



12. Graphical Grammars

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

Graph Editor

- A grammar is represented by a set of **graphs**
- A graph is a set of **nodes**
- Two special nodes: the **initial node** and the **terminal node**
- Nodes are **labeled**
- Nodes can be connected
- Connections are **oriented**

File > New > Grammar > Graphical Editor > Syntax

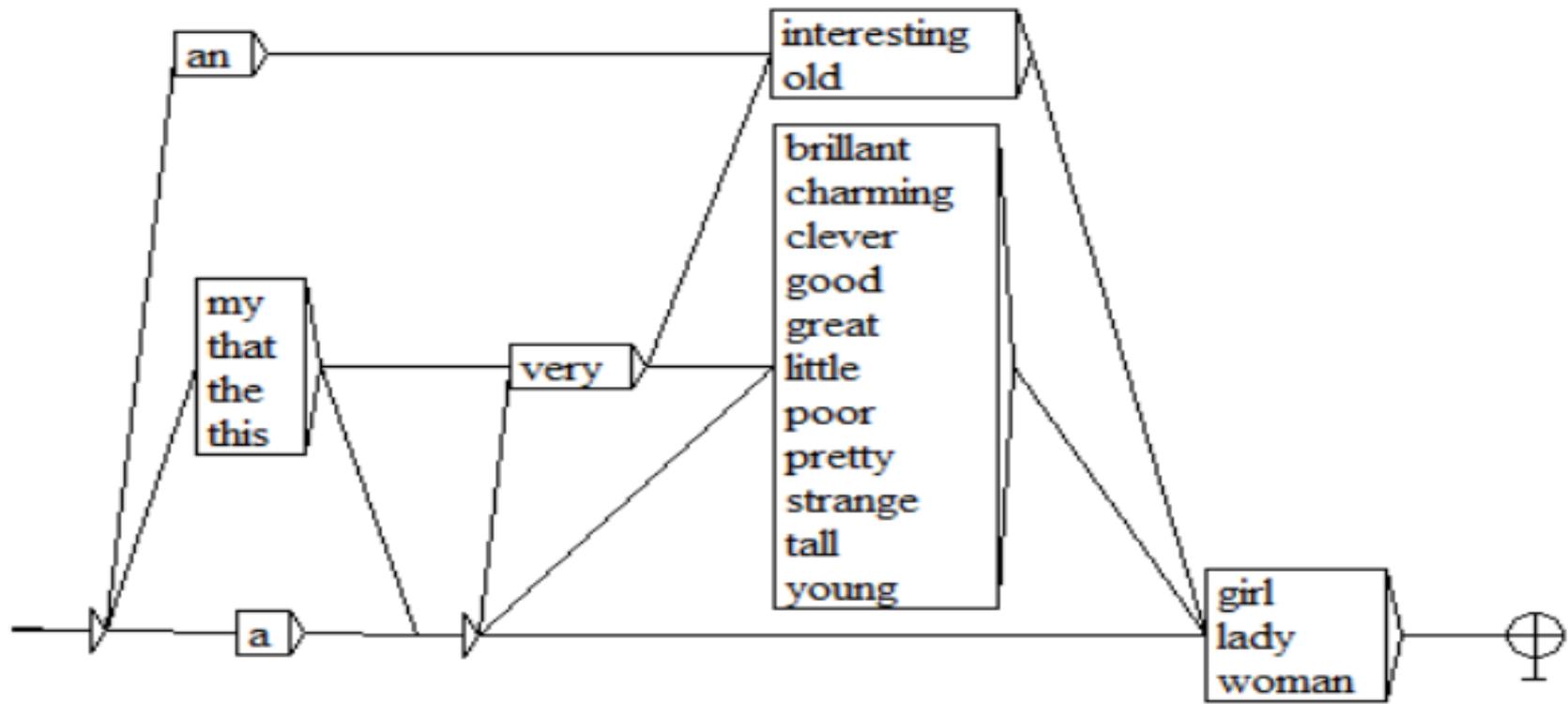
The screenshot shows the NooJ Community Edition interface. The title bar reads "NooJ Community Edition - [C:\Users\Max\Documents\...]". The menu bar includes "File", "Edit", "Lab", "Project", "Windows", "Info", "TEXT", and "GRAMMAR". The main window title is "English/English syntactic grammar." The central area displays a syntactic tree diagram for the sentence "a young lady". The root node is a circle with a cross, connected to a triangle node labeled "a". This "a" node branches into two triangle nodes: "young" and "lady". The "lady" node is connected to another circle node with a cross. A status bar at the bottom left shows "0.5 sec" in red, and a "Cancel" button is visible.

TEXT > Locate > NooJ Grammar

Graph Editor

- Ctrl-Click in empty space to create a node
- Ctrl-Click in a node to start editing its label
- Ctrl-Enter to validate the edition of a label
- **WARNING: do not double-click!**
- Click two nodes to connect them
- **WARNING: Connections are oriented!**
- Click two nodes to disconnect them
- **UNDO (Ctrl-Z) and REDO (Ctrl-Y)**

English/English syntactic grammar.



0.7 sec

Cancel

C:\Users\Max\Documents\Noo\en\Projects_The portrait of...

