

# XIV UNL School

Patras, March 10-14, 2014

# Day #2

Morning

- Corpus
  - NLization
- Dictionaries (I)
  - UNL>NL Dictionary

Lunch break

Afternoon

- Grammar: Morphology
  - Inflectional Paradigms

Corpus

# 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.@trial = teachers
- Some attributes are context-sensitive
  - teacher.@multal = many teachers
  - water.@multal = much water
  - book.@most = most of the book
  - book.@most.@pl = most books



# Important!

## (relations)

- There can be more than one way of realizing a relation
  - $\text{pos}(\text{book}, \text{John}) = \text{book of John, John's book}$
- Relations constitute a hierarchy, where lower relations may be replaced by upper relations + attribute
  - $\text{src}(\text{book}, \text{library}) = \text{book from the library}$
  - $\text{plc}(\text{book}, \text{library.}@from) = \text{book from the library}$
- Relations are not commutative
  - $\text{cnt}(\text{evidence}, \text{absence}) = \text{evidence of absence}$
  - $\text{cnt}(\text{absence}, \text{evidente}) = \text{absence of evidence}$
- The relation is always defined by the target
  - $\text{relation}(\text{source}, \text{target}) = \text{target is relation of source}$ 
    - $\text{pos}(\text{book}, \text{John}) = \text{John is possessor of book (John's book)}$
    - $\text{ben}(\text{book}, \text{John}) = \text{John is beneficiary of book (book for John)}$
    - $\text{cnt}(\text{book}, \text{John}) = \text{John is content of book (book about John)}$
    - $\text{mat}(\text{book}, \text{stone}) = \text{stone is the material of book (book of stone)}$
    - $\text{and}(\text{John}, \text{Mary}) = \text{Mary is coordinated to John (John and Mary)}$

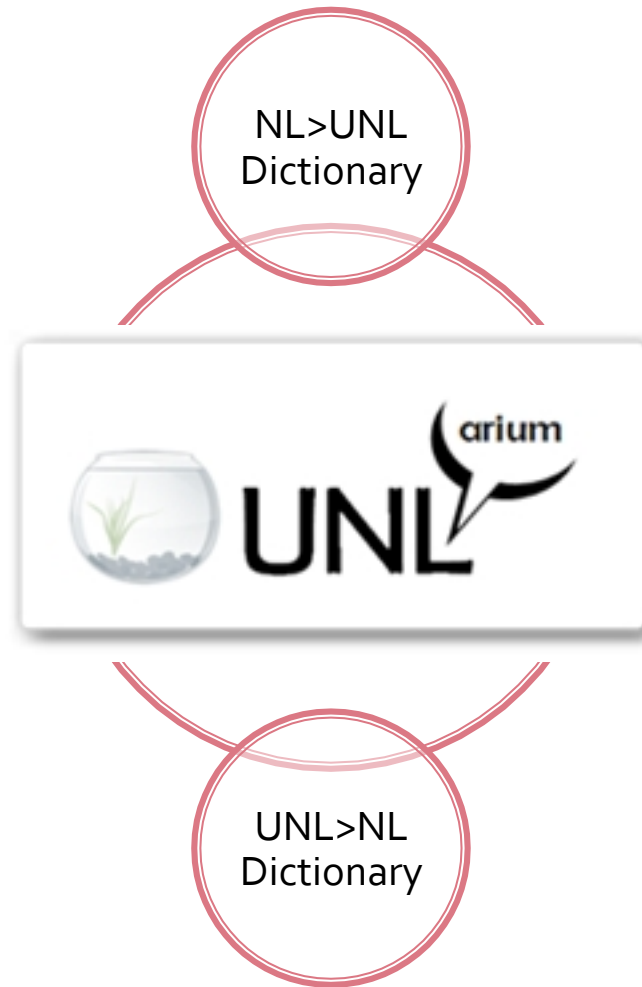
# Discussion #2

---

# Dictionaryes

---

# Dictionaries



Activity

# Aesop-3

# Aesop-3

- To create the UNL>NL Dictionary of the project AESOP-3 available at the Sandbox

