RESOURCES
UNLarium
VALERIE
ROAD MAP

Ronaldo Martins
UNDL Foundation

Faça você mesmo

Objetivo

- Traduzir o texto abaixo de inglês para português

Alice's Adventures in Wonderland

I - Down the rabbit-hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do. Once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice, "without pictures or conversations?"

Estado da arte

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do. Once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice, "without pictures or conversations?"

Alice estava começando a ficar muito cansado de ficar sentado por sua irmã no banco, e de não ter nada para fazer. Uma ou duas vezes ela espiou o livro em sua irmã estava lendo, mas não tinha imagens ou conversas em que, "e o que é o uso de um livro", pensou Alice, "sem figuras nem diálogos?"

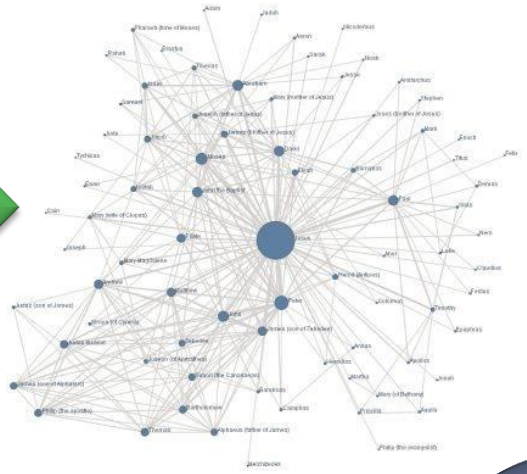
Alice estava começando a obter muito cansado do assento por sua irmã no banco, e de não não ter nada fazer. Uma vez que ou duas vezes peeped no livro que sua irmã lia, mas não teve nenhuma retrato ou conversação nele, "e o que é o uso de um livro," pensamento Alice, "sem retratos ou conversações?"

GOOGLE

SYSTRAN

Metodologia

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do. Once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice, "without pictures or conversations?"



Alice estava começando a ficar muito cansada de estar sentada ao lado de sua irmã e não ter nada para fazer: uma vez ou duas ela dava uma olhadinha no livro que a irmã lia, mas não havia figuras ou diálogos nele e "para que serve um livro", pensou Alice, "sem figuras nem diálogos?".

UNLization

- UNL Editor
- IAN
- SEAN

NLization

- EUGENE

UNLization

- UNL Editor
  - Dicionário EN-UNL
- IAN ou SEAN
  - Dicionário EN-UNL
  - Gramática EN-UNL
  - Knowledge Base*
  - Memória de Tradução*

Resultado

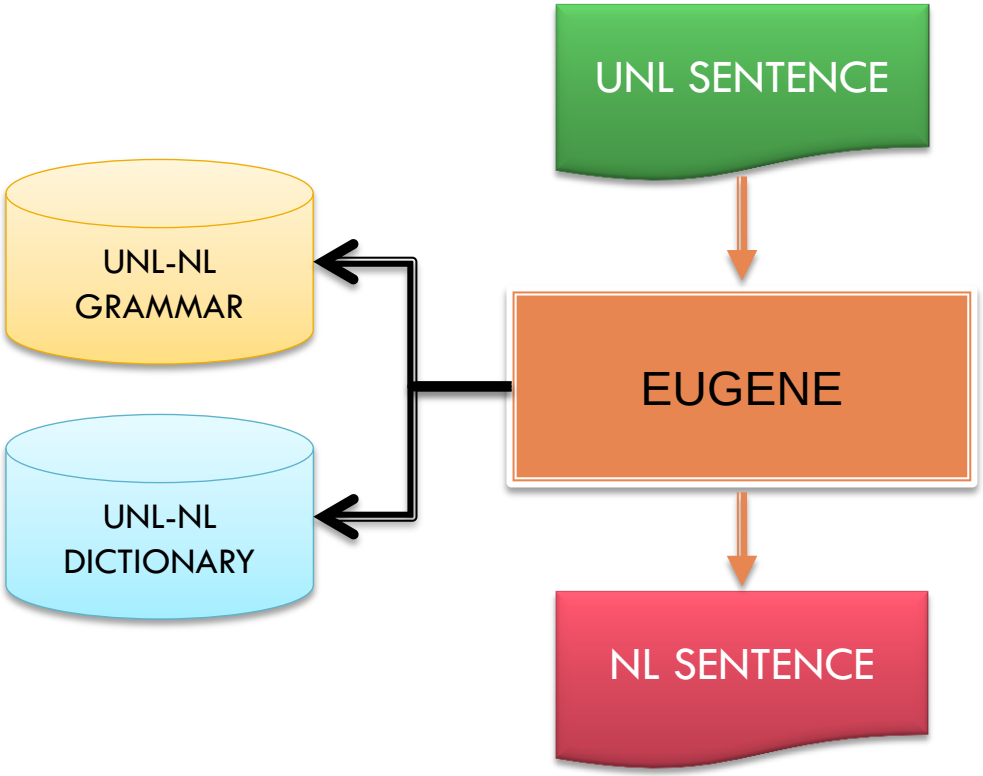
[S:1]
{org}
ALICE'S ADVENTURES IN WONDERLAND
{/org}
{unl}
mod(100796315:71.@pl.@entry, Alice:29)
plc(100796315:71.@pl.@entry, 105632175:19)
{/unl}
[/S]

