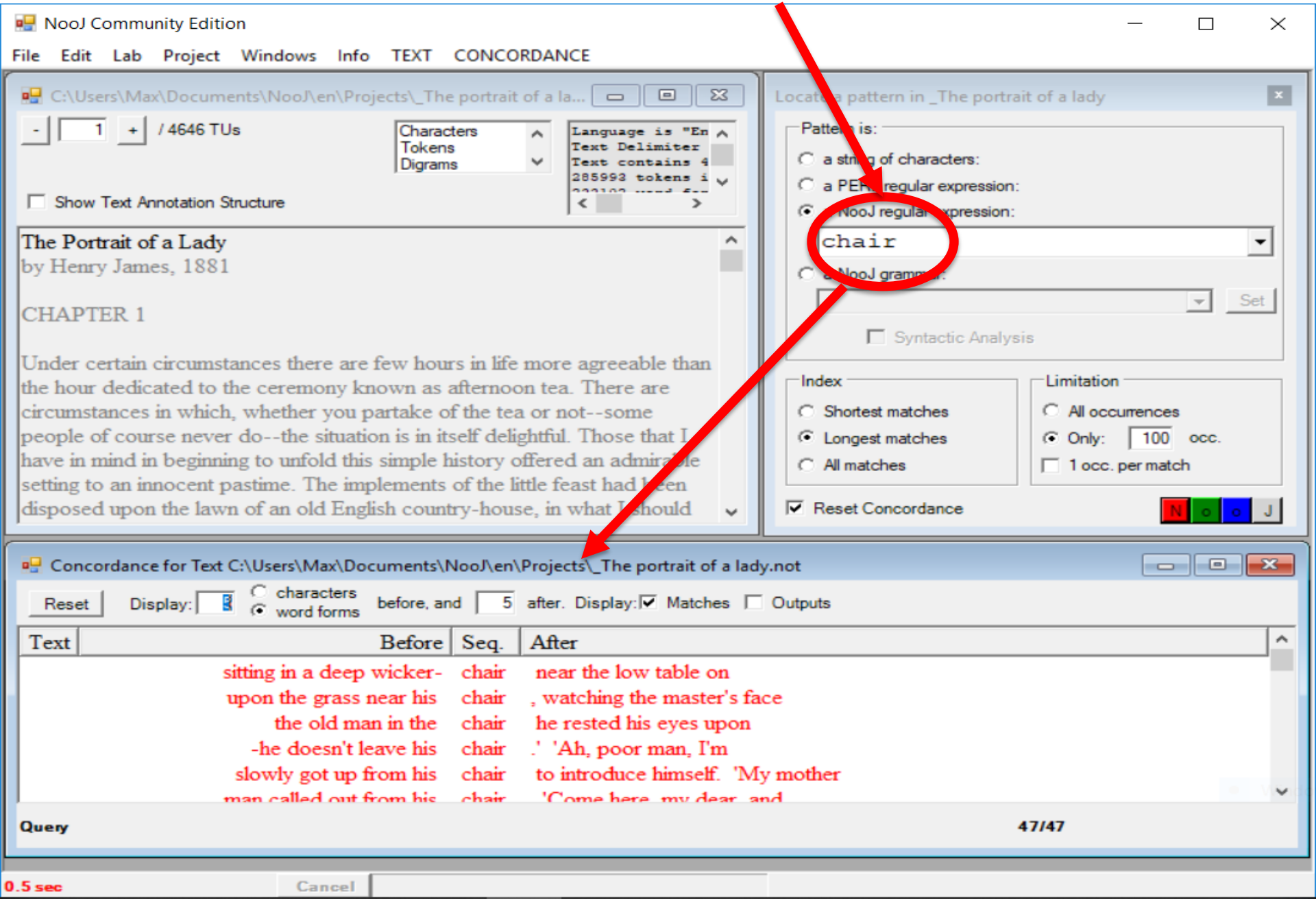
Funded by
the European Union

6. Queries, regular grammars and concordances

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

TEXT > Locate


NooJ Community Edition
File Edit Lab Project Windows Info TEXT CONCORDANCE
C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a la...
1
/ 4646 TUs
Characters
Tokens
Digrams
Show Text Annotation Structure
The Portrait of a Lady
by Henry James, 1881
CHAPTER 1
Under certain circumstances there are few hours in life more agreeable than the hour dedicated to the ceremony known as afternoon tea. There are circumstances in which, whether you partake of the tea or not--some people of course never do--the situation is in itself delightful. Those that I have in mind in beginning to unfold this simple history offered an admirable setting to an innocent pastime. The implements of the little feast had been disposed upon the lawn of an old English country-house, in what I should
Locate a pattern in _The portrait of a lady
Pattern is:
a string of characters:
a PERL regular expression:
NooJ regular expression:
chair
a NooJ grammar:
Set
Syntactic Analysis
Index
Shortest matches
Longest matches
All matches
Limitation
All occurrences
Only: 100 occ.
1 occ. per match
Reset Concordance
Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a lady.not
Reset
Display:
characters
word forms
before, and 5 after. Display: Matches Outputs
Text
Before
Seq.
After
sitting in a deep wicker- chair near the low table on
upon the grass near his chair , watching the master's face
the old man in the chair he rested his eyes upon
-he doesn't leave his chair .' 'Ah, poor man, I'm
slowly got up from his chair to introduce himself. 'My mother
Query
47/47
0.5 sec
Cancel

