

Geneva, July 2nd

# XII UNL School

## Day #2



# Day #2

- ~~Welcome~~
- ~~Context~~
- ~~Normalization Grammar~~
- Closed-Class Dictionary
  - Universal Words
  - Lexical Resources
  - Dictionaries
  - Closed-Class Dictionary

# Universal Words

---

# Universal Words

- nodes in the UNL graph
- temporary (not to be included in the UNL Dictionary)
  - 3.14159
  - [www.undlfoundation.org](http://www.undlfoundation.org)
  - H<sub>2</sub>O
- 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

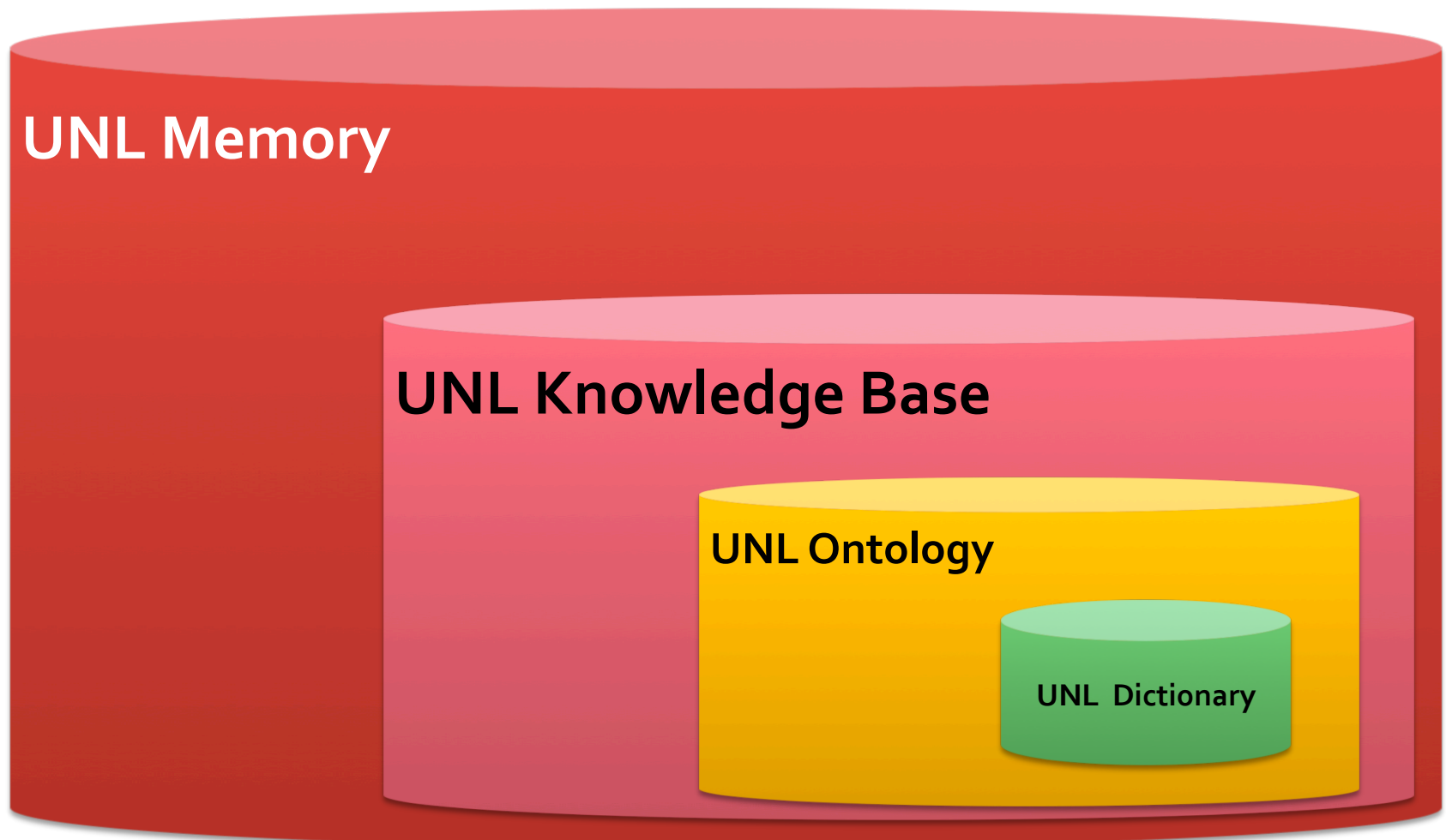
## UCI (Uniform Concept Identifier)

