Ronaldo Martins
UNDL Foundation

Dictionaries

Dictionaries

- UNL-NL Dictionary (Analysis)
  - Enumerative (word forms)
    - [foot] [2883] "10028466 5" (NOU,STE,WRD,**SNG**,M1) <eng,0,0>;
      [feet] [2883] "10028466 5" (NOU,STE,WRD,**PLR**,M1) <eng,0,0>;
- NL-UNL Dictionary (Generation)
  - Generative (base forms)
    - [foot] [2883] "10028466 5" (NOU,STE,WRD,SNG,M1,**FLX(PLR:="feet";)**) <eng,0,0>;

Building dictionaries

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 : (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>

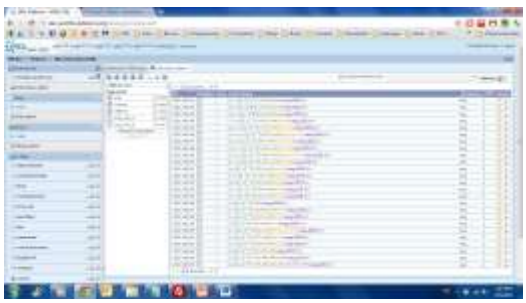
[NLW] {ID} "UW" (ATTR , …) < LG , FRE , PRI >; COMMENTS

- 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.

Examples

[book]{}"book" (N,NOU,SNG,M2)<eng,0,0>;
[books]{}"book" (N,NOU,PLR,M2)<eng,0,0>;
[car]{}"car" (N,NOU,SNG,M2)<eng,0,0>;
[cars]{}"car" (N,NOU,PLR,M2)<eng,0,0>;
[dr.]{}"dr"(N,ABB,Mo)<eng,0,0>;
[girl]{}"girl" (N,NOU,SNG,M2)<eng,0,0>;
[glass]{}"glass" (N,NOU,SNG,M2)<eng,0,0>;
[Bill]{}"Bill"(N,PPN,SNGT,Mo)<eng,0,0>;
[Edward]{}"Edward"(N,PPN,SNGT,Mo)<eng,0,0>;
[Geneva]{}"Geneva"(N,PPN,SNGT,Mo)<eng,0,0>;
[alone]{}"alone"(J,ADJ,Mo)<eng,0,0>;
[available]{}"available"(J,ADJ,Mo)<eng,0,0>;
[beautiful]{}"beautiful"(J,ADJ,Mo)<eng,0,0>;
[early]{}"early"(A,AAV,Mo)<eng,0,0>;
[never]{}"never"(A,AAV,Mo)<eng,0,0>;
[o'clock]{}"o'clock"(A,AAV,Mo)<eng,0,0>;
[arrive]{}"arrive"(V,VER,TSTI,INF,M17)<eng,0,0>;
[arrives]{}"arrive"(V,VER,TSTI,3PS,PRS,M17)<eng,0,0>;
[arrived]{}"arrive"(V,VER,TSTI,PAS,M17)<eng,0,0>;
[arrived]{}"arrive"(V,VER,TSTI,PTP,M17)<eng,0,0>;
[arriving]{}"arrive"(V,VER,TSTI,GER,M17)<eng,0,0>;

Using dictionaries (IAN and EUGENE)



Task #2: NL-UNL Dictionary

- Goal
  - To create entries for NL-UNL Dictionary for Reference Corpus
- Methodology
  - Create the corresponding entries in the dictionary format to all the words appearing in your wordlist:
    - Do not fill in the field between { } (but don't forget to include it)
    - Use the English word as the UW, when applicable
    - Define only the basic features (use the tagset to define the features)
    - The frequency and priority should be set to 0 (zero);
    - An example dictionary (of English) may be downloaded from the Wiki
  - Upload the dictionary to IAN (UNLWEB>UNLDEV>PROJECTS>IAN>DICTIONARY EDITOR)
  - Share it with the user "martins"
- DURATION: 4 hours