

7. Advanced queries

This series of tutorials is based upon work from COST Action
Multi3Generation CA18231, supported by COST
(European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation, cf. www.cost.eu

Enhancing NooJ grammars:

1. Lexical symbols

house | houses

be | being | am | is | are | was | were

Enhancing NooJ grammars:

1. Lexical symbols

house | houses = **<house>**

be | being | am | is | are | was | were = **<be>**

- Lexical symbols are lexical entries written between brackets.
- A lexical symbol matches all the forms that are associated with the lexical entry: variants, inflected forms, derived forms
- Lexical symbols refer to dictionaries: **<tables>** will not match anything in any text, because *tables* is not a valid lexical entry.

Lexical symbols: inflection + derivation

<demonstration>

=

demonstration | demonstrations |
demonstrate | demonstrates | demonstrated | demonstrating
| demonstrator | demonstrators |
demonstratable

... only works if *demonstration* is an entry of a dictionary, associated with an inflectional paradigm, and the derivational paradigms to produce the verb *demonstrer*, the noun *démonstrateur*, and the adjective *démonstrable*

Lexical symbols: inflection + derivation

<demonstration>

=

demonstration | demonstrations |
demonstrate | demonstrates | demonstrated | demonstrating
| demonstrator | demonstrators |
demonstratable

<king> = king | kings | queen | queens |
roi | reine | czar | tsar | tsarine |...

... only works if *king* is an entry of a dictionary and is associated with an inflectional paradigm, and the set of variants

Lexical symbols: inflection + derivation

<manifester>

=

manifeste | manifeste | manifeste | ... |
manifestation | manifestations |
manifestant | manifestante | manifestants |
manifestantes |
remanifeste | remanifeste | remanifestons | ... |
manifestable | manifestables |
manifestabilité | manifestabilités | ...

Lexical symbols

<читати> =

читати | читатиму | читатиме | читатимеш | читатимемо | читатимете | читатимуть |
прочитати | прочитав | прочитала | прочитало | прочитали | прочитаю | прочитаєш |
 прочитає | прочитаємо | прочитаєте | прочитають | прочитай | прочитайте |
 прочитамо | прочитай | прочитайте | прочитавши | прочитаний | прочитаного |
прочитаного | прочитаний | прочитаному | прочитаним | прочитаному | прочитанім
| прочитане | прочитаного | прочитаний | прочитаному | прочитаним | прочитаному
| прочитанім | прочитана | прочитаної | прочитану | прочитаній | прочитаною |
 прочитаній | прочитані | прочитаних | прочитаних | прочитані | прочитаним |
 прочитаними | прочитаних | читав | читала | читало | читали | читаю | читаєш |
 читає | читаємо | читаєте | читають | читай | читайте | читаймо | читаючи |
читаючий | читаючого | читаючого | читаючий | читаючому | читаючим | читаючому
| читаючім | читаюче | читаючого | читаючий | читаючому | читаючим | читаючому
| читаючім | читаюча | читаючої | читаючу | читаючій | читаючою | читаючій |
 читаючі | читаючих | читаючих | читаючі | читаючим | читаючими | читаючих |
читаний | читаного | читаного | читаний | читаному | читаним | читаному | читанім
| читане | читаного | читаний | читаному | читаним | читаному | читанім | читана |
читаної | читану | читаній | читаною | читаній | читані | читаних | читаних | читані |
 читаним | читаними | читаних

Enhancing NooJ grammars:

2. Syntactic symbols

in | on | from | to | inside | over ...

Enhancing NooJ grammars:

2. Syntactic symbols

<PREP> = in | on | from | to | inside | over ...

Enhancing NooJ grammars:

2. Syntactic symbols

<PREP> = in | on | from | to | inside | over ...

<N> (nouns), **<A>** (adjectives), **<V>** (verbs), **<DET>** (determiners), **<PREP>** (prepositions), **<CONJ>** (conjunctions)...

... only works if lexical entries are associated with these category (Part-Of-Speech) codes.

- Check the list of codes in the documentation of the dictionary you are using

Enhancing NooJ grammars:

2. Syntactic symbols

<PREP> = in | on | from | to | inside | over ...

<N>, <A>, <V>, <DET>, <PREP>, <CONJ>...

... Users are free to add dictionaries of terms and invent new categories, e.g.,

<MEDIC>, <POLITICS>, <POSITIVE>, <PSYVERB>...

3. Features and Properties

- Developers of dictionaries or grammars can invent any feature, e.g., **+PR**, **+Medic**

- Developers can also define properties in a special file: **properties.def**, e.g.,

N_Nb = sing | plur ;

(Nouns have property “Nb”; its potential values are “sing” or “plur”)

- Properties will be useful to check agreements, e.g.:

<\$N\$Nb=\$V\$Nb>

(The value of the noun property “Nb” must be equal to the value of the verb property “Nb”)

Enhancing NooJ grammars:

3. Features and Properties

- Lexical symbols, e.g. **<be+PR>**, **<cousin-s>**
- Syntactic symbols, e.g. **<N+Hum>**, **<V+INF>**,
<DET+p>
- Combining properties, e.g.
<N+Hum-f+p>, **<V+F-3+s>**
<cousin-Hum+p>, **<have-INF+s>**
<MEDICINE+Disease-Pneumo>

Regular grammars: recapitulation

<be> <PREP> <N>

<V+G> <ADV>* <PREP> <N+Abst>

(bank | credit | debit | cash) <card>

<DET> ((very|<E>) <A> | <E>) <A>* <N>

Exercise: A Lexical Field

- Looking for all occurrences of the theme *death* :

Exercise: A Lexical Field

- Looking for all occurrences of the theme *death* :

<assassinate> | <decease> | <die> | <eradicate> |
<exterminate> | <kill> | <obliterate> | <perish> |
<suicide> | <strangle> |

<corpse> | <death> | <extermination> | <fatality> | *funerals* |
grim reaper | <guillotine> | <massacre> | <mortality> |
<murder> | <slaughter> | <tomb> | <widower> |

fatal | *deadly* | *morbid* | *mortal* |

<kick> *the bucket* | <buy> *the farm* | <put> <WF>* *to death* |
<wipe> <WF>* *out*

<kick> the bucket | <buy> the farm | <put>
<WF>* to death | <wipe> <WF>* out





CONGRATULATIONS



You know how to write complex regular grammars and apply them to a text