- ❑ 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)

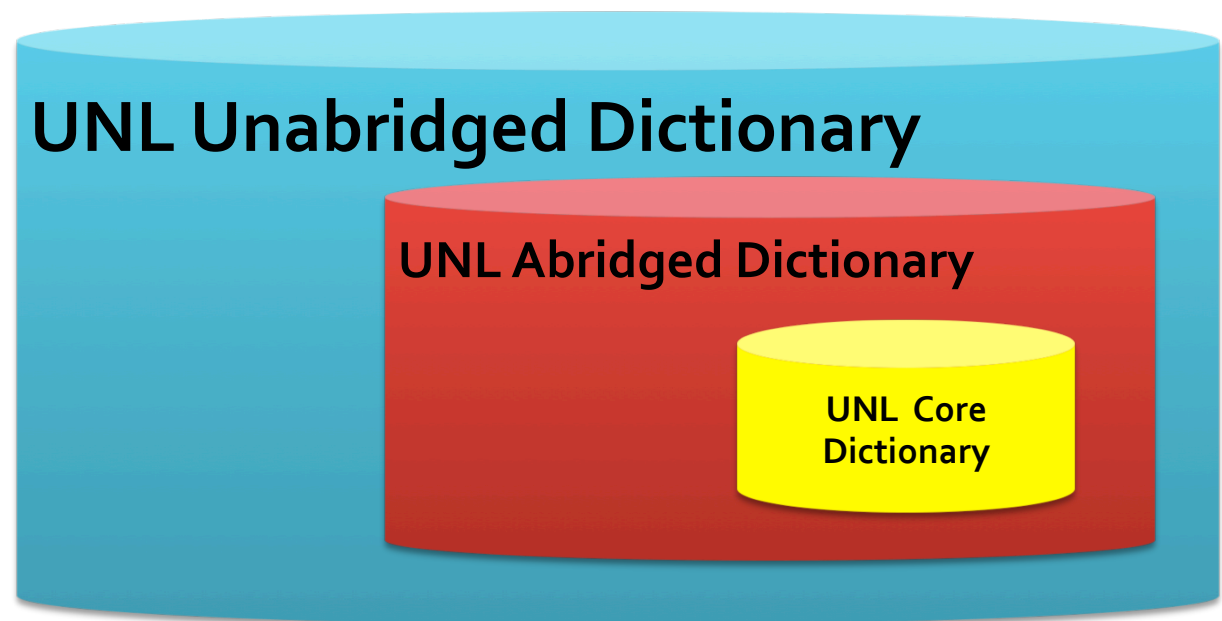
# Lexical Resources

---

# UNL Lexical Resources

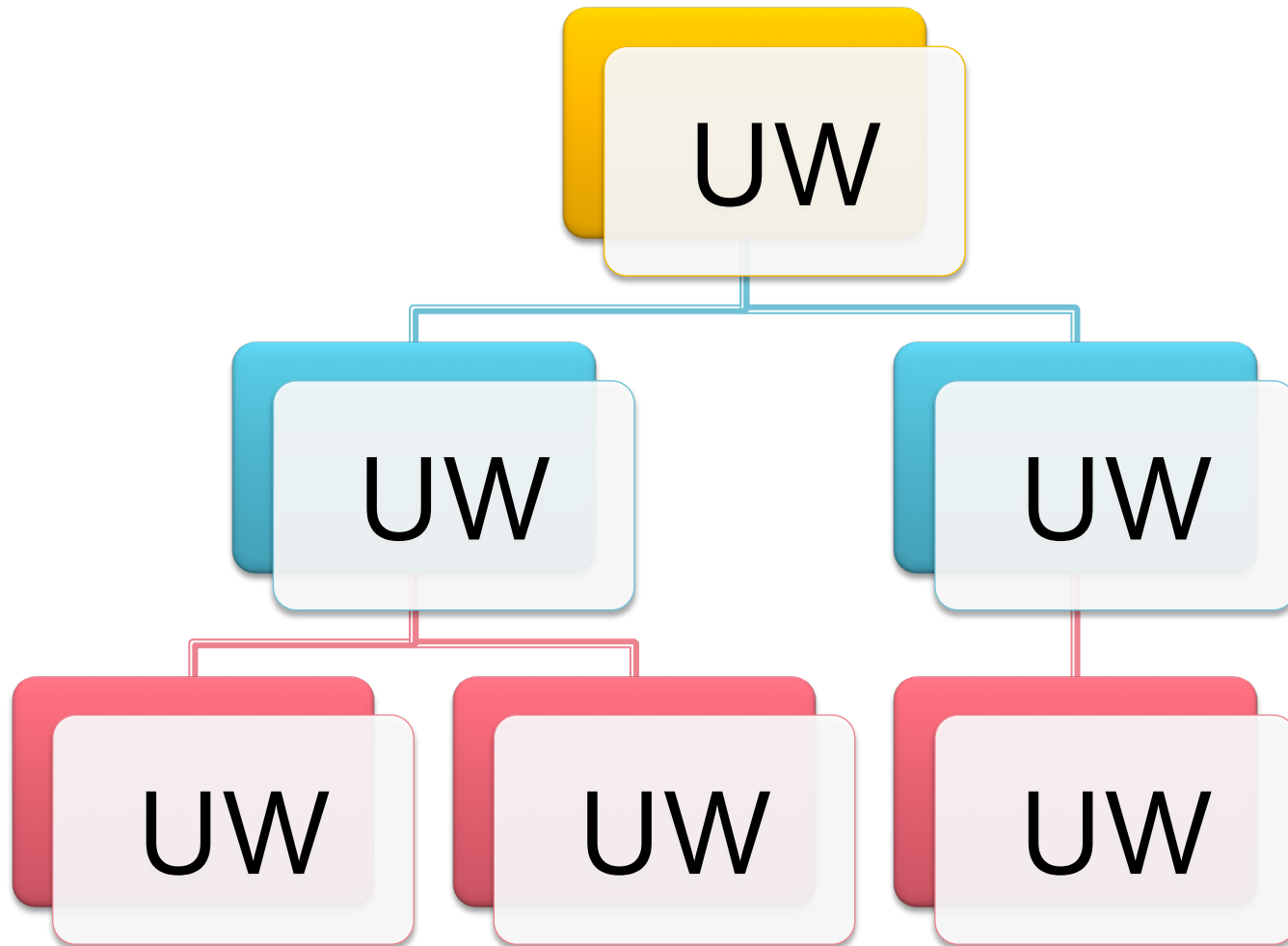


# UNL Dictionary

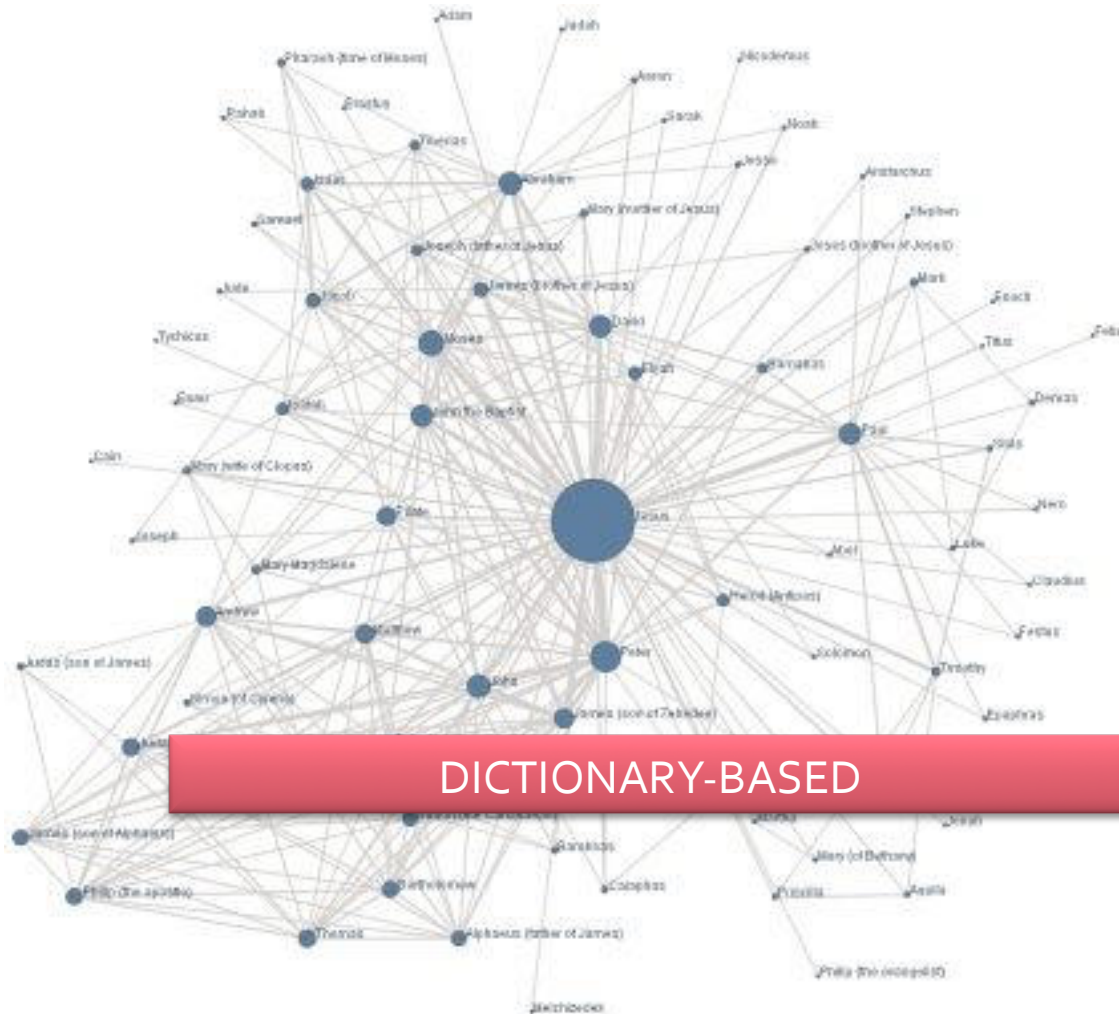