4086 / 4646 TUs

Characters Tokens Digrams

Language is Text Delimi Text contai 285992 toke

Show Text Annotation Structure

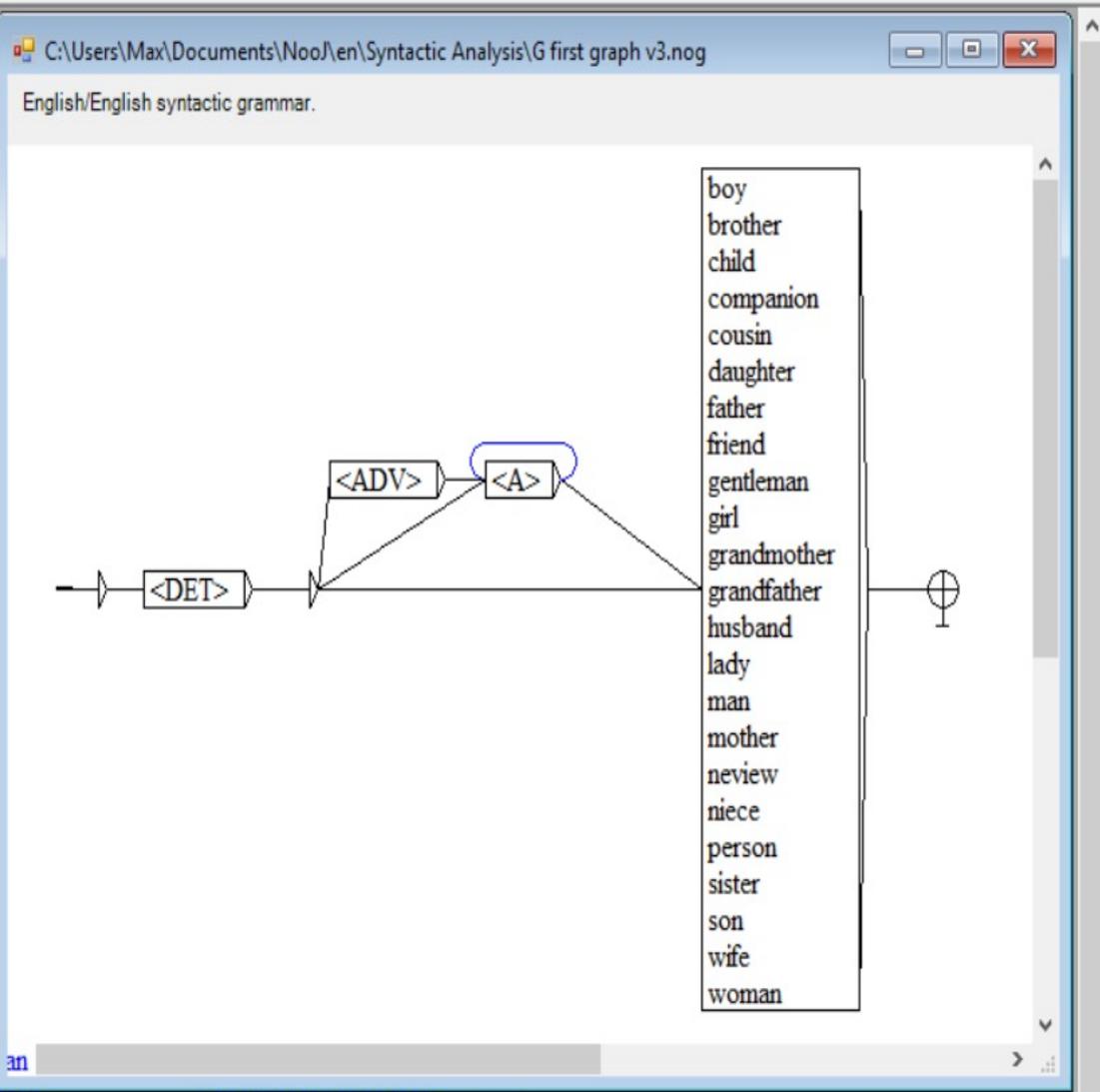
Isabel sat there looking up at her, without rising; her face was almost a

Concordance for Text C:\Users\Max\Documents\Noo\en\Projects_The portrait of a la

Reset Display: 5 characters before, and 5 after. Display: Matches

Text	Before	Seq.
conclusion. She was provided with	a trusty companion	
his splendid range of suggestion,	a very agreeable m	
Touchett answered. 'Lord Warburton's	a very amiable you	
about you.' Isabel thought her	a very attractive pe	
unhappy as she-I've	a very bad husban	
his hand. His cousin was	a very brilliant girl	
briefly silent. 'I call Henrietta	a very brilliant wor	
you're mistaken. He's	a very clever man	
a thoroughly grand man and	a very dear friend	
worthy of Isabel.' 'You're	a very devoted frie	
now, in her own eyes,	a very different pe	
had seen hitherto, but was	a very different pe	
of these impressions. 'He's	a very fine man	
a very amiable young man--	a very fine young r an	
life and she had been	a very fortunate person	
her in. She has been	a very good friend	