The Chomsky-Schützenberger hierarchy

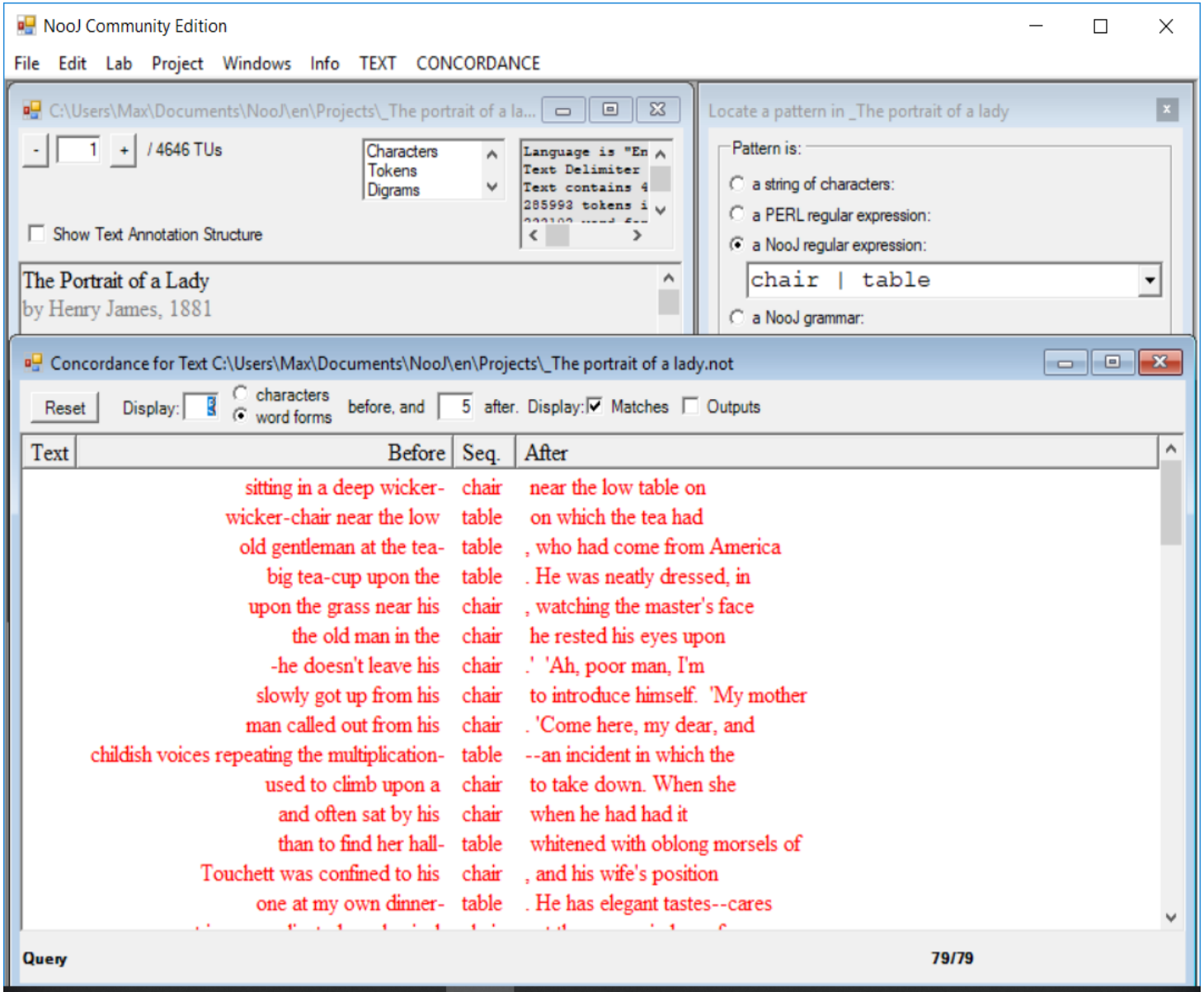
- A natural language comprises an infinite set of sentences
- How to describe this infinite set?
- Chomsky and Schützenberger designed a mathematical tool to describe languages: generative grammars.
- There are four types of generative grammars:
  - Regular grammars (Type 3)
  - Context-free grammars (Type 2)
  - Context-sensitive grammars (Type 1)
  - Unrestricted grammars (Type 0)
- Regular grammars are the least powerful;
  Unrestricted grammars are the most powerful

Regular Grammars

- Regular grammars are equivalent to Type 3 Generative Grammars in Chomsky-Schützenberger's hierarchy
- Regular grammars are the least powerful types of grammars, but are still very useful for the description of many linguistic phenomena, in particular morphology.
- Many computer tools understand regular grammars: awk, grep, lex, perl, sed, etc.
- There are also linguistic tools better adapted to describe morphological phenomena in natural languages, *e.g.*, XFST