# Important!

- You must propose lemmas corresponding to the UW (unambiguous) and not the English word (ambiguous)
  - table, tabular array (a set of data arranged in rows and columns)
    - French: tableau (~~table~~)
- Your lemmas must correspond to LEXICAL UNITS (i.e., words that can be found as headwords or sub-entries in ordinary dictionaries)
  - ilunga (a person who is ready to forgive any abuse for the first time, to tolerate it a second time, but never a third time)
    - English: ilunga or NONE (OVERSPECIFIED)
- Your lemmas must be lemmas (citation form) and not word forms (inflections)
  - friend (a person you know well and regard with affection and trust)
    - Spanish: amigo (~~amiga, amigos, amigas~~)
- Register is used only when the word does not belong to the standard use of the language (colloquialism, archaism, jargon, slang, etc.)

SLURRP



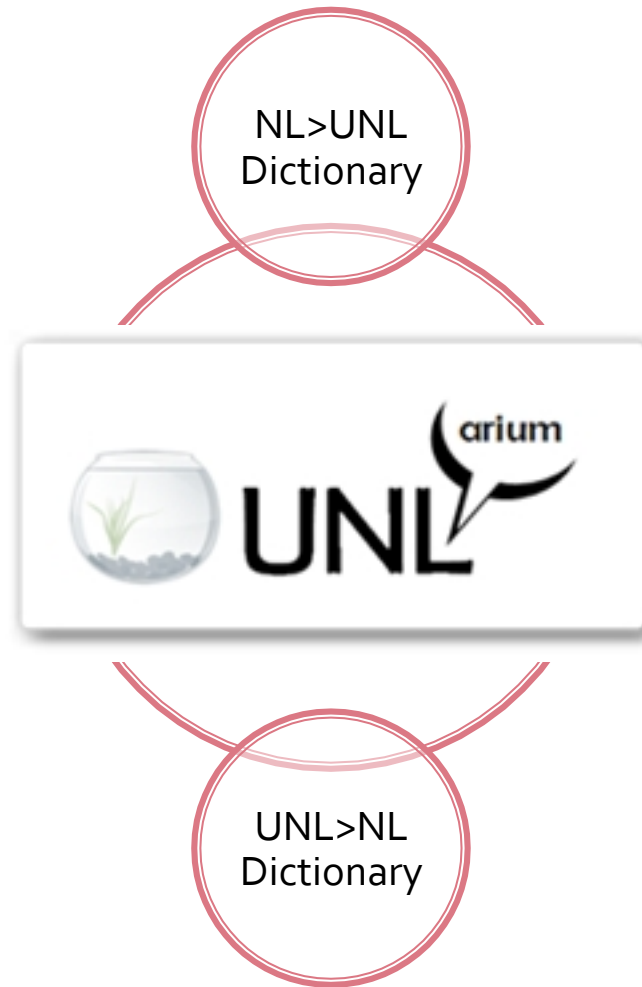


Grammar

# Morphology

---

# Dictionaries



# Generative Dictionary

base form > word forms



# Step #1

- To identify all possible word forms
  - WORD FORM = ROOT (STEM) + AFFIXES
    - AFFIXES
      - PREFIXES
      - INFIXES
      - SUFFIXES
      - CIRCUMFIXES (PREFIX + SUFFIX)
- However: what is a "word form"?
  - nation x nation-**s**
  - nation x nation-**al**
  - nation x nation-**al-ise**
  - nation x **de**-nation-**al-ise**
  - nation x **de**-nation-**al-ise-d**

# Lexicalization x Grammaticalization of Linguistic Categories

## Gender representation in English

NO GENDER	blood, sea		
ISOLANT	DIFFERENT ROOTS	boy	girl
AGGLUTINANT	SAME ROOT + DIFFERENT SUPPORTING WORDS	male teacher	female teacher
INFLECTIONAL	SAME ROOT + DIFFERENT AFFIXES	actor	actress

# Different types of affixation

## INFLECTION

(ROOT + AFFIX = same UW)

- book > books
- write > wrote

## DERIVATION

(ROOT + AFFIX = new UW)