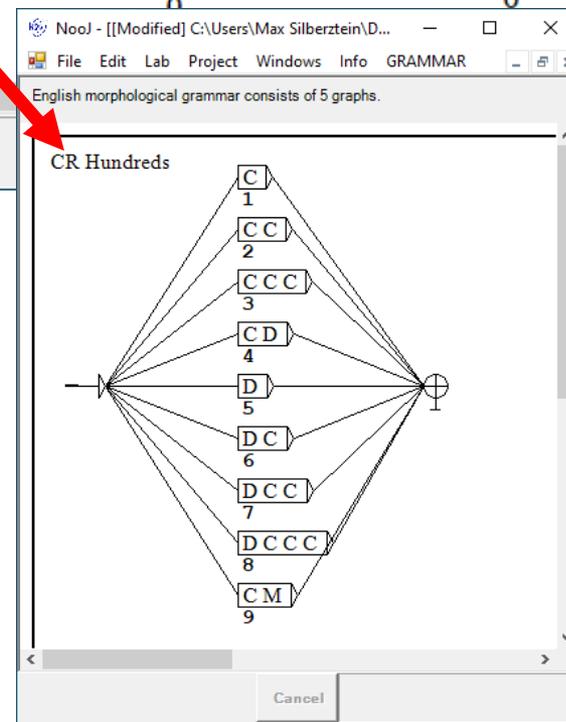
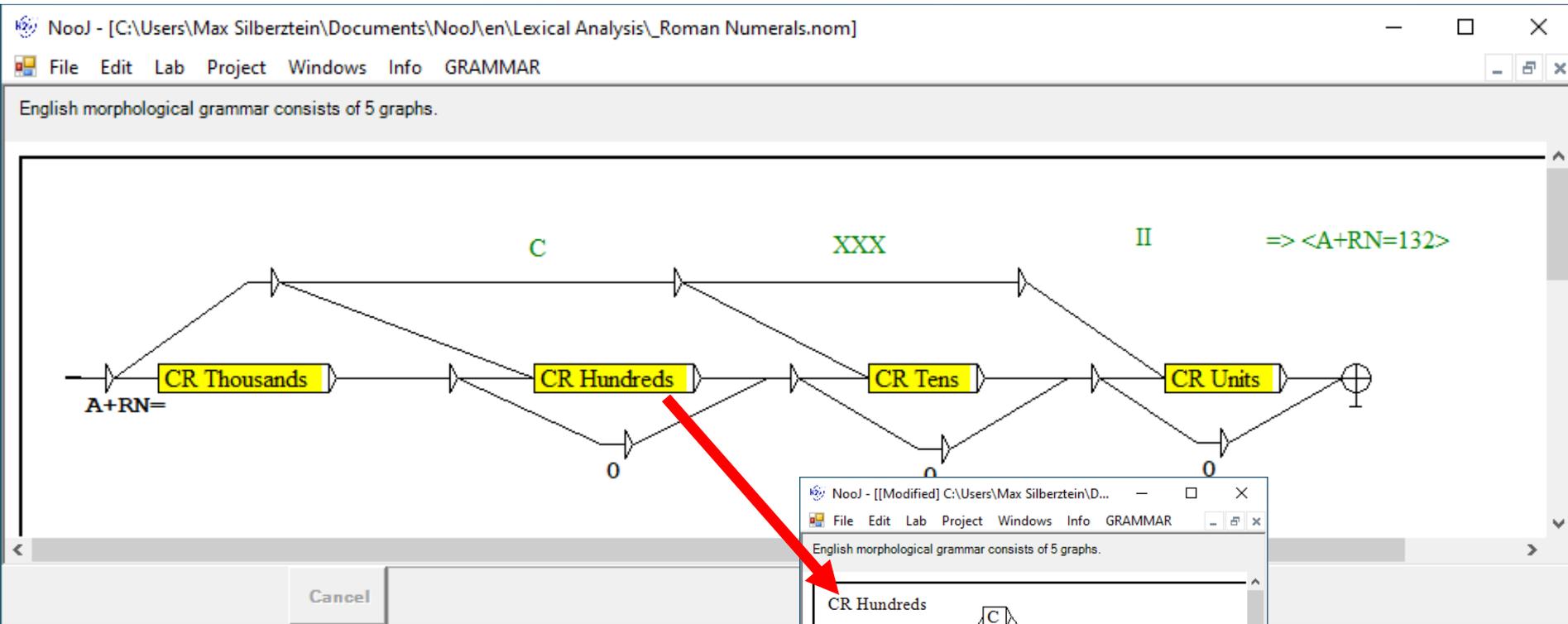
Query 2378/2378



Regular and Context-Free grammars

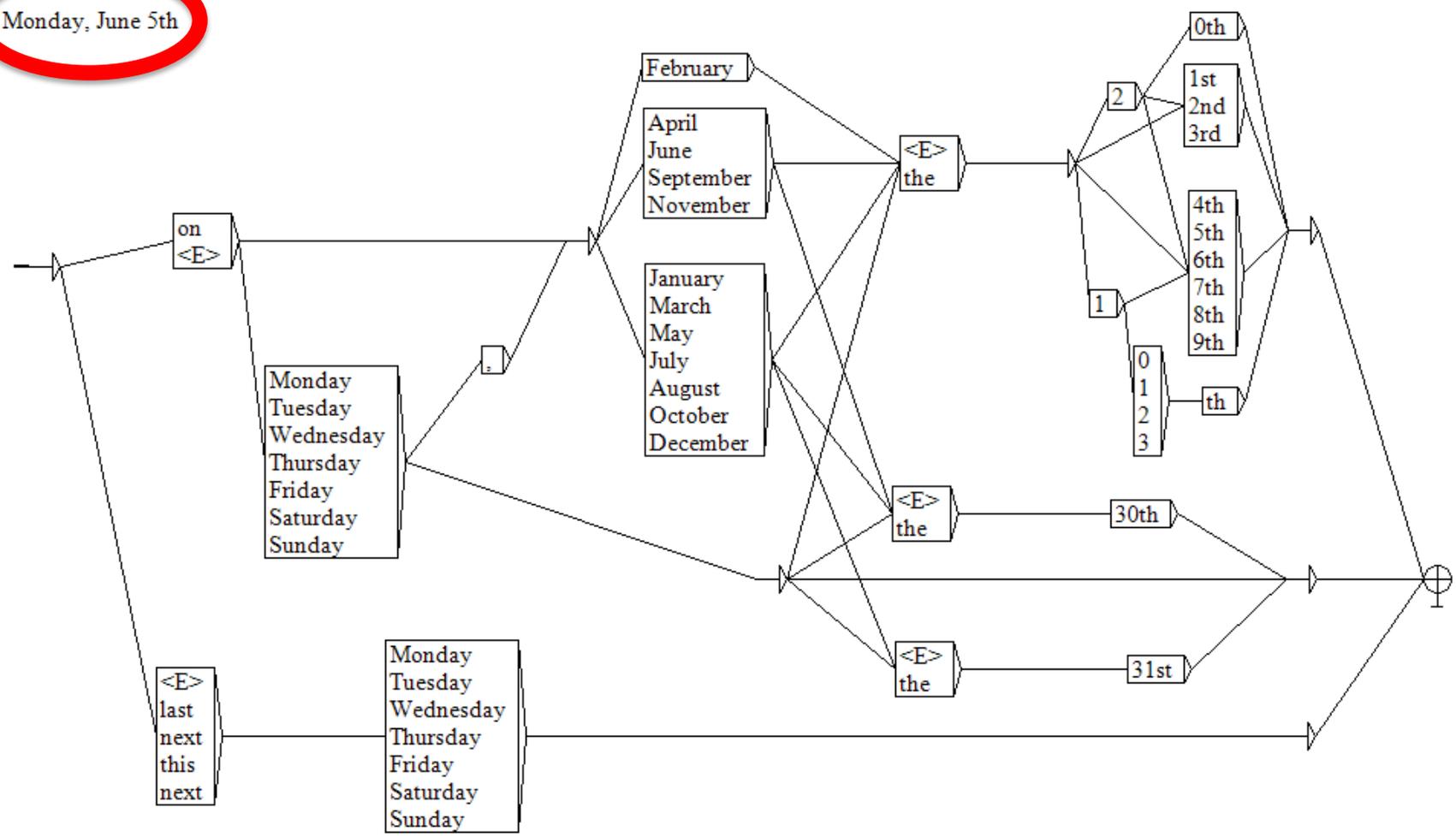
- A graphical grammar that contains only one graph is equivalent to a regular grammar
- A graphical grammar that contains multiple graphs is equivalent to a context-free grammar
- The main graph is named **Main**
- To create a new graph: GRAMMAR > New Graph
- To delete current graph: GRAMMAR > Delete Graph
- To display the structure of a grammar: GRAMMAR > Show Structure
- To refer to an embedded graph, add prefix “:” in front of its name in a node, e.g., **:NP**