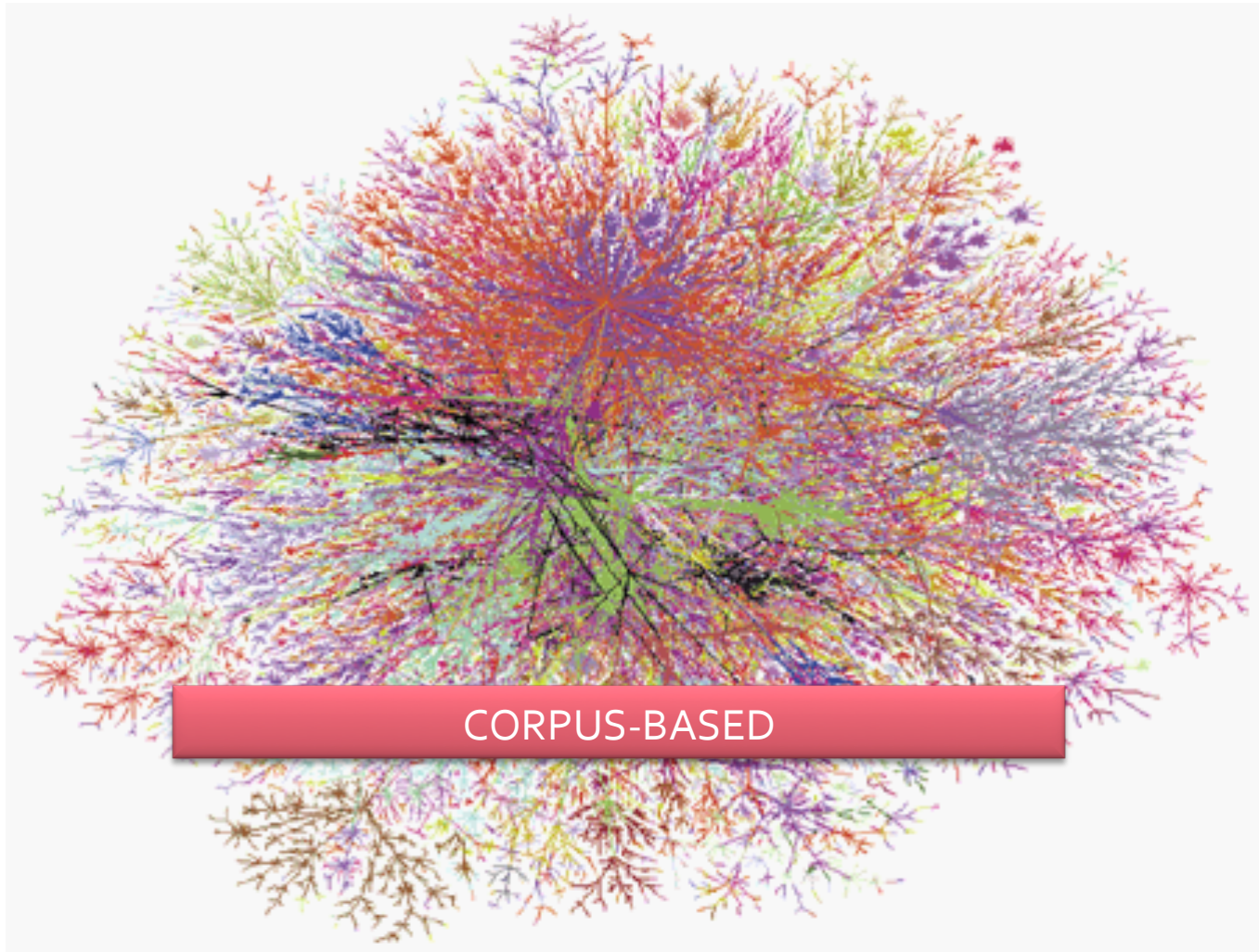
# UNL Ontology



# UNL Knowledge Base



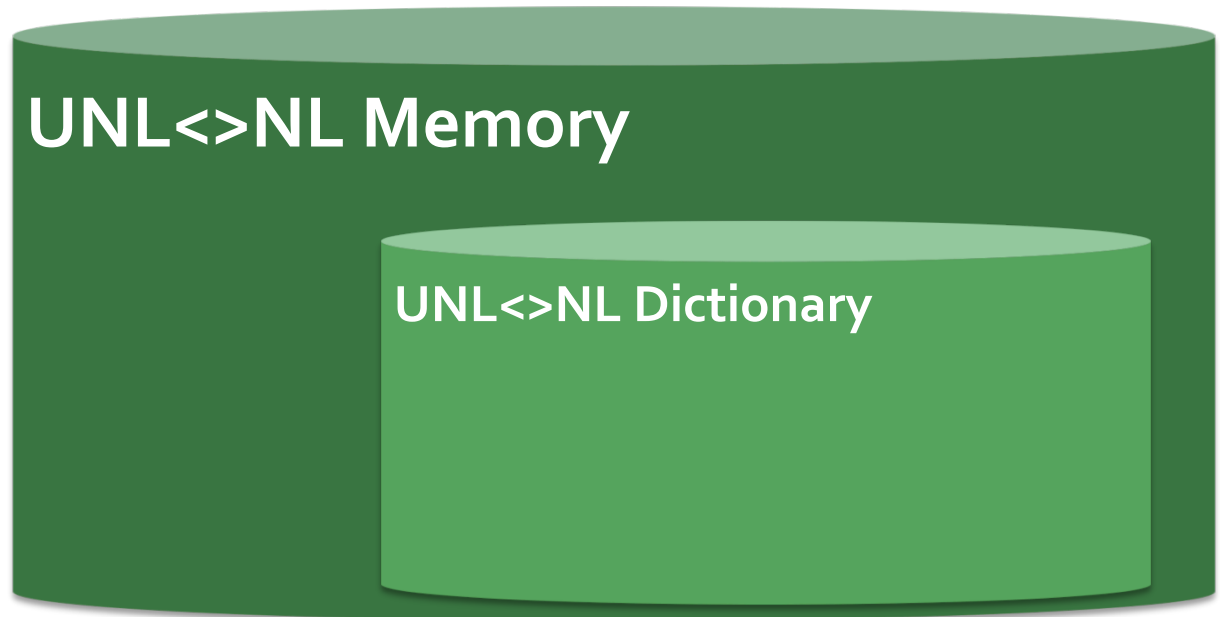
# UNL Memory



# NL Lexical Resources



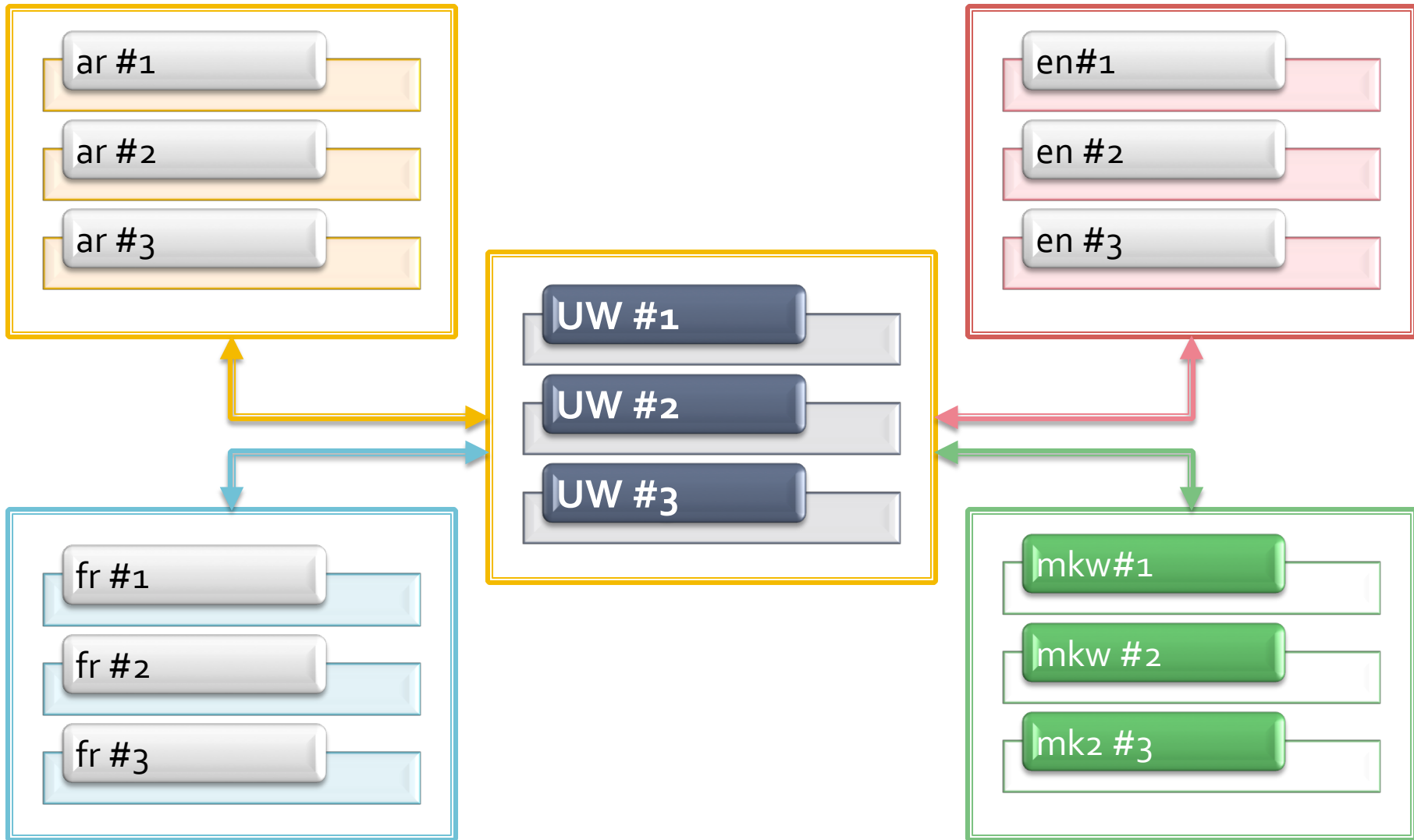
# UNL<>NL Lexical Resources



# UNL<>NL Dictionaries

---

# UNL<->NL Dictionaries



# Types of Dictionary

- Default Dictionary
  - punctuation signs, regular expressions to deal with URL's, e-mail addresses, etc.
- Closed-Class Dictionary
  - determiners, prepositions, conjunctions, etc.
- Open-Class Dictionary
  - nouns, verbs, adverbs, adjectives



# Open-Class Dictionary

- NL-UNL Dictionary (Analysis)

- Enumerative (word forms)

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

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

- UNL-NL Dictionary (Generation)

- Generative (base forms)

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



# Dictionary Specs

[www.unlweb.net/wiki/Dictionary\\_Specs](http://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 : (PLR:="oo":"ee")
    - Replacement
      - <ATTRIBUTE>":="<SOURCE>":"<TARGET>;
      - PLR:="oo":"ee";
    - Left appending
      - <ATTRIBUTE>":=<LEFT ADDITION>"<"<LEFT DELETION>;
      - NOT:="<un>;
    - Right appending
      - <ATTRIBUTE>":=<RIGHT DELETION">"<RIGHT ADDITION>;
      - PLR:="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