[S:2]
{org}
DOWN THE RABBIT-HOLE
{/org}
{unl}
pos(109304465:78.@def.@down, 102324045:91)
plc(00:75.@entry, 109304465:78.@def.@down)
{/unl}
[/S]

[S:3]
{org}
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do.
{/org}
{unl}
aoj(302431728:22.@inceptive.@inchoative.@past.@plus.@progressive, Alice:79)
obj(302431728:22.@inceptive.@inchoative.@past.@plus.@progressive, 202374099:17)
obj(302431728:22.@inceptive.@inchoative.@past.@plus.@progressive, 202203362:99)
plc(202203362:99, 00:31)
plc(202374099:17, 110602985:75.@by)
pur(202374099:17, 109213565:23.@def)
pur(00:31, 202560585:39)
and(202374099:17, 202203362:99)
mod(110602985:75.@by, Alice:79)
{/unl}
[/S]

[S:4]
{org}
Once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice, "without pictures or conversations?"
{/org}
{unl}
or(400065294:29, 400118869:66)
man(202165146:45.@past.@anterior.@entry, 400118869:66)
agt(202165146:45.@past.@anterior.@entry, 400065294:29)
agt(202165146:45.@past.@anterior.@entry, Alice:23)
obj(202165146:45.@past.@anterior.@entry, 106410904:34.@def)
mod(110602985:38, Alice:23)
agt(200625119:48.@past.@progressive.@relative, 110602985:38)
obj(200625119:48.@past.@progressive.@relative, 106410904:34.@def)
aoj(202203362:24.@past.@contrast, 106410904:34.@def)
obj(202203362:24.@past.@contrast, 103931044:12.@no)
or(202203362:24.@past.@contrast, 107133701:95.@no)
or(107133701:95.@no, 103931044:12.@no)
aoj(200689344:19.@past.@and, Alice:23)
obj(200689344:19.@past.@and, :01)
or(107133701:65.@without.@generic, 103931044:10.@without.@generic)
mod:01(100947128:49.@def.@interrogative, 106410904:82.@indef)
mod:01(106410904:82.@indef, 107133701:65.@without.@generic)
mod:01(106410904:82.@indef, 103931044:10.@without.@generic)
aoj:01(100947128:49.@def.@interrogative, 00:46.@wh)
{/unl}
[/S]

NLization


UNL SENTENCE
UNL-NL GRAMMAR
EUGENE
UNL-NL DICTIONARY
NL SENTENCE

UNL-NL Dictionary

- UNL-NL Dictionary Specs
  - UNL-NL 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 two-character language code according to ISO 639-1.
- 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.