Enhanced Context-Free Grammars



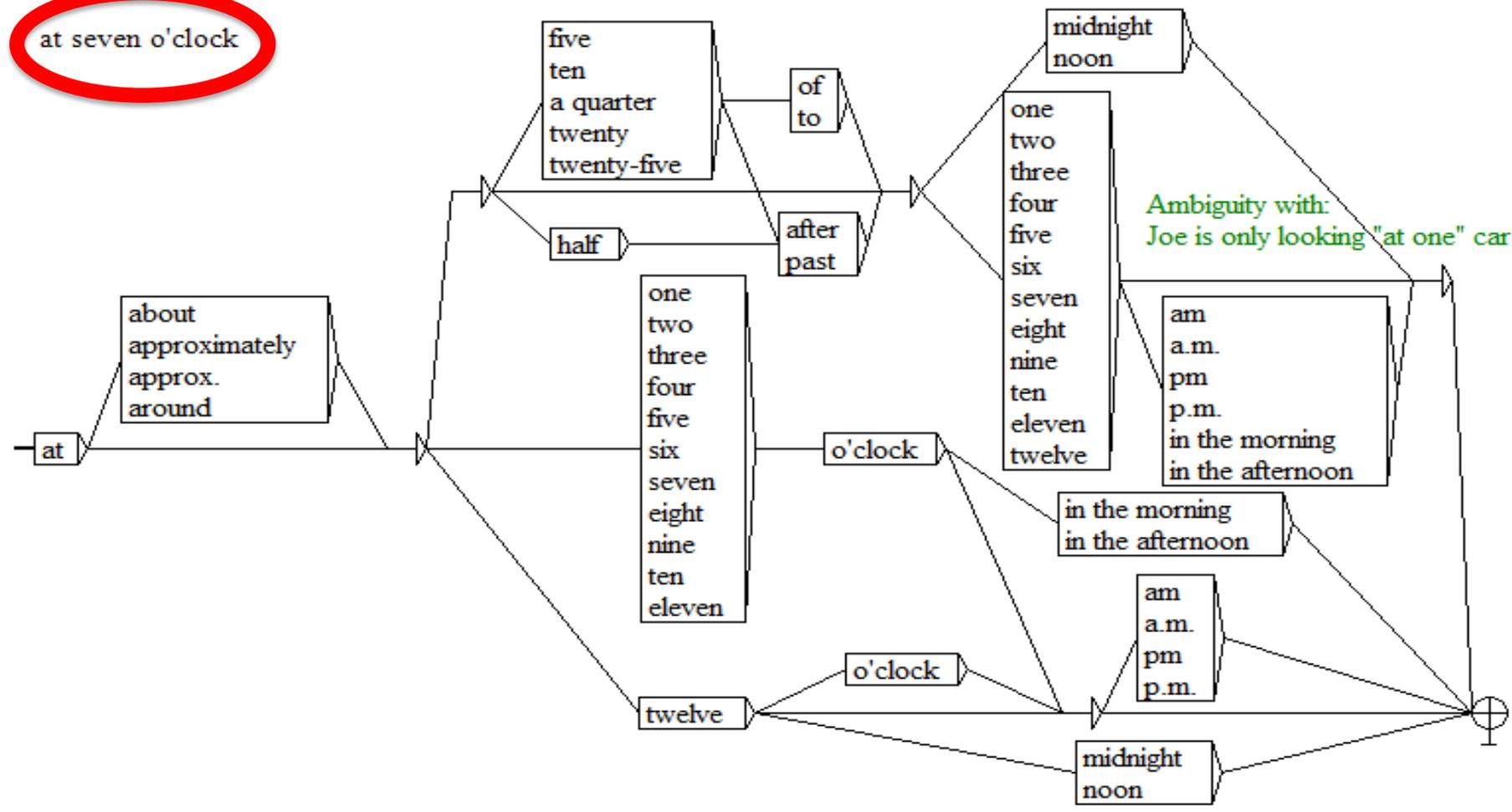
English (United States)/English (United States) syntactic grammar consists of 20 graphs.

