



13. Context-Sensitive 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

Context-Sensitive Grammars

- Context-sensitive grammars are equivalent to Type 1 Generative Grammars in Chomsky-Schützenberger's hierarchy
- Context-sensitive grammars can be more powerful than regular and context-free grammars.
- They can recognize any context-sensitive language, are very well-adapted to describe distributional and agreement constraints between linguistic units, and can be processed by linear bounded automata.
- Lexical Functional Grammars (LFG) are equivalent to context-sensitive grammars.

Context-Sensitive Grammars

- In NooJ, context-sensitive grammars are context-free grammars + variables and constraints.
- Context-sensitive grammars have two components: the description of sequences and the description of agreement constraints.

Type 1 : Context-Sensitive Grammars

A grammar that recognizes words that contains the same number of “a”, “b” and “c”

The screenshot shows the NooJ interface with a grammar file named `a^n b^n c^n.nog`. The grammar is defined as:

French /French syntactic grammar.

A B C $\langle \$B\$LENGTH=\$A\$LENGTH \rangle$
 $\langle \$C\$LENGTH=\$A\$LENGTH \rangle$

The diagram illustrates the grammar rules for generating strings with equal numbers of 'a', 'b', and 'c'. It shows three non-terminals: A , B , and C . Each non-terminal is represented by a box containing the character it generates (a, b, or c) with a self-loop arrow above it, indicating that each can be repeated. These are connected by hyphens, representing concatenation. To the right, there are two context-sensitive constraints: $\langle \$B\$LENGTH=\$A\$LENGTH \rangle$ and $\langle \$C\$LENGTH=\$A\$LENGTH \rangle$, which ensure that the number of 'b's and 'c's matches the number of 'a's.

Contract for: `a^n b^n c^n`

Check

```
# NooJ
# Enter examples, *counter-examples and #comments
#
a a a b b b c c c
a a a b b c c c
```

Cancel

Type 1 : Context-Sensitive Grammars

A grammar that recognizes words constituted by a number of "a" that is a power of 2, e.g., 2, 4, 8, 16...

The screenshot shows the NoJ software interface. The main window displays a grammar graph for the language a^{2^n} . The graph starts with a root node labeled "Phrase" on the left. A line connects it to a node labeled "a a". From "a a", two lines branch out to a node with a circle containing a plus sign (\oplus) and a node with a circle containing a cross (\otimes). The \oplus node is connected to a node labeled "Phrase" in a yellow box, which is enclosed in red parentheses and labeled "X" below. The \otimes node is connected to another node labeled "Phrase" in a yellow box, also enclosed in red parentheses and labeled "Y" below. A line connects the \otimes node to a node labeled $\langle \$X=\$Y \rangle$. Below the graph is a scroll bar. A second window titled "Contract for: a^{2^n} " is open, showing a "Check" button and a text area with the following content:

```
# 12 n'est pas une puissance de 2 :  
a a a a a a a a a a a a a a  
# 16 est une puissance de 2 :  
a a a a a a a a a a a a a a a
```

At the bottom of the NoJ window is a "Cancel" button.

Type 1 : Context-Sensitive Grammars

A grammar that recognizes reduplications, e.g., bye bye, chouchou, papa