Example

Alice

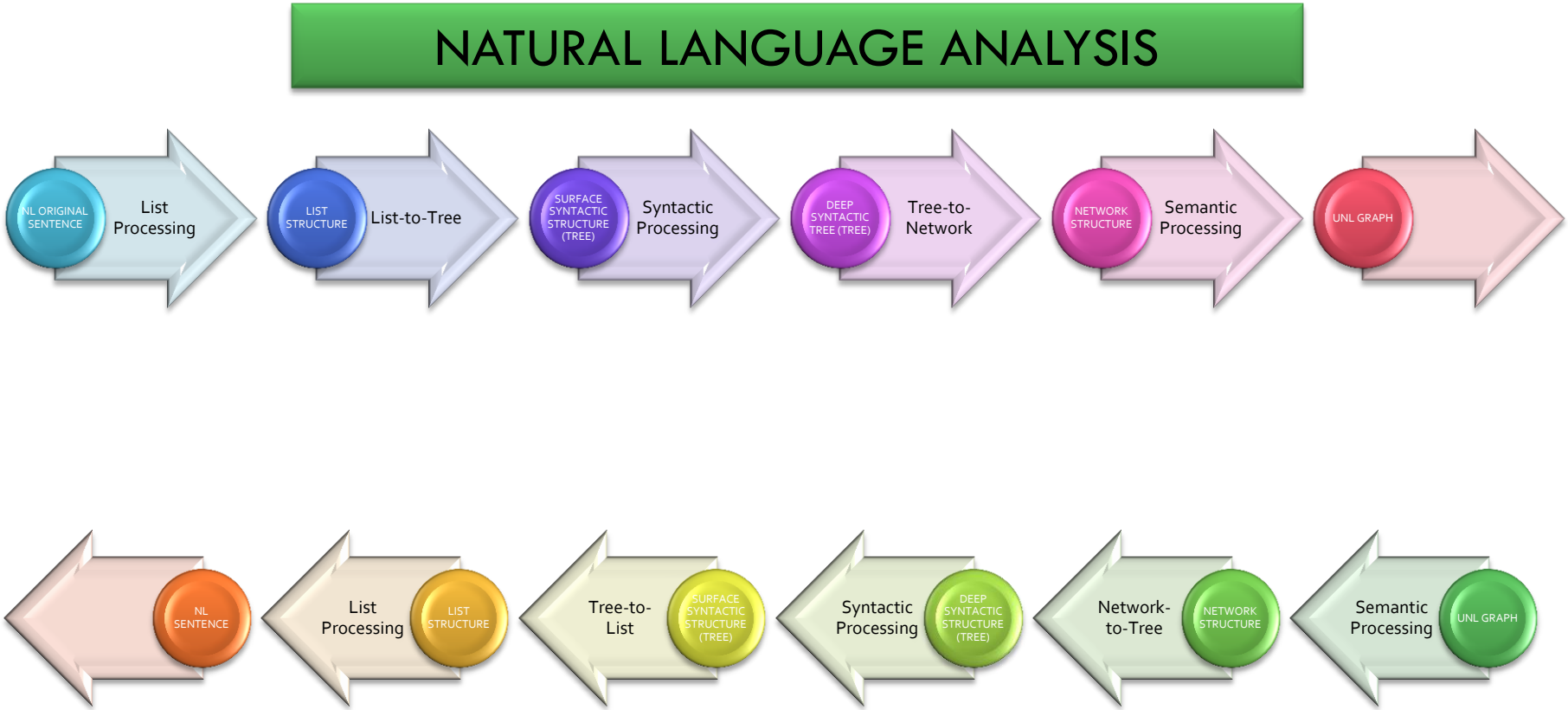
[Alice] {1} "Alice" (N,PPN,FEM) <pt,0,0>;
[aventura] {2} "100796315" (N,NOU,FEM) <pt,0,0>;
[beira-rio] {3} "109213565" (N,NOU,FEM) <pt,0,0>;
[cansado] {4} "302431728" (J,ADJ) <pt,0,0>;
[coelho] {5} "102324045" (N,NOU,MCL) <pt,0,0>;
[conversa] {6} "107133701" (N,NOU,FEM) <pt,0,0>;
[duas vezes] {7} "400065294" (A,AAV) <pt,0,0>;
[espiar] {8} "202165146" (V,VER) <pt,0,0>;
[fazer] {9} "202560585" (V,VER) <pt,0,0>;
[imagem] {10} "103931044" (N,NOU,FEM) <pt,0,0>;
[irmã] {11} "110602985" (N,NOU,FEM) <pt,0,0>;
[ler] {12} "200625119" (V,VER) <pt,0,0>;
[livro] {13} "106410904" (N,NOU,MCL) <pt,0,0>;
[nada] {14} "00" (R,NPR) <pt,0,0>;
[País das Maravilhas] {15} "105632175" (N,PPN,FEM) <pt,0,0>;
[pensar] {16} "200689344" (V,VER) <pt,0,0>;
[sentar] {17} "202374099" (V,VER) <pt,0,0>;
[ter] {18} "202203362" (V,VER) <pt,0,0>;
[toca] {19} "109304465" (N,NOU,FEM) <pt,0,0>;
[uma vez] {20} "400118869" (A,AAV) <pt,0,0>;
[utilidade] {21} "100947128" (N,NOU,FEM) <pt,0,0>;