Monday, June 5th



English /English syntactic grammar consists of 20 graphs.

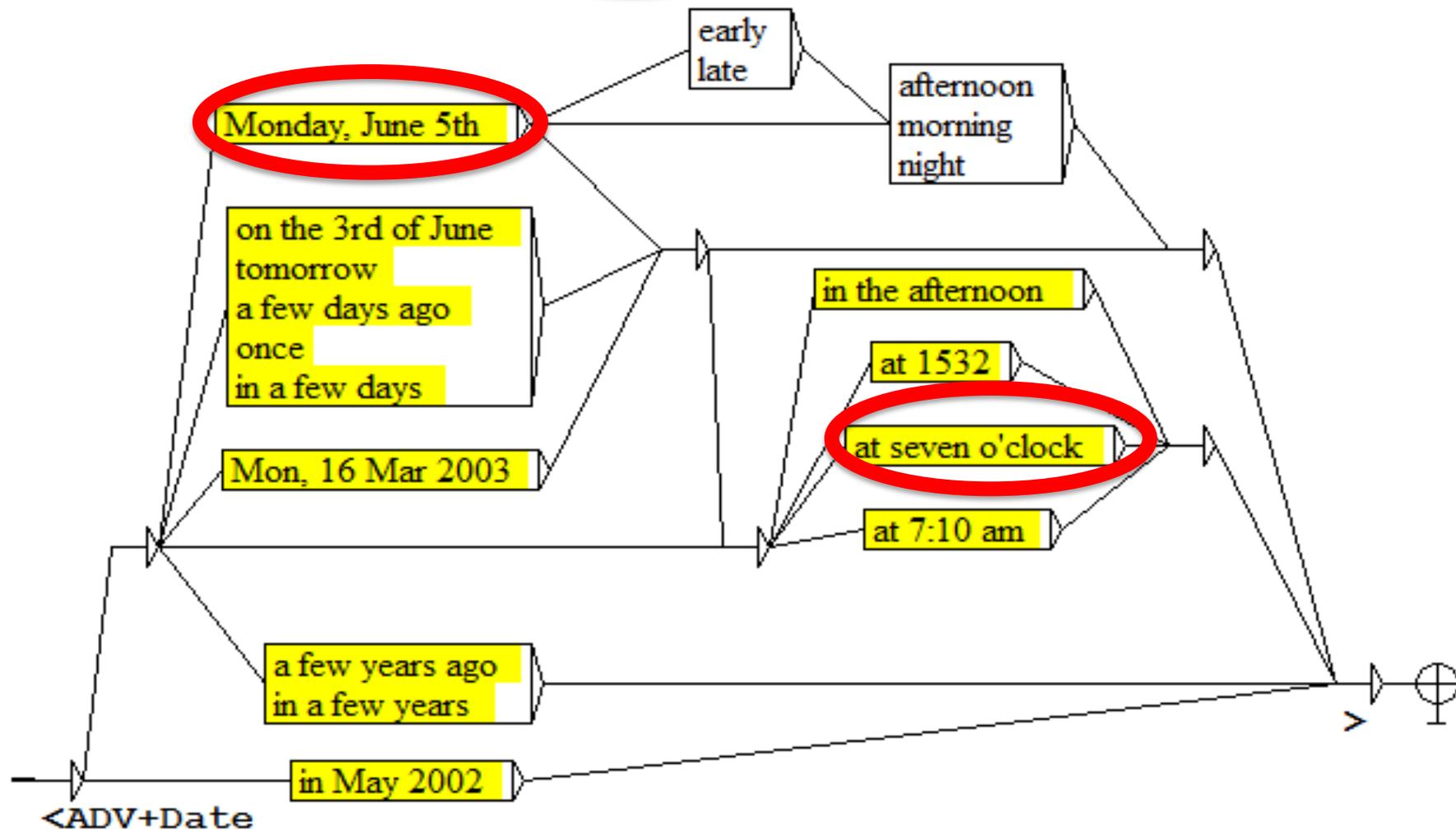
at seven o'clock



Ambiguity with: Joe is only looking "at one" car

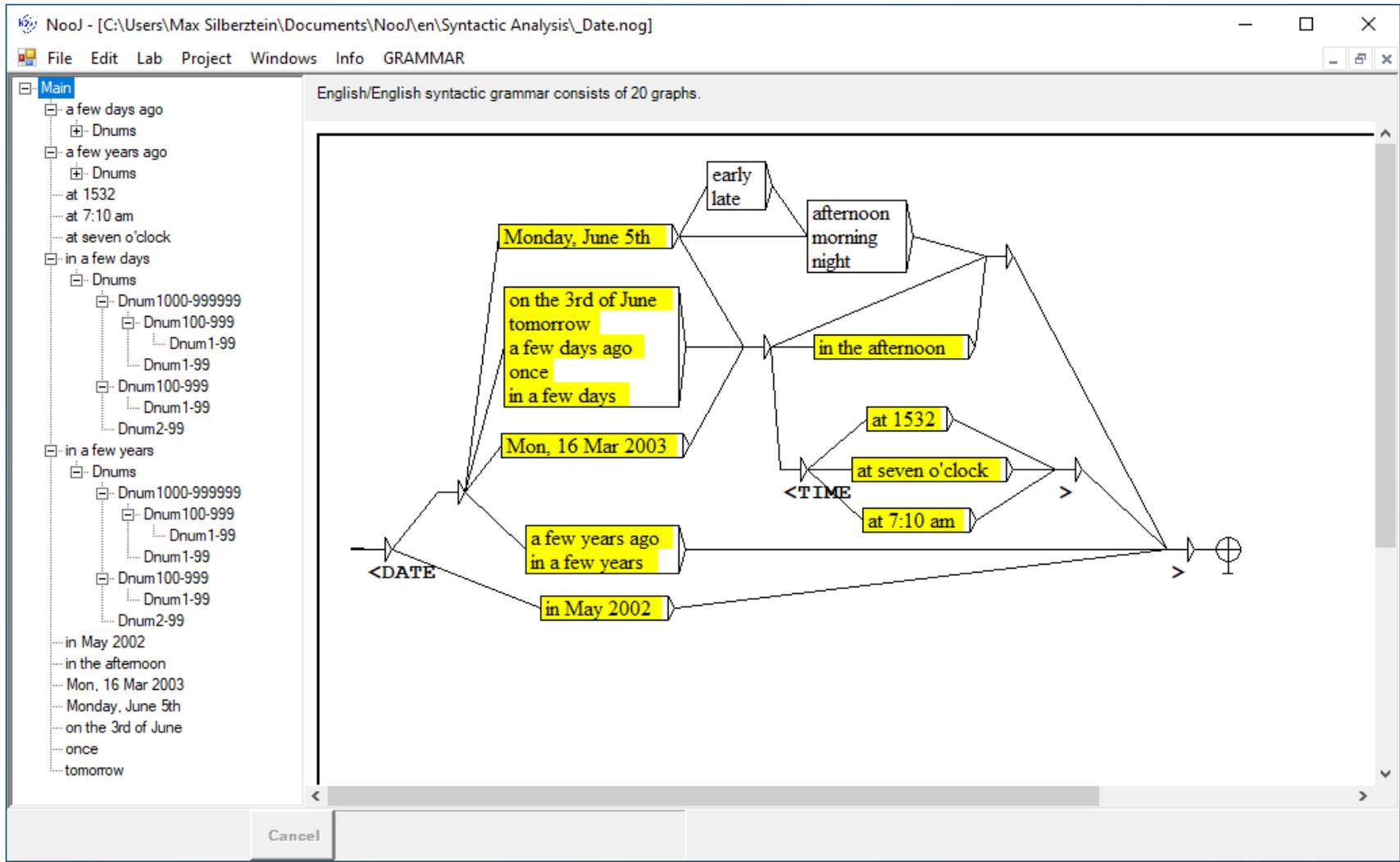
Cancel

English /English syntactic grammar consists of 20 graphs.



Context-Free Grammars

- A **Main** graph, plus any number of embedded graphs
- Recursion is allowed



Debug a grammar: green, blue, or red results

NooJ Community Edition - [C:\Users\Max\Documents\NooJ\en\Syntactic Analysis_Date.nog]

File Edit Lab Project Windows Info GRAMMAR

English/English syntactic grammar consists of 20 graphs.

at seven o'clock

five
ten
a quarter
twenty
twenty-five

of
to

midnight
noon

one
two
three
four
five
six
seven
eight
nine
ten
eleven
twelve

half

after
past

about
approximately
approx.
around

at

one

in the morning

am
a.m.
pm
p.m.
in the morning
in the afternoon

problem:
John looks <at one> TV

Enter expression:

1 analysis. Click a solution below to display the corresponding path:

Perfect Match Partial Match Failure

Paths

("Main" ("at seven o'clock" at five to ten in the morning))	<ADV+Date>

6.7 sec Cancel

Associate a grammar with a contract: green, or red results

NooJ Community Edition - [C:\Users\Max\Documents\NooJ\en\Syntactic Analysis_Date.nog]

File Edit Lab Project Windows Info GRAMMAR

English/English syntactic grammar consists of 20 graphs.

at seven o'clock

five
ten
a quarter
twenty
twenty-five

of
to

midnight
noon

problem:
John looks <at one> TV

one
two
three
four
five
six
seven
eight
nine
ten
eleven
twelve

about
approximately
approx.
around

half

after
past

am
a.m.
pm
p.m.
in the morning
in the afternoon

at

one
two

in the morning
in the afternoon

Contract

Check

March 23rd
*Tuesday, February the 30st
on Monday 14th
on the 22nd of May, 2002
a long time ago
a couple of years ago
two years ago
two months ago

6.7 sec

Cancel

Grammars that recognize named entities for institutions, dates, durations, addresses, professions, organizations, etc.
Organizations, Hela Fehri, Univ. of Sfax

