

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

The screenshot shows the NooJ Community Edition interface. The main window displays the text of 'The Portrait of a Lady' by Henry James. A red arrow points from the 'TEXT > Locate' menu to the 'Locate a pattern in' dialog box. In this dialog, the 'Pattern is' section has 'a NooJ regular expression' selected, and the pattern 'chair' is entered in the text field. Another red arrow points from the 'chair' text field to the concordance table below. The concordance table shows the results of the search, with the word 'chair' highlighted in red in the original image. The table has columns for 'Text', 'Before', 'Seq.', and 'After'.

Locate a pattern in _The portrait of a lady

Pattern is:

- ☐ a string of characters:
- ☐ a Perl regular expression:
- ☒ a NooJ regular expression:
- ☐ a NooJ grammar:

chair

☐ 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 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	
man called out from his	chair	'Come here, my dear, and	

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**

The screenshot displays the NooJ Community Edition interface. The main window shows the text 'The Portrait of a Lady' by Henry James, 1881. A search window titled 'Locate a pattern in _The portrait of a lady' is open, showing the pattern 'chair | table' selected under 'a NooJ regular expression:'. Below this, a concordance window titled 'Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a lady.not' is open. It shows a table with columns 'Text', 'Before', 'Seq.', and 'After'. The table lists occurrences of the words 'chair' and 'table' in the text, with the words highlighted in red. The status bar at the bottom shows 'Query' and '79/79'.

NooJ Community Edition

File Edit Lab Project Windows Info TEXT CONCORDANCE

C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a la... / 4646 TUs

Show Text Annotation Structure

The Portrait of a Lady
by Henry James, 1881

Locate a pattern in _The portrait of a lady

Pattern is:

☐ a string of characters:

☐ a PERL regular expression:

☒ a NooJ regular expression:

chair | table

☐ a NooJ grammar:

Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a lady.not

Reset Display: ☐ characters before, and ☐ 5 after. Display: ☒ Matches ☐ Outputs

☒ word forms

Text	Before	Seq.	After
sitting in a deep wicker-	chair	near the low table on	
wicker-chair near the low	table	on which the tea had	
old gentleman at the tea-	table	, who had come from America	
big tea-cup upon the	table	. He was neatly dressed, in	
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	
man called out from his	chair	. 'Come here, my dear, and	
childish voices repeating the multiplication-	table	--an incident in which the	
used to climb upon a	chair	to take down. When she	
and often sat by his	chair	when he had had it	
than to find her hall-	table	whitened with oblong morsels of	
Touchett was confined to his	chair	, and his wife's position	
one at my own dinner-	table	. He has elegant tastes--cares	

Query 79/79

Regular grammars in NooJ

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

The screenshot displays the NooJ Community Edition interface. The top window shows the file path and search options. The bottom window shows the concordance results for the pattern 'his chair'.

Top Window: Locate a pattern in _The portrait of a lady

- Language is "En"
- Text Delimiter: " "
- Text contains 4 tokens
- 285993 tokens in 40000 words
- Pattern is:
- Options: ☐ a string of characters, ☐ a PERL regular expression, ☒ a NooJ regular expression, ☐ a NooJ grammar.

Bottom Window: Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a lady.not

Reset Display: ☒ characters before, and 5 after. Display: ☒ Matches ☐ Outputs

Text	Before	Seq.	After
lay upon the grass near	his chair	, watching the master's face	
infirm--he doesn't leave	his chair	. 'Ah, poor man, I'm	
he slowly got up from	his chair	to introduce himself. 'My mother	
old man called out from	his chair	. 'Come here, my dear, and	
uncle, and often sat by	his chair	when he had had it	
Mr. Touchett was confined to	his chair	, and his wife's position	
and that he remained in	his chair	a long time beyond the	
He simply fell back in	his chair	and returned his father's	
And Ralph got up from	his chair	and wandered over to the	
out.' Ralph leaned back in	his chair	with folded arms; his eyes	
beside her, leaning back in	his chair	, was Mr. Gilbert Osmond. They	
man murmured, dropping back in	his chair	and feeling his small moustache	
fidget as he sat in	his chair	. It made him feel hot	
Osmond only leaned back in	his chair	listening. Isabel didn't look	

Query 15/15

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)**



NooJ

English A Corpus processor used in the Digital Humanities - A Linguistic Development Environment - A linguistic Engine for developing Natural Language Processing Software Applications.