UNL-NL Grammar

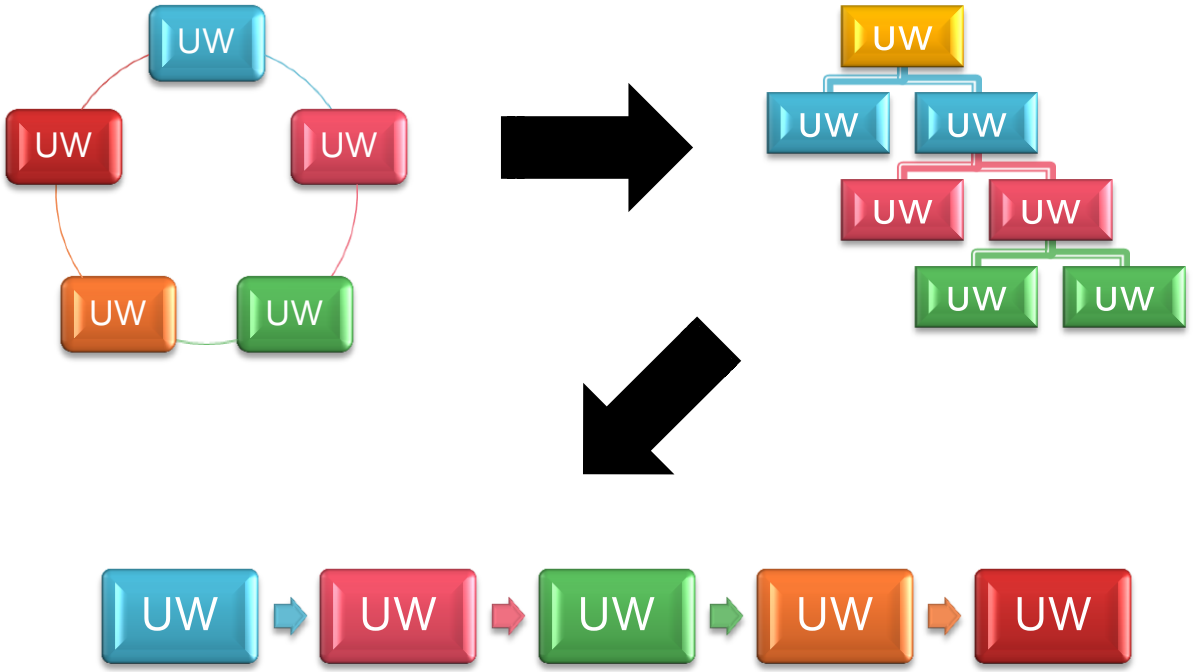
- UNL-NL Grammar Specs
  - a plain text file (.txt)
  - one entry per line
  - there are two types of rules:
    - **TRANSFORMATION RULES** are used to generate natural language sentences out of UNL graphs and vice-versa.
    - **DISAMBIGUATION RULES** are used to improve the performance of transformation rules by constraining their applicability.