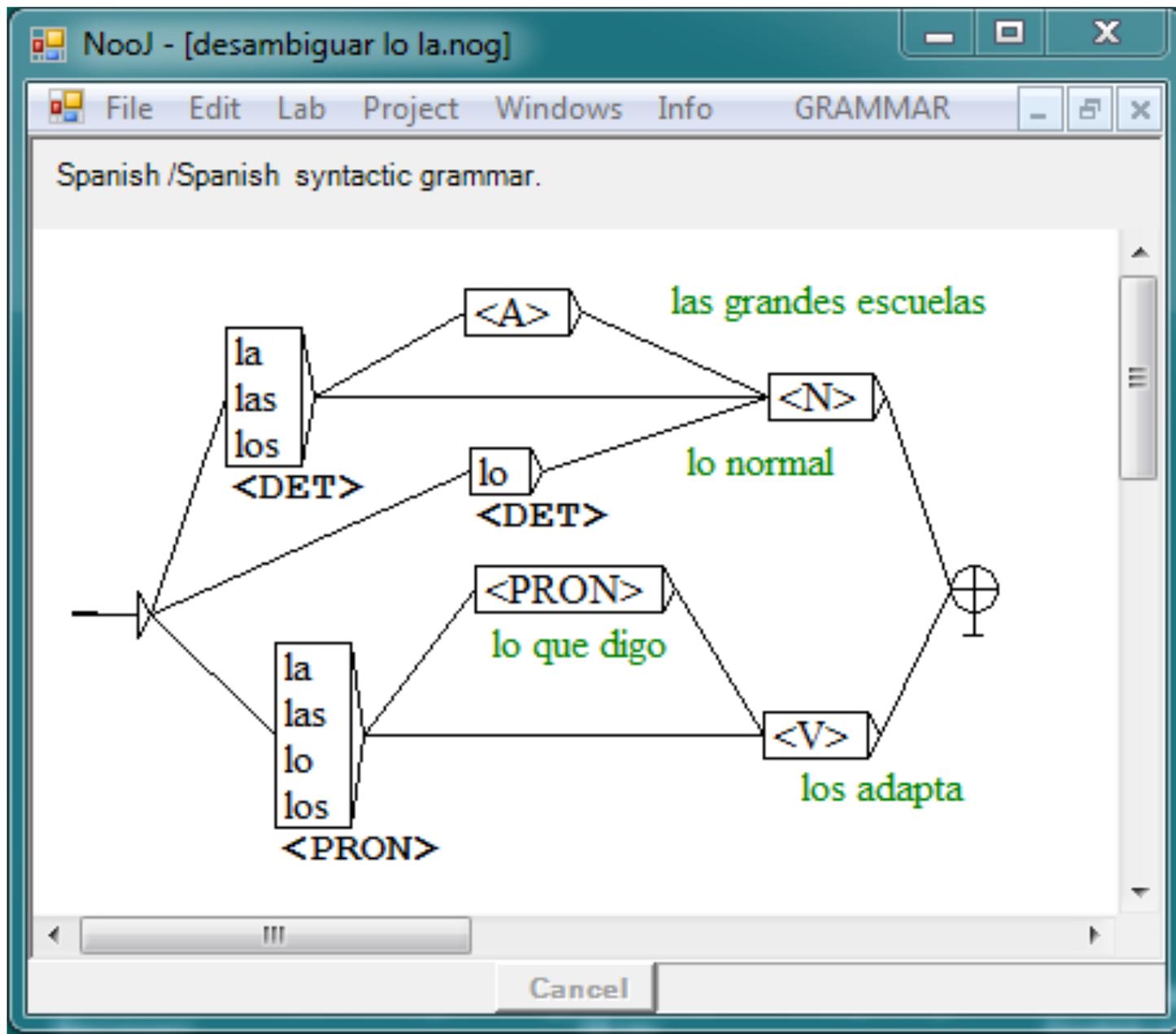
Concordance for Corpus Corpus corpusorg.noc

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

word forms

Text	After	Seq.
text 1.not	مدينة الزقازيق " محافظة الشرقية " 8- جامعة	جامعة الزقازيق
text 1.not	أسيوط "الصعيد" 10- جامعة المنيا ----- المنيا	جامعة أسيوط
text 1.not	المنصورة " محافظة الدقهلية " 13- جامعة قناة	جامعة المنصورة
text 1.not	طريق الاسماعيليه " الدراسة باللغة الانجليزية	جامعة مصر الدولية
text 1.not	مدينة 6 اكتوبر أولا : التخصصات العلمية	جامعة 6 اكتوبر
text 2.not	- أكدال ص.ب. 554 شاله أكدال	جامعة محمد الخامس
text 2.not	ص.ب. 19-9167 زنقة طارق بن	جامعة الحسن الثاني
text 2.not	ص.ب. 524 وجدة - المركب الجامعي	جامعة محمد الأول
text 2.not	للبنزول والمعادن ص.ب 31261 الظهران	جامعة الملك فهد
text 2.not	92 شارع 9 أفريل 1938- تونس- 1007 تليفون: 21671-562700/567322 فاكس	جامعة تونس
text 2.not	بقرطاج 4. (جامعة الحقوق الاقتصادية و	جامعة 7 نوفمبر
text 2.not	(الجنوب) طريق المطار كم 8.5 صفاقس	جامعة صفاقس
text 2.not	دمشق، البرامكة- الجمهورية العربية السورية	جامعة دمشق
text 2.not	حلب، الجمهورية العربية السورية تليفون	جامعة حلب
text 2.not	ص.ب (50) الخوض 123- سلطنة عمان	جامعة السلطان قابوس
text 2.not	بيروت، ص.ب 115020 بيروت تليفون	جامعة بيروت العربية
text 2.not	ص.ب-21- محافظة النجف - الكوفة	جامعة الكوفة
text 2.not	جامعة الموصل-العراق تليفون: 96460-810733 فاكس	جامعة الموصل
text 2.not	الجادرية - بغداد - العراق تليفون : 9641-7760735 فاكس	جامعة بغداد
text 2.not	ص ب 321 - الرمز البريدي 11115 تليفون	جامعة الخرطوم
text 2.not	تليفون : 0811)3-227986 (جوبا فاكس : 00249-11-222142) ص.ب	جامعة جوبا

Dozens of disambiguation grammars



Opinion Mining, Le Monde 2002

NooJ

File Edit Lab Project Windows Info

Locate a pattern in lm2002.noc

Pattern is:

- a string of characters:
- a PERL regular expression:
- a NooJ regular expression:
- a NooJ grammar:

Syntactic Analysis

Index

- Shortest matches
- Longest matches
- All matches

Limitation

- Only: matches
- All matches
- 1 example per match

Reset Concordance

Concordance for Corpus Corpus lm2002.noc

Reset Display: characters word forms before, and after. Display: Matches Outputs

Text	Before	Seq.	After
MS 2002-12.not	s'est ouverte au tourisme	prudemment	, après av
MS 2002-12.not	au Festival de Cannes, manie	habillement	l'art du pe
MS 2002-12.not	une sonde martienne, en précisant	prudemment	: " Les va
MS 2002-12.not	Irak, ainsi que le fonds	généreusement	mis à sa
MS 2002-12.not	la Cinémathèque française s'article	adroitement	autour de
MS 2002-12.not	de les indemniser en jouant	habillement	de ces o
		courageusement	chaque j
		sagement	coupé. J
		cruellement	éprouvée
		cruellement	de donné
		cruellement	la vie inte
		Imprudemment	avancé,
		cruellement	d'eau. Le
		bêtement	au milieu
		cruellement	la supériorité
		cruellement	défaut au
		cruellement	d'équiper
		imprudemment	déclaré C
		bêtement	, dans l'ar
		cruellement	. Chacun