- glory > glorify
- write > writer

## COMPOSITION

(ROOT + ROOT = one UW)

- write > will write, write down

# Instructions (I)

## ■ INFLECTIONS

- Are not simply "concatenated" to roots
  - "John's" is not an inflection of "John"
  - "Populusque" is not an inflection of "Populus"
- Are not generated by "periphrases"
  - "will go" is not an inflection of "go"
  - "have gone" is not an inflection of "go"
  - "little boy" is not an inflection of "boy"
  - "more beautiful" is not an inflection of "beautiful"
- Do not change the "category" of the word
  - "to google" is not an inflection of "google"
- Do not change the "meaning" of the word
  - "booklet" is not an inflection of "book"
  - "writer" is not an inflection of "write"

# Instructions (II)

## ■ INFLECTIONS

- may involve changes in the root
  - άνθρωπος > ανθρώπους
  - foot > feet
  - be > am
- may be limited to certain words
  - big > bigger
- may be done with the morpheme zero
  - fish (singular) > fish (plural)



# Step #2

- To analyse the inflections:
  - rosa = nominative singular = NOM&SNG
  - rosae = nominative plural = NOM&PLR
  - rosam = accusative singular = ACC&SNG

YOU HAVE TO USE THE TAGSET!

- bonum = masculine accusative singular
- ...
- bonarum = feminine genitive plural = FEM&GNT&PLR
- bonas = feminine accusative plural = FEM&ACC&PLR
- ...
- bonum = neutral nominative singular = NEU&NOM&SNG
- ...

# Be careful!

- Include all and only INFLECTIONAL CATEGORIES
  - rosa = ~~FEM&NOM&SING~~
    - There is no MCL for "rosa" (i.e., "rosa" is not inflectional in gender)

# Step #3

- To simplify as much as possible
  - Latin
    - bonis = MCL&DAT&PLR
    - bonis = FEM&DAT&PLR
    - bonis = NEU&DAT&PLR
    - THEREFORE:
      - bonis = DAT&PLR (irrespective of gender)
  - English
    - loved = 1PS&PAS&IND
    - loved = 2PS&PAS&IND
    - loved = 3PS&PAS&IND
    - loved = 1PP&PAS&IND
    - loved = 2PP&PAS&IND
    - loved = 3PP&PAS&IND
    - THEREFORE
      - loved = PAS&IND

Activity

# Inflectional Schemas

# Inflectional Schema

- To identify the inflectional schema of all the words introduced in the project AESOP-3
  - Go to SANDBOX>DICTIONARY>BROWSE
  - Copy the list of words
  - For each POS (A, N, J, V)
    - List all possible INFLECTIONS
    - Analyse them into inflectional schemas
    - Simplify the inflectional schemas
    - Introduce the inflectional schema to SANDBOX>GRAMMAR>LANGUAGE SETTINGS>INFLECTIONAL SCHEMA

# Example of inflectional schema

- NOM&SNG
- NOM&PLR
- VOC&SNG
- VOC&PLR
- ACC&SNG
- ACC&PLR
- GNT&SNG
- GNT&PLR
- DAT&SNG
- DAT&PLR
- ABL&SNG
- ABL&PLR

# Inflectional Languages

"usual" behavior

## LEXICALIZED

- PART OF SPEECH
- LEXICAL CATEGORY
- ABSTRACTNESS (N)
- ALIENABILITY (N)
- ANIMACY (N)
- CARDINALITY (N)
- DEFINITENESS (N)
- DEGREE (J, A)
- GENDER (N)
- POLARITY
- REGISTER
- SOLIDARITY
- STATUS
- TRANSITIVITY (V)

## GRAMMATICALIZED

- NOUNS
  - NUMBER
  - CASE
  - DEGREE
- ADJECTIVES
  - GENDER
  - NUMBER
  - CASE
- VERBS
  - NUMBER
  - PERSON
  - TENSE
  - ASPECT
  - MOOD
  - VOICE
- ADVERBS
  - NONE