Transformation Rules

Transformation Rules

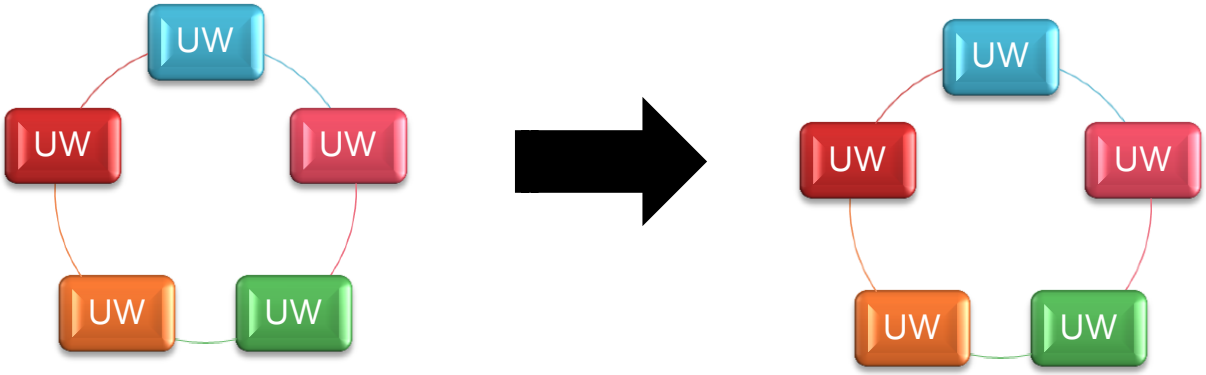


Natural Language Generation



NETWORK-TO-NETWORK (NN)

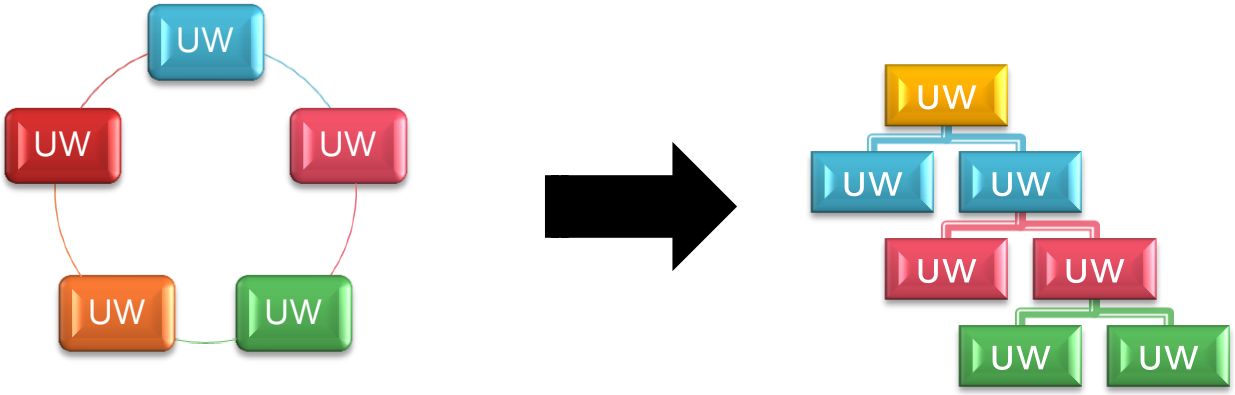
Transformation Rules



ACTION	RULE
ADD RELATION	SEM(A;B):=+SEM(C;D);
DELETE RELATION	SEM(A;B):=-SEM(A,B);
REPLACE RELATION	SEM(A;B):=SEM(C;D);
MERGE RELATION	SEM(A;B),SEM(C;D):=SEM(E;F);
DIVIDE RELATION	SEM(A;B):=SEM(C;D),SEM(E;F);

NETWORK-TO-TREE (NT)