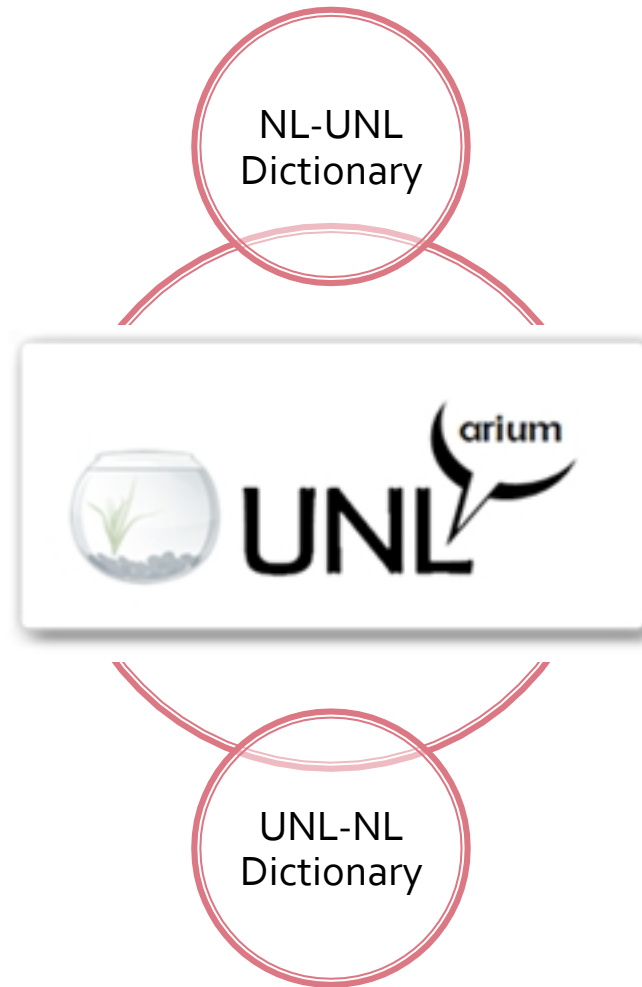
- [China]{24} "China(iof>Asian country)" (NOU, WRD, SNG, Po, Fo) <eng,0,0>;
- [choose]{106} "to choose(icl>to decide)" (POS=VER, LEX=WRD, PAR=M1, FRA=Y76, FLX(3PS&PRS&IND:=o>"s"; PAS:="chose"; PTP:="chosen"; GER:="choosing";)) <eng,0,0>;
- [clear-eyed]{25} "clear-eyed(icl>discerning)" (POS=ADJ, LEX=WRD, 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



# Closed-Class Dictionary

---

# Structure (I)

- Determiners (LEX=D, UW = "",att)
  - Articles: the, a
  - Quantifiers: all, any, few, many, much, etc.
  - Demonstrative determiners: this, that, same, etc.
  - Interrogative determiners: what, which, etc.
  - Possessive determiners (UW="oo"): my, your, etc.
- Auxiliaries (LEX=I, UW = "",att)
  - Auxiliary verbs ("do", "have", "will")
  - Modal auxiliary verbs ("can", "should", "must", etc.)

# Structure (II)

- Copula and linking verbs (LEX=V, UW = "")
  - be, look, etc.
- Pronouns (LEX=R, UW="oo")
  - PERSONAL PRONOUNS (POS=PPR): I, you, ...
  - POSSESSIVE PRONOUNS (POS=SPR): mine, yours, ...
  - REFLEXIVE PRONOUNS (POS=FPR): myself, yourself, ...
  - RECIPROCAL PRONOUNS (POS=CPR): each other, one another
  - INDEFINITE PRONOUNS (POS=NPR): anybody, nobody, ...
  - INTERROGATIVE PRONOUNS (POS=IPR): who, where, ...
  - DEMONSTRATIVE PRONOUNS (POS=DEP): this, there, ...
  - RELATIVE PRONOUNS (POS=RPR): who, that,....

# Structure (III)

- Specifier Adverbs (LEX=A, UW="", att)
  - not, very, almost, etc.
- Prepositions (LEX=P, UW="", att, rel)
  - in, on, above, from, etc.
- Conjunctions (LEX=C, UW="", att, rel)
  - and, or, because, etc.
- Affixes (LEX=F, UW="")
  - st, nd, 's, etc.

# Be careful with homonyms:

- Determiners (modify) ≠ Pronouns (replace)
  - his book ≠ the book is his
  - that book ≠ the book is that
  - both books ≠ both are books

# Exercise #3

- Create the closed-class dictionary for your native language (you may use the English Closed-Class Dictionary as an "inspiration")
- Upload the dictionary to the UNLarium (UNLARIUM>DICTIONARY>IMPORT)