Inverse Document Frequency

Export Report

87.5 sec

Cancel

Automatic detection of grammatical errors and production of correction suggestions

The screenshot displays the Noo Community Edition interface. The main window shows a document titled "Demo Document.not" with the text: "Remember when you were a careless eight year old kid riding a bike with your friends, racing each other around the neighborhood? Remember that feeling of absolute freedom as you felt the wind in your my fr com I was lapto". The words "eight year old" are highlighted in red. A "Concordance" window is open, showing a table with columns "Text", "Before", "Seq.", and "After". The entry for "eight year old" shows the suggestion "Add dashes" between "eight" and "year". A third window, titled "G Dnum N old.nog", displays a syntactic grammar diagram. The diagram shows a sequence of grammatical categories: a start symbol, a box containing "<DET+Dnum>" and "<NB>", a box containing "<N+s>", a list of adjectives ("high", "large", "long", "tall", "old"), and a final terminal symbol. A "Cancel" button is visible at the bottom of the interface.

0 sec

Automatic detection of grammatical errors

Nool Community Edition

File Edit Lab Project Windows Info GRAMMAR

C:\Users\Max\Documents\Nool\en\Projects\Demo Docume... / 8 TUs

Characters Tokens Digrams

Show Text Annotation Structure

That is when I **decided getting** on the bike. I haven't regretted my decision so far. One of the best things about cycling is that the bike is perfect for exercising. Just cycling to and from work or to the shop.

8 13 15

when,ADV I,PRO+Pers=1+Nb=s decide,V+Te

Concordance for Text C:\Users\Max\Documents\Nool\en\Projects\De

Reset Display: 5 characters before, and 5 after. Display word forms

Text	Before	Seq.	After
meetings. That is when I	decided getting/to get	on th	
mountain bike trails so I	could spent/spend	the d	
all of that if I	hadn't decide/decided	to try	

Query 3/3

22.8 sec Cancel

0.2 sec Cancel

Nool - [C:\Users\Max\Documents\Nool\en\Syntactic Analysis\2 G could Vpp.nog]

File Edit Lab Project Windows Info Max TEXT GRAMMAR

English/English syntactic grammar.

could should would

<V+PP> \$THIS_

All I have is here

<be> <V+PP> <N> +EXCLUDE

You have notice You have noticed

<ADV>

<have>

no very +EXCLUDE

<V+PP> \$THIS_ V+PP

<be>

<ADV>

<V+G> <V+PP> <N> +EXCLUDE

I was also scare I was also home

<decide> <hope> <want>

to <V+INF> +EXCLUDE <V+G> to \$THIS_

<V+PP> \$THIS_ V+PP

<V+PP> \$THIS_ V+PP

<V+PP> \$THIS_ V+PP

<V+PP> \$THIS_ V+PP

- Lexical problems: (unknown words)
- Syntactic problems (auxiliary verbs+ compound adjectives)
- Punctuation problems (missing space + extra space)

NooJ Community Edition

File Edit Lab Project Windows Info CONCORDANCE

C:\Users\Max\Documents\NooJ\en\Projects\Demo Document.not

- 3 + / 8 TUs

Characters Tokens Digrams

Language is "English(en)".
Text Delimiter is: \n (NEWLINE)
Text contains 8 Text Units (TUs).
729 tokens including:

Show Text Annotation Structure

That is when I **decided getting** on the bike. I haven't regretted my decision so far. One of the best things about cycling is that the bike is perfect for exercising. Just cycling to and from work or to the shops every day is enough to keep you healthy and happy. Besides, **its** incredibly liberating to be able to get anywhere without losing time in traffic jams. Also don't forget about the pollution while reducing also traffic congestion and the At some point, **I** realized that I started to use the bike with friends and to head out for coffee on weekends. I appreciate what you are seeing around you. You can distance. That daily distance I rode to work was no decent mountain bike trails so I **could spent** the day Because I enjoyed it.

"I wish I could bike all the way round the world" I said OK, **I** knew I couldn't ride across the oceans. I came continents, **from** coast to coast. The more I thought about If I will do this, **I** will have to thoroughly prepare, **I** thought over towering mountains and across burning deserts? language?

After a few months of training, **I** set off. This was hard people cheered me on when they **heard** about my idea

Concordance for Text C:\Users\Max\Documents\NooJ\en\Projects\Demo Doc...

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

Before	Seq.	After
and yet you can cover	alot/+Unknown word	of distance. That daily distan
cheered me on when they	heard/+Unknown word	about my journey. The news;
month later, I found myself	petaling/+Unknown word	back toward the place where
meetings. That is when I	decided getting/to get	on the bike. I haven
mountain bike trails so I	could spent/spend	the day going up and
all of that if I	hadn't decide/decided	to try biking instead of
when you were a careless	eight year old/Add dashes	kid riding a bike with
I could continue on my	way ./Delete space	I never gave up on
a bike with your friends	,racing/Add a space	each other around the night
way as a grown up	,until/Add a space	my friends presented me a
brand-new bike. At first	,I/Add a space	was a bit skeptical about
a couple of days later	,I/Add a space	changed completely my mind

GRAM = G Lexical Mistakes 31/31



CONGRATULATIONS



You know how to develop complex grammars
with the graph editor to develop various NLP
applications