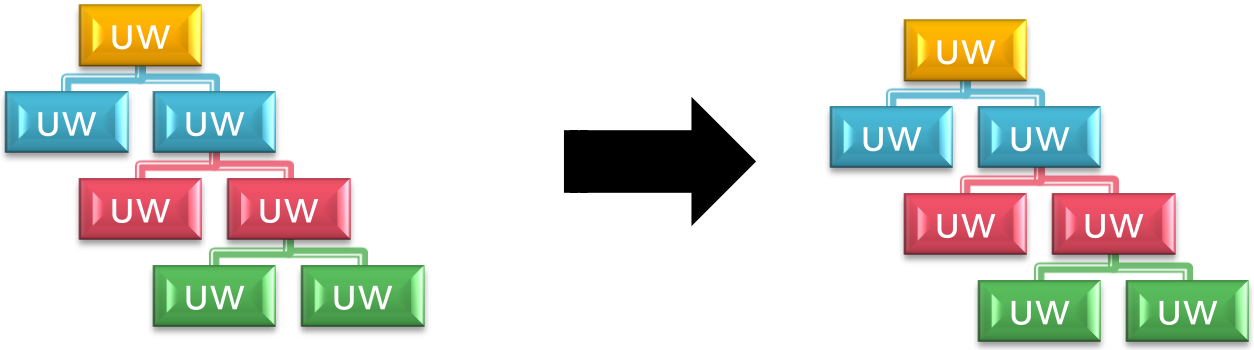
Transformation Rules



ACTION	RULE
REPLACE	SEM(A;B):=SYN(C;D);

TREE-TO-TREE (TT)

Transformation Rules

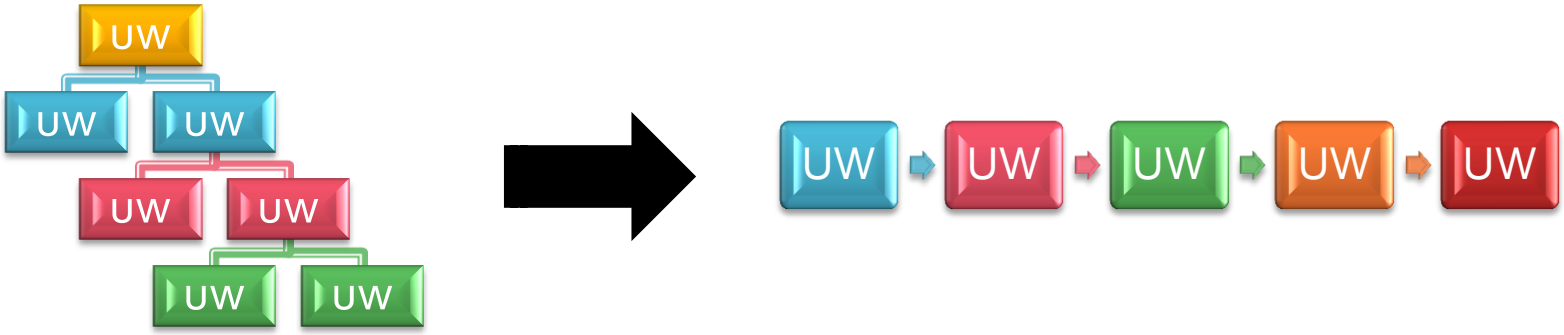


ACTION	RULE
ADD RELATION	SYN(A;B):=+SYN(C;D);
DELETE RELATION	SYN(A;B):=-SYN(A;B);
REPLACE RELATION	SYN(A;B):=SYN(C;D);
MERGE RELATION	SYN(A;B),SYN(C;D):=SYN(E;F);
DIVIDE RELATION	SYN(A;B):=SYN(C;D),SYN(E;F);

ACTION	RULE
ADD NODE	SYN(A;B):=SYN(A;B;C);
DELETE NODE	SYN(A;B):=SYN(A);

TREE-TO-LIST (TL)

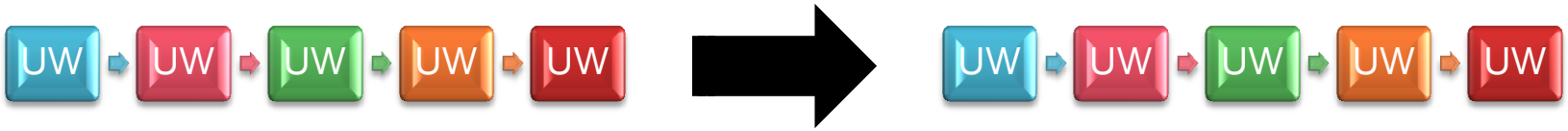
Transformation Rules



ACTION	RULE
REPLACE	SYN(A;B):=(C);