# Inflections

- Inflectional categories
  - NUMBER: NUM={SNG|PLR|DUA|...}
  - GENDER: GEN={MCL|FEM|COM|NEU|INV}
  - CASE: CAS={NOM|ACC|DAT|...}
  - ABSOLUTE TENSE: ATE={PRS|PAS|FUT|...}
  - RELATIVE TENSE: RTE={RPT|RPS|RFT|...}
  - ASPECT: ASP={PFV|NPFV|CAU|...}
  - MOOD: MOO={IND|IMP|SUB|...}
  - VOICE: VOI={ACV|PSV|MIV}
  - PERSON: PER={1PS|2PS|...}



# Important!

- Be careful about prescriptive biases:
  - Irregular inflections are inflections
    - mouse > mice (inflection): mouse.@pl
    - actor > actress (inflection): actor.@female
- Open issues
  - undo = undo x do.@not
  - redo = redo x do.@again
  - cow = cow x bull.@female

# Inflectional Rules

---

# Step #4

- To identify "inflectional paradigms", i.e., to group words according to their inflectional behavior

# Example



# Inflections

## INFLECTIONAL PARADIGMS

- AFFIXATION  
(PREFIXATION,  
SUFFIXATION,  
INFIXATION) TO THE BASE  
FORM
- REGULAR OR SEMI-  
REGULAR
- GRAMMAR
- book>books

## INFLECTIONAL RULES

- AFFIXATION  
(PREFIXATION,  
SUFFIXATION,  
INFIXATION) TO THE BASE  
FORM
- IRREGULAR
- DICTIONARY
- mouse>mice

# Step #5

- To create inflectional paradigms for the regular or semi-regular groups of words
  - A-rules

# Affixation (A-rules)

- Prefixation
  - AFFIX<BF;
    - NOT:="un"<o;
      - do>undo
- Suffixation
  - BF>AFFIX;
    - PLR:=o>"s";
      - book>books
- Infixation
  - [BF]:AFFIX;
  - AFFIX<[BF];
  - [BF]>AFFIX;
- Replacement
  - BF:AFFIX;

# Example of inflectional paradigm

- NOM&SNG:="a">"a"; rosa>rosa
- NOM&PLR:="a">"ae"; rosa>rosae
- VOC&SNG:="a">"a"; rosa>rosa
- VOC&PLR:="a">"ae"; rosa>rosae
- ACC&SNG:="a">"am"; rosa>rosam
- ACC&PLR:="a">"as"; rosa>rosas
- GNT&SNG:="a">"ae"; rosa>rosae
- GNT&PLR:="a">"arum"; rosa>rosarum
- DAT&SNG:="a">"ae"; rosa>rosae
- DAT&PLR:="a">"is"; rosa>rosis
- ABL&SNG:="a">"a"; rosa>rosa
- ABL&PLR:="a">"is"; rosa>rosis



Activity

# Inflectional Paradigms

# Inflectional Paradigms

- Create inflectional paradigms for the groups of regular or semi-regular words appearing in the corpus
  - Go to UNLARIUM>GRAMMAR>INFLECTIONAL PARADIGM
- Test your paradigm

# Inflectional Paradigms

- Only inflectional categories
  - NUM, GEN, CAS, ...
  - ~~DEG, DEF, POL, ...~~
- Only affixation
  - kill>kills, killed, killing
  - ~~kill>will kill, is killing, have killed, ...~~
- Only regular
  - book>books, bus>buses, ...
  - ~~mouse>mice, child>children, ...~~

# Important!

- You can only use the tagset
  - masculine = MCL
- Attributes are conjoined through &
  - MCL&SNG
- The order of attributes is not important
  - MCL&SNG = SNG&MCL
- Rules are not cumulative
  - FEM:=2>"es"; (actor>actress)
  - PLR:=0>"es"; (actor>actores)
- Rules must be mutually exclusive
  - ~~PLR:=0>"s"; (actor>actors)~~
  - ~~FEM&PLR:=0>"es"; (actor>actors>actresses)~~
- In case of alternative forms, use ALT
  - PLR:=0>""; (fish>fish)
  - PLR&ALT:=0>"es"; (fish>fishes)