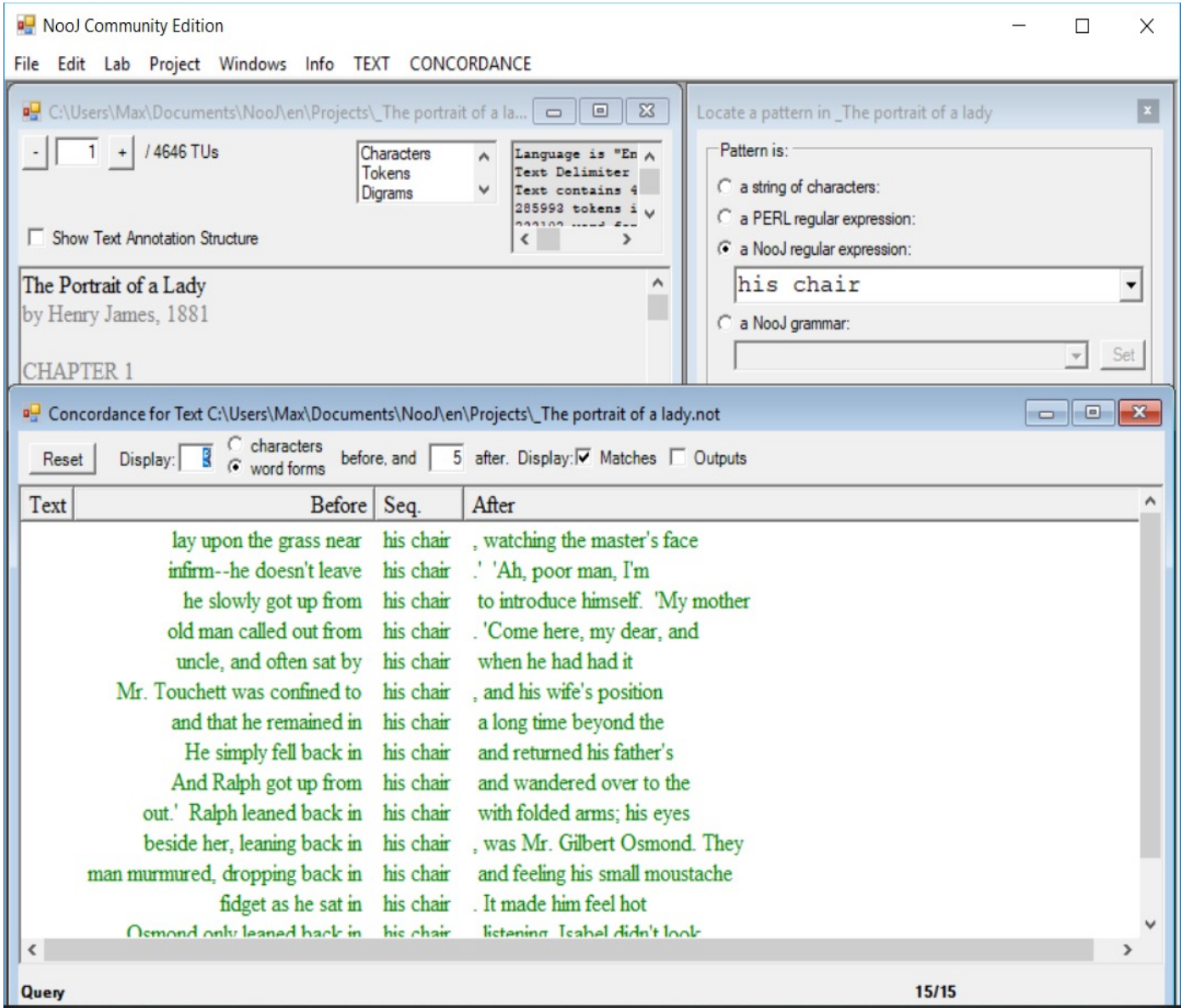
Regular grammars in NooJ

1. Disjunction, *e.g.*, **chair | table**



Regular grammars in NooJ

2. Concatenation, *e.g.*, **his chair**



Regular grammars in NooJ

3. Priority

Concatenations has priority over disjunctions

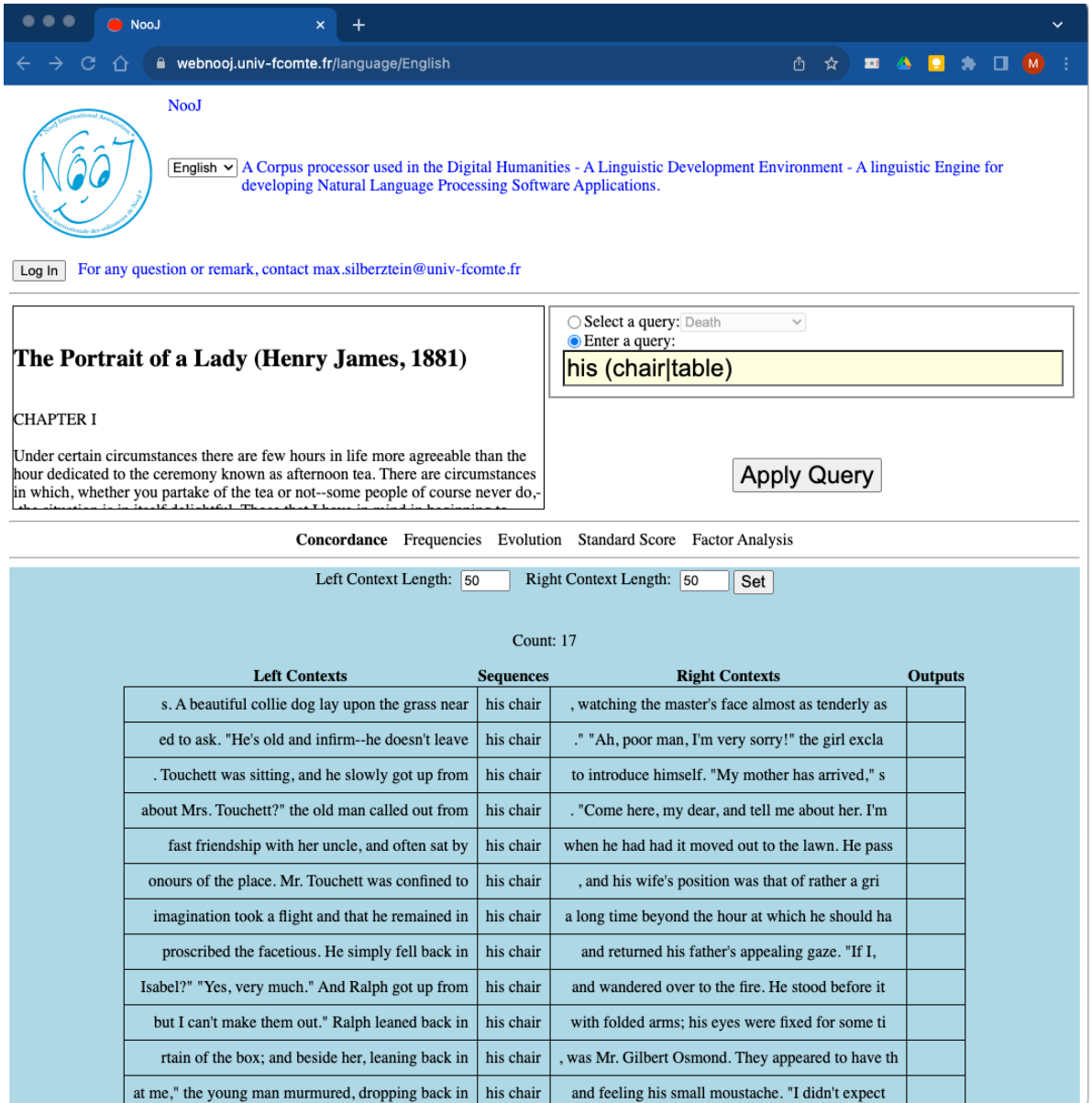
his chair | table

vs.

his (chair | table)

Regular grammars in NooJ

3. Priority, *e.g.*, his (chair|table)



Left Contexts	Sequences	Right Contexts	Outputs
s. A beautiful collie dog lay upon the grass near	his chair	, watching the master's face almost as tenderly as	
ed to ask. "He's old and infirm--he doesn't leave	his chair	." "Ah, poor man, I'm very sorry!" the girl excla	
. Touchett was sitting, and he slowly got up from	his chair	to introduce himself. "My mother has arrived," s	
about Mrs. Touchett?" the old man called out from	his chair	. "Come here, my dear, and tell me about her. I'm	
fast friendship with her uncle, and often sat by	his chair	when he had had it moved out to the lawn. He pass	
onours of the place. Mr. Touchett was confined to	his chair	, and his wife's position was that of rather a gri	
imagination took a flight and that he remained in	his chair	a long time beyond the hour at which he should ha	
proscribed the facetious. He simply fell back in	his chair	and returned his father's appealing gaze. "If I,	
Isabel?" "Yes, very much." And Ralph got up from	his chair	and wandered over to the fire. He stood before it	
but I can't make them out." Ralph leaned back in	his chair	with folded arms; his eyes were fixed for some ti	
rtain of the box; and beside her, leaning back in	his chair	, was Mr. Gilbert Osmond. They appeared to have th	
at me," the young man murmured, dropping back in	his chair	and feeling his small moustache. "I didn't expect	

Regular grammars in NooJ

4. Factorization

- Left factorization:

his chair | his table = his (chair | table)