LIST-TO-LIST (LL)

Transformation Rules



ACTION	RULE
ADD	(A):=(A),(B); or (A):=(B),(A);
DELETE	(A):=-(A);
REPLACE	(A):=(B);
MERGE	(A),(B):=(C);
DIVIDE	(A):=(B),(C);

General Properties of
Transformation Rules

PRIORITY
Rules should be applied serially, according to the order defined in the grammar. The first rule will be the first to be applied, the second will the second, and so on.

RECURSIVENESS
Rules should be applied recursively as long as their conditions are true.

COMPREHENSIVENESS
Grammars should be applied comprehensively as long as there is at least one applicable rule.

ACTION
The rules may add or delete values to the source and the target nodes, but only in the right side items:

agt(a;b):=agt(+c;);
agt(a;b):=agt(;-b);

General Properties of
Transformation Rules

CONSERVATION
Rules affect only the information clearly specified. No relation, node or feature is deleted unless explicitly informed.

For instance, in the examples below, the source node of the "agt" relation preserves, in all cases, the value "a". The only change concerns the feature "c", which is added to the source node of the "agt" in the first two cases; and the feature "b", which is deleted from the target node in the third case.

agt(a;b):=agt(c;);
agt(a;b):=agt(+c;);
agt(a;b):=agt(;-b);

General Properties of
Transformation Rules

In any case, the ADD and DELETE rules (i.e., when the right side starts with "+"or "-") **preserve** the items in the left side, except for the explicitly deleted ones:

INPUT:	agt(a;b), obj(a;c), tim(a;d)
RULE:	agt(a;b),^mod(a;e):=+mod(a;e);
OUTPUT:	agt(a;b), obj(a;c), tim(a;d), mod(a;e)

INPUT:	agt(a;b), obj(a;c), tim(a;d)
RULE:	agt(a;b):=-agt(a;b);
OUTPUT:	obj(a;c), tim(a;d)

General Properties of
Transformation Rules

The REPLACE, MERGE and DIVIDE rules affect only their designated scopes. In that sense, NN may only replace, merge or divide semantic relations; TT may only replace, merge or divide syntactic relations; and LL may only replace, merge or divide list nodes. All other information is preserved, unless explicitly informed.

INPUT:	agt(a;b), cob(a;c)
RULE:	cob(;):=obj(;);
OUTPUT:	agt(a;b), obj(a;c)

INPUT:	agt(a;b), cob(a;c)
RULE:	cob(a;):=obj(-a,+d;);
OUTPUT:	agt(a;b), obj(d;c)

General Properties of
Transformation Rules

CONJUNCTION

Both the left and the right side of the rule may have as many items as necessary, provided that they are interlinked by ",", as exemplified below:

SEM(A;B),SEM(C;D),SEM(E;F):=SEM(G;H),SEM(I;J),SEM(K;L);

DISJUNCTION

The left side of the rules may bring disjuncts, but not the right side.

{SEM(A;B),SEM(C;D)},^SEM(E;F):=+SEM(E;F);

SEM(A;B),{SEM(C;D),SEM(E;F)}:=-SEM(A;B);

agt(VER,{V01,V02};NOU,^SNG}:=;

General Properties of
Transformation Rules

EXTENDIBILITY
The left side of the rules may bring wildcard characters, such as the ones indicated in Page 4.

"cit?(??)":=[[city(icl>metropolis)]];
"cit*":=[[city(icl>metropolis)]];
" #### " :=+YEAR;

General Properties of
Transformation Rules

CONCISION
Rules should be as small as possible. In that sense, the source and the target nodes may be simple placeholders or indexes:

cob(;):=obj(;);
tim(%01;[[in]]),obj([[in]];%02):=tim(%01;%02);
tim(VER,%01;[[in]]),obj([[in]];NOU,%02):=tim(%01;%02);

By default, the first node to appear in the left side is %01, the second is %02, and so on. The same to the right side. Therefore, the last rule above may be rewritten as:

tim(VER;[[in]]),obj([[in]];NOU):=tim(;%04);

In the DELETE rules, the right side may be omitted in case of deletion of the entire left side:

obj(PRE;):=;

General Properties of
Transformation Rules

READABILITY
There can be blank spaces between variables and symbols. Comments can be added after the ";".

obj (;) := ; this rule deletes every "obj" relation.