[Log In](#) For any question or remark, contact max.silberstein@univ-fcomte.fr

The Portrait of a Lady (Henry James, 1881)

CHAPTER I

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 itself delightful. There that I have intended to begin to

Select a query: Death

Enter a query:

his (chair|table)

Apply Query

Concordance Frequencies Evolution Standard Score Factor Analysis

Left Context Length: 50 Right Context Length: 50 Set

Count: 17

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 excl	
. 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

(her|his) (chair|table)



NooJ

English

A Corpus processor used in the Digital Humanities - A Linguistic Development Environment - A linguistic Engine for developing Natural Language Processing Software Applications.

Log In

For any question or remark, contact max.silberstein@univ-fcomte.fr

The Portrait of a Lady (Henry James, 1881)

CHAPTER I

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--

Select a query: Death

Enter a query:

(her|his) (chair|table)

Apply Query

Concordance Frequencies Evolution Standard Score Factor Analysis

Left Context Length: 50 Right Context Length: 50 Set

Count: 25

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, "", $\langle E \rangle$

word $\langle E \rangle$ = word

$\langle E \rangle$ word = word

Many linguistic applications:

- “s” is the plural suffix, e.g.: table “s” = tables
- $\langle E \rangle$ is the singular suffix, e.g.: table $\langle E \rangle$ = table
- Optional linguistic units, e.g.: a (little | $\langle E \rangle$) 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

The screenshot displays the NooJ software interface with the following components:

- Main Window:** Shows the text "The Portrait of a Lady" by Henry James, 1881, and "CHAPTER 1". The text is displayed in a large font.
- Search Panel (Right):** Titled "Locate a pattern in _The portrait of a lady". It contains a "Pattern is:" section with three radio buttons: "a string of characters:", "a PERL regular expression:", and "a NooJ regular expression:". The "a NooJ regular expression:" option is selected, and the pattern "a (young | <E>) lady" is entered in the text field. Below this, there is a dropdown menu for "a NooJ grammar:" set to "C:\Users\max\Documents\NooJ\en\Syntactic Analysis_death.nog" and a "Set" button. A "Syntactic Analysis" checkbox is also present.
- Search Options (Bottom Right):** Includes an "Index" section with three radio buttons: "Shortest matches", "Longest matches" (selected), and "All matches". A "Limitation" section has two radio buttons: "All occurrences" (selected) and "Only: 100 occ." (with a "100" input field). A "1 occ. per match" checkbox is also present. A "Reset Concordance" checkbox is checked.
- Concordance Window (Bottom):** Titled "Concordance for Text C:\Users\max\Documents\NooJ\en\Projects_The portrait of a lady.not". It has a "Reset" button, a "Display:" section with a "5" input field and a "characters word forms" checkbox, and a "before, and 15 after. Display: [x] Matches [] Outputs" section. The concordance table has four columns: "Text", "Before", "Seq.", and "After". The "Text" column contains the original text, and the "Seq." column contains the sequence of words matched by the pattern. The "After" column contains the text following the match.