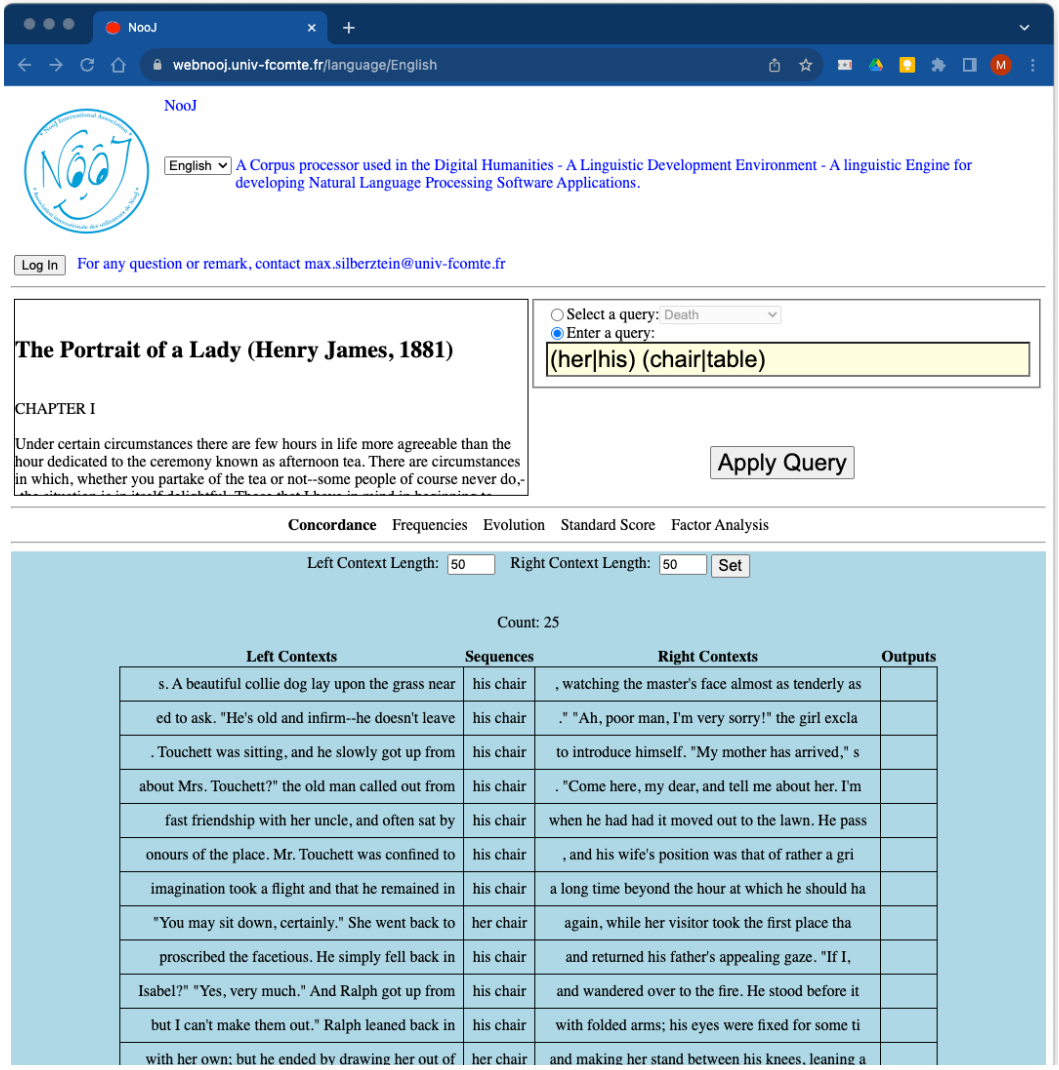
- Right factorization:

his chair | her chair = (his | her) chair

Regular grammars in NooJ

4. Factorization

(her|his) (chair|table)



Left Contexts	Sequences	Right Contexts	Outputs
s. A beautiful collie dog lay upon the grass near	his chair	, watching the master's face almost as tenderly as	
ed to ask. "He's old and infirm--he doesn't leave	his chair	." "Ah, poor man, I'm very sorry!" the girl excla	
. Touchett was sitting, and he slowly got up from	his chair	to introduce himself. "My mother has arrived," s	
about Mrs. Touchett?" the old man called out from	his chair	. "Come here, my dear, and tell me about her. I'm	
fast friendship with her uncle, and often sat by	his chair	when he had had it moved out to the lawn. He pass	
onours of the place. Mr. Touchett was confined to	his chair	, and his wife's position was that of rather a gri	
imagination took a flight and that he remained in	his chair	a long time beyond the hour at which he should ha	
"You may sit down, certainly." She went back to	her chair	again, while her visitor took the first place tha	
proscribed the facetious. He simply fell back in	his chair	and returned his father's appealing gaze. "If I,	
Isabel?" "Yes, very much." And Ralph got up from	his chair	and wandered over to the fire. He stood before it	
but I can't make them out." Ralph leaned back in	his chair	with folded arms; his eyes were fixed for some ti	
with her own; but he ended by drawing her out of	her chair	and making her stand between his knees, leaning a	

Regular grammars in NooJ

5. Neutral element for the concatenation

- ε, empty string, empty word, "", **<E>**

word <E> = word
<E> word = word

Many linguistic applications:

- "s" is the plural suffix, *e.g.*: table "s" = tables
- <E> is the singular suffix, *e.g.*: table <E> = table
- Optional linguistic units, *e.g.*: a (little | <E>) table

