Geneva, July 5th

# XII UNL School

## Day #5

# Day #5

- ~~Welcome~~
- ~~Context~~
- ~~Normalization Grammar~~
- ~~Closed-Class Dictionary~~
- ~~Open-Class Word List~~
- ~~Corpus~~
- ~~Bruno-A1~~
- ~~NC-A1~~
- Evaluation and Discussion

# Evaluation & discussion

# Evaluation & Discussion

- Workshop
- Infrastructure

# Workshop

- Problems & Solutions
  - What should be preserved in the next editions?
  - What should be changed in the next editions?

# Infrastructure

- UNLARIUM, VALERIE, IAN, EUGENE, WIKI
  - What should be changed?
  - Which would be the features to be included?

# Follow-up &
# Next steps

# FoR-UNL (modified)

| LEVEL | DICTIONARIES | | | GRAMMARS | | | |
|---|---|---|---|---|---|---|---|
| | GD | ND | AD | UC | | NC | |
| | | | | ANA | GEN | ANA | GEN |
| A1 | 2,000 | 2,000 | 2,000 | 100 | 100 | 100 | 100 |
| A2 | 3,000 | 3,000 | 3,000 | 300 | 300 | 300 | 300 |
| B1 | 5,000 | 5,000 | 5,000 | 500 | 500 | 500 | 500 |
| B2 | 5,000 | 5,000 | 5,000 | 500 | 500 | 500 | 500 |
| C1 | 5,000 | 5,000 | 5,000 | 500 | 500 | 500 | 500 |
| C2 | 5,000 | 5,000 | 5,000 | 500 | 500 | 500 | 500 |

# Status (A1)

| LANGUAGE | GD | ND | AD | UGA | UGG | NGA | NGG |
|---|---|---|---|---|---|---|---|
| Arabic | 100% | 100% | | | | | |
| Armenian | 49% | - | | | | | |
| Bulgarian | 100% | 23% | | | | | |
| Chinese | 100% | 100% | 100% | | | | |
| Kannada | - | - | | | | | |
| Khmer | 50% | - | | | | | |
| Malay | 100% | 100% | 100% | | | | |
| Panjabi | - | - | | | | | |
| Ukrainian | 100% | 100% | | | | | |

# Post-workshop tasks
## deadline = 30/09/2013

- Open-Class Word List (3,000 entries)
  - LEMMA + LEX
- Corpus NC-A1
  - Original corpus: 5-10 original articles from the Wikipedia about culture-specific subjects (minimum of 5,000 words), in separate files, in plain text format with UTF-8 encoding
  - List of at least 1,000 noun phrases appearing in the corpus

# NP's

- the length of the NP must be equal or greater than 2 words (one-word NP's must be excluded): Geneva
- NP's must not contain foreign words: the city of Genève (note that "the city of Geneva" is OK)
- NP's must be continuous (there cannot be any extra-content, e.g., parentheses, inside the NP): the second most populous city in Switzerland (after Zurich) (note that the NP will be "the second most populous city in Switzerland")
- NP's must not contain verbs, even when used as nouns, adjectives or adverbs: French-speaking part of Switzerland, numerous international organizations, including the headquarters of many of the agencies of the United Nations and the Red Cross (in the latter case, there will be 2 NP's: "numerous international organizations" and "the headquarters... Red Cross")
- NP's must be original (no change should be made to the original text from the Wikipedia)
- NP's must ignore nesting (only the longest NP must be considered): "the headquarters of many of the agencies of the United Nations and the Red Cross" must be treated as a single NP (the inner NP's, such as "the agencies of the United Nations and the Red Cross" must not be extracted from the longer NP)
- NP's must be unique (repetitions must be ignored)
- NP's must be provided one per line in a plain text file, with UTF-8 encoding.

# Follow-up

- BRUNO-A1:
  - 2,000 entries (around 4,000 UNLdots)
  - number of subcategorization frames > 15
  - number of paradigms > 15
- NC-A1:
  - 1,000 entries (3,000 UNLdots)

# This week

| USER | Balance* |
|---|---|
| A | USD 174.00 |
| B | USD 254.10 |
| C | USD 71.50 |
| D | USD 104.50 |
| E | USD 545.50 |
| F | USD 99.75 |
| G | USD 84.25 |
| H | USD 206.40 |
| I | USD 383.00 |

# Near future (2013)

- Projects (reopening in August 1st)
  - GD-A1 (former MIR-A1 (unl>nl dic))
  - ND-A1 (former MIR-A1 (nl dic))
- III UNL Olympiad (UC-A2)
  - CFP: August 1st, DEADLINE: November 1st
- XIII UNL School (Yerevan, Sep 2013)
  - CFP: August 1st
- XIV UNL School (Kuwait, Dec 2013)
  - CFP: October 1st

# Finally

# Thank you very much!