Text	Before	Seq.	After
The Portrait of		a Lady	by Henry James, 1881 CHAPTER 1 Under certain circumstances there are few hours in
The person in question was		a young lady	who seemed immediately to interpret the greeting of the small beast. He advanced with g
apartment was presently occupied by		a lady	who paused there and looked very hard at our heroine. She was a plain, elderly
interest in the advent of		a young lady	who was evidently not insipid. If he was considerably disposed, something told him, here
good opportunities--better than what		a young lady	would naturally have. I'm of an inquisitive disposition, though you mightn't think it
very accurate. We once had		a lady	who wrote novels staying here; she was a friend of Ralph's and he asked
't a lord; you're		a lady	, my dear, but I ain't a lord. Now over here I don't think
in expressing his admiration of		a young lady	who had confided in his hospitality. She was right in trusting to his good manners
act of making love to		a young lady	who, on careful inspection, should be found to present remarkable analogies with herself.
man can't judge for		a young lady	. I'm glad you didn't ask me before you made up your mind. I
understand about her. Is she		a Lady	? 'She's a capital good girl.' I don't like the way you say that
had been Royalty--stood like		a lady	-in-waiting. 'Well, I never, Miss Molyneux!' said Henrietta Stackpole. 'If I wanted to go

Query 36/36

0.3 sec Cancel

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 indefinite 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

The screenshot displays the NooJ Community Edition interface. The main window shows the text 'The Portrait of a Lady' by Henry James, 1881. A search window titled 'Locate a pattern in _The portrait of a lady' is open, showing the pattern 'the <WF>* lady'. A concordance window titled 'Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a lady.not' is also open, showing the results of the search. The concordance table has columns 'Text', 'Before', and 'Seq.'. The 'Text' column contains the original text, and the 'Seq.' column contains the extracted phrases. The 'Before' column is empty. The 'Seq.' column contains the following phrases: 'The Portrait of a Lady', 'the lady', 'the young lady', 'the expression apply more particularly to the young lady', 'The person in question was a young lady', 'the independent young lady', 'the young lady', 'the young lady', 'The young lady', 'the young lady', 'the doorway of this apartment was presently occupied by a lady', 'the old lady', 'the depths of this young lady', 'the advent of a young lady', 'the young lady', and 'the taste of my own lady'.

NooJ Community Edition

File Edit Lab Project Windows Info TEXT CONCORDANCE

C:\Users\Max\Documents\NooJ\en\Projects_The portrait of a la... / 4646 TUs

Characters Tokens Digrams

Language is "En" Text Delimiter Text contains 4 285993 tokens i 333100 words f...

Show Text Annotation Structure

The Portrait of a Lady by Henry James, 1881

Locate a pattern in _The portrait of a lady

Pattern is:

☐ a string of characters:

☐ a PERL regular expression:

☒ a NooJ regular expression:

the <WF>* lady

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.
companions to remark that apparently		The Portrait of a Lady
her,' said Lord Warburton. 'Is		the lady
's not yet settled. Does		the young lady
sensible than that of defiance.		the expression apply more particularly to the young lady
's Mrs. Touchett's niece--		The person in question was a young lady
diverted, and he trotted toward		the independent young lady
the old man. 'I suppose		the young lady
him altogether,' he then replied.		the young lady
father's name?' 'Yes,' said		The young lady
the office; and in fact		the young lady
go abroad.' 'And you want		the doorway of this apartment was presently occupied by a lady
to please than Edith; but		the old lady
's quickly-stirred interest in		the depths of this young lady
'And now tell me about		the advent of a young lady
these things were much to		the young lady
		the taste of my own lady

Query 89/89

Regular grammars in NooJ

Recapitulation

1. Disjunction, e.g., **chair | table**
2. Concatenation, e.g., **his chair**
3. Priority, e.g., **his chair | table** \neq **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)

NooJ

webnoo.univ-fcomte.fr/language/English



English ▾ A Corpus processor used in the Digital Humanities - A Linguistic Development Environment - A linguistic Engine for developing Natural Language Processing Software Applications.

Log In

For any question or remark, contact max.silberstein@univ-fcomte.fr

The Portrait of a Lady (Henry James, 1881)

CHAPTER I

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

Select a query: Death ▾

Enter a query:

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

Apply Query

Concordance Frequencies Evolution Standard Score Factor Analysis

Left Context Length: 50 Right Context Length: 50 Set

Count: 158

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 resem	
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