SCOPE GENERATION
Inside a relation, nodes can be relations (i.e., hypernodes) as well.

SEM(A;SEM(C;D)):=SEM(A;C),SEM(C;D));
SEM(A;C),SEM(C;D):=SEM(A;SEM(C;D));

General Properties of
Transformation Rules

COMMUTATIVITY

Inside the same side of the rule, the order of the factors does not affect the end result, except for list-processing rules (LL, LT and TL).

SEM(A;B):=SEM(C;D),SEM(E;F); = SEM(A;B):= SEM(E;F),SEM(C;D);
SYN(A;B):=SYN(C;D),SYN(E;F); = SYN(A;B):= SYN(E;F),SYN(C;D);

But:
(A):=(B),(C); ≠ (A):=(C),(B);
SYN(A;B):=(C),(D); ≠ SYN(A;B):=(D),(C);
(C),(D):=SYN(A;B); ≠ (D),(C):=SYN(A;B);

Additionally, the order of the features inside a relation does not affect the end result, but the order of the nodes is non-commutative.
SEM(VER,TRA ; NOU,MCL) = SEM(TRA,VER ; MCL,NOU)
But:
SEM(VER,TRA ; NOU,MCL) ≠ SEM(NOU,MCL ; VER,TRA)

General Properties of
Transformation Rules

DICTIONARY ATTRIBUTES
Dictionary attributes can be used as variables with the value "&".

SYN(A,^num ; B,num=&01):=SYN(A,num=&01;);

DICTIONARY RULES
Dictionary rules can be triggered by "!".

(@pl, pl:="feet"):=(!pl,-@pl);
(@pl, pl:="oo":"ee"):=(!pl,-@pl);
(@pl, pl:="y">"ies"):=(!pl,-@pl);
(@pl, pl:=1>"ies"):=(!pl,-@pl);
(@pl, pl=*):=(!pl,-@pl);

Example

LIST-TO-LIST

- aventura.@pl > aventuras
- conversa.@pl > conversas
- imagem.@pl > imagens

(@pl):=(-@pl,"a">"as","m">"ns");

NETWORK-TO-LIST

- pos(toca, coelho) = toca do coelho
  - pos(toca,coelho)
  - NA(toca;PC(de;coelho));
  - (toca)(PC(de;coelho));
  - (toca)(de)(coelho)

  - pos(N,%x;N,%y):=NA(%x;PC(“de”;%y)); (NT)
  - NA(%x;%y):=(%x)(%y); (TL)
  - PC (%x;%y):=(%x)(%y); (TL)

  - pos(N,%x;N,%y):=(%x)("de")(%y);

http://www.ronaldomartins.pro.br/alice

Atividade

Resultado

Alice estava começando a ficar muito cansado de ficar sentado por sua irmã no banco, e de não ter nada para fazer. Uma ou duas vezes ela espiou o livro em sua irmã estava lendo, mas não tinha imagens ou conversas em que, "e o que é o uso de um livro", pensou Alice, "sem figuras nem diálogos?"

Alice estava começando a obter muito cansado do assento por sua irmã no banco, e de não não ter nada fazer. Uma vez que ou duas vezes peeped no livro que sua irmã lia, mas não teve nenhuma retrato ou conversação nele, “e o que é o uso de um livro,” pensamento Alice, “sem retratos ou conversações?”

Alice estava começando a ficar muito cansada de sentar ao lado de sua irmã à beira-rio, e de ter nada para fazer. Uma vez ou duas vezes tinha espiado o livro que sua irmã estava lendo, mas o livro não tinha nenhuma imagem ou nenhuma conversa, e Alice pensou qual é a utilidade de um livro sem imagens ou sem conversas?

ronaldo.martins@undlfoundation.org
http://www.unlweb.net

Muito obrigado!