Regular grammars in NooJ

5. Neutral element for the concatenation

a lady | a young lady

Regular grammars in NooJ

5. Neutral element for the concatenation

a lady | a young lady
= a (lady | young lady)

Regular grammars in NooJ

5. Neutral element for the concatenation

a lady | a young lady
= a (lady | young lady)
= a (**<E>** lady | young lady)

Regular grammars in NooJ

5. Neutral element for the concatenation

a lady | a young lady
= a (lady | young lady)
= a (**<E>** lady | young lady)
= a ((**<E>** | young) lady)

Regular grammars in NooJ

5. Neutral element for the concatenation

a young lady | a lady

= a (young lady | lady)

= a (young lady | **<E>** lady)

= a ((young | **<E>**) lady)

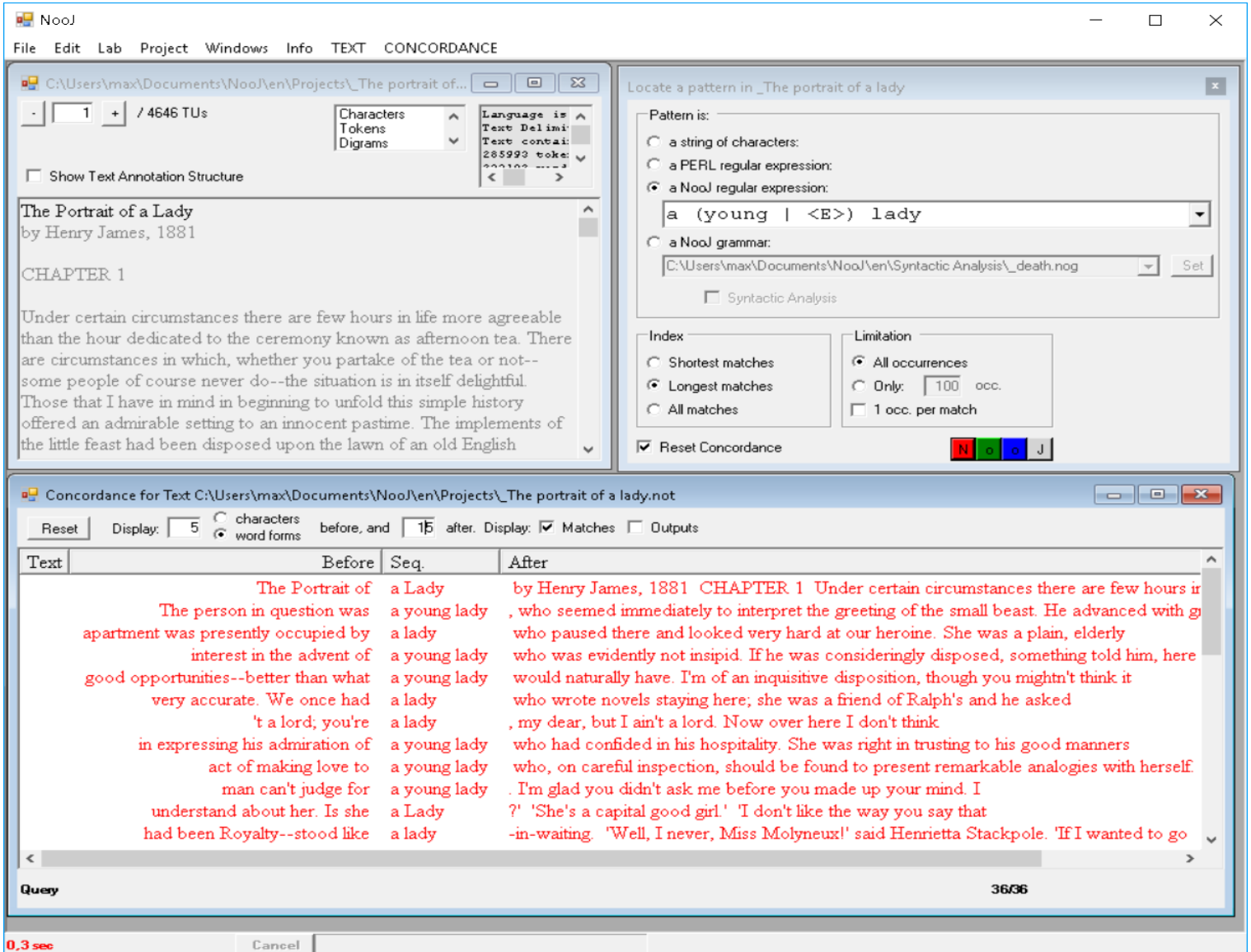
= a (young | **<E>**) lady

- To describe optional linguistic sequences:

(XXX | <E>)

Regular grammars in NooJ

5. Neutral element for the concatenation



Regular grammars in NooJ

6. the Kleene operator

a (<E> | very | very very | very very very | ...) pretty table

Regular grammars in NooJ

6. the Kleene operator *

a (<E> | very | very very | very very very | ...) pretty table

= a very* pretty table

Regular grammars in NooJ

6. the Kleene operator

Linguistic applications:

- A declarative sentence is constituted by a subject, a verb and any number of complements:

  Subject Verb Complement*

  Joe gave (Mary) (an apple) (in the train) (yesterday)...

- Present Perfect is constituted by the auxiliary verb *have*, potentially followed by an undefinite number of adverbs, followed by a verb in the past participle:

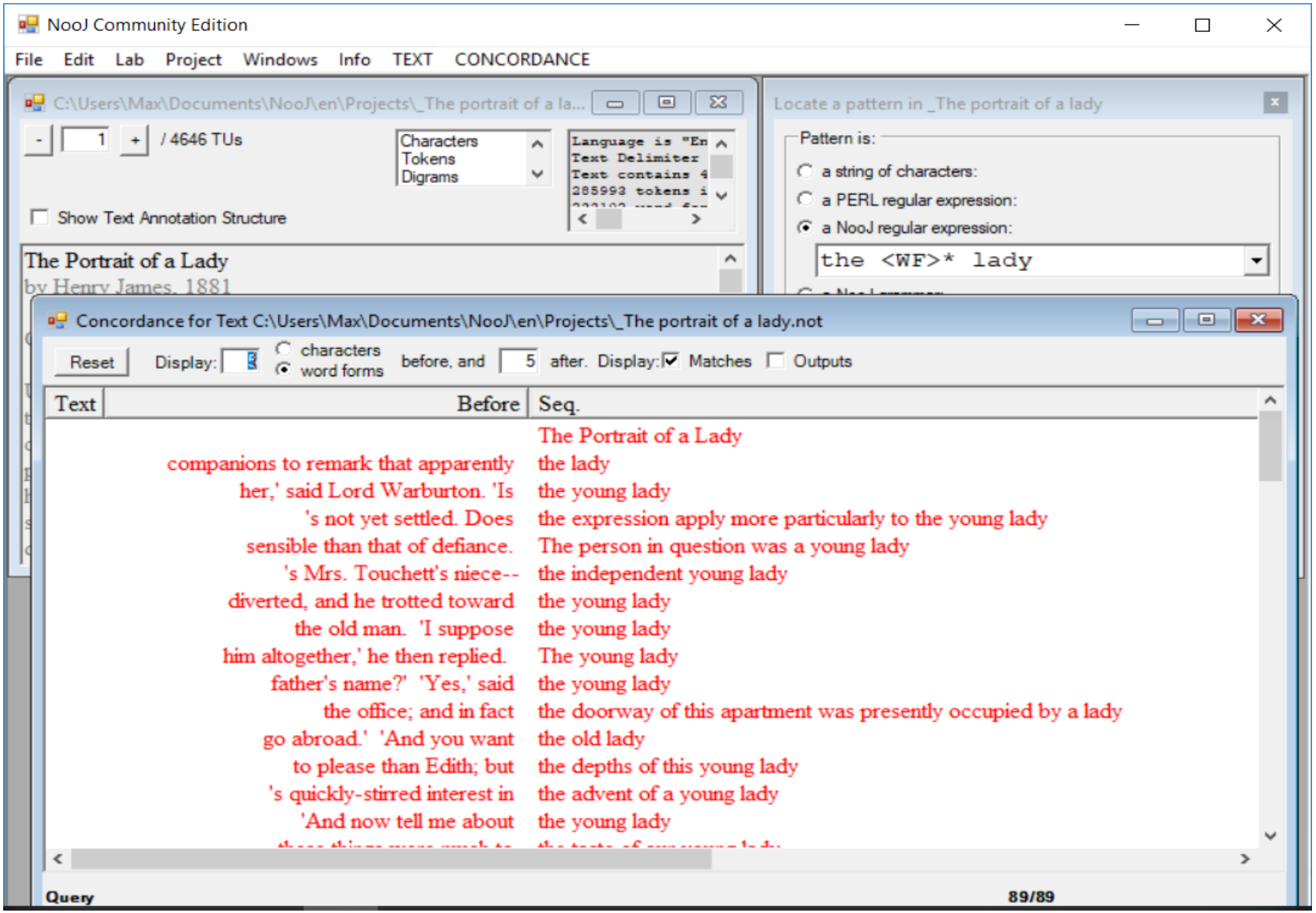
  have Adverb* Past-Participle

  It has (never) (before) done a Becket play

Regular grammars in NooJ

6. the Kleene operator

the <WF>* lady



Regular grammars in NooJ
Recapitulation

1. Disjunction, e.g., chair | table

2. Concatenation, e.g., his chair

3. Priority, e.g., his chair | table ≠ his chair | his table

4. Factorize, e.g., his chair | his table = his (chair | table)

5. Empty String <E>, e.g., table <E> = table

6. Kleene Operator, e.g., a very* pretty house

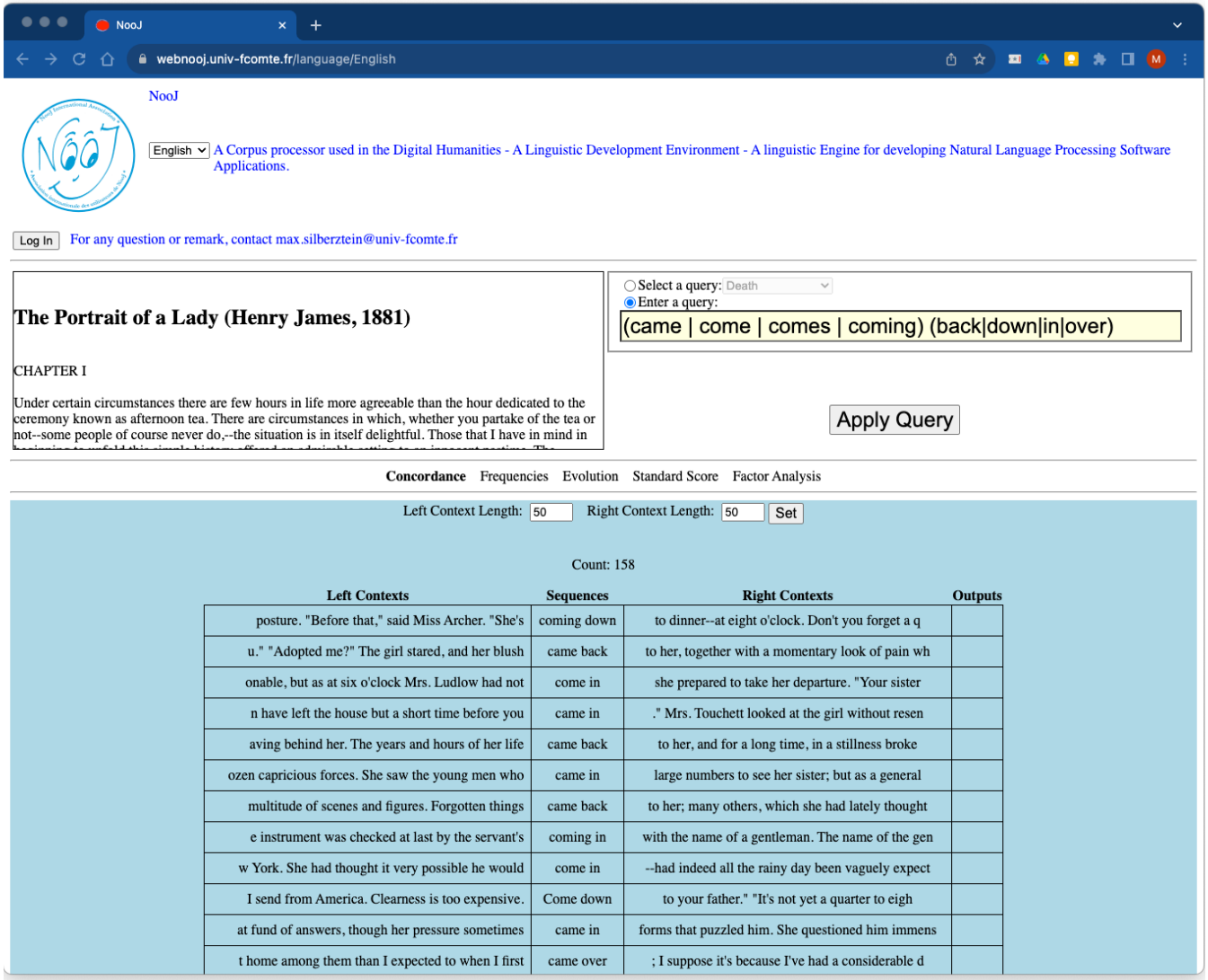
Exercise

Find occurrences of the verb *to come* followed by: *back*, *down*, *in*, *over*

Exercise

Find occurrences of the verb *to come* followed by: *back*, *down*, *in*, *over*

(came | come | comes | coming) (back|down|in|over)



Left Contexts	Sequences	Right Contexts	Outputs
posture. "Before that," said Miss Archer. "She's	coming down	to dinner--at eight o'clock. Don't you forget a q	
u." "Adopted me?" The girl stared, and her blush	came back	to her, together with a momentary look of pain wh	
onable, but as at six o'clock Mrs. Ludlow had not	come in	she prepared to take her departure. "Your sister	
n have left the house but a short time before you	came in	." Mrs. Touchett looked at the girl without resen	
aving behind her. The years and hours of her life	came back	to her, and for a long time, in a stillness broke	
ozen capricious forces. She saw the young men who	came in	large numbers to see her sister; but as a general	
multitude of scenes and figures. Forgotten things	came back	to her; many others, which she had lately thought	
e instrument was checked at last by the servant's	coming in	with the name of a gentleman. The name of the gen	
w York. She had thought it very possible he would	come in	--had indeed all the rainy day been vaguely expect	
I send from America. Clearness is too expensive.	Come down	to your father." "It's not yet a quarter to eigh	
at fund of answers, though her pressure sometimes	came in	forms that puzzled him. She questioned him immens	
t home among them than I expected to when I first	came over	; I suppose it's because I've had a considerable d	





CONGRATULATIONS

You know how to write simple queries and apply them to a text or a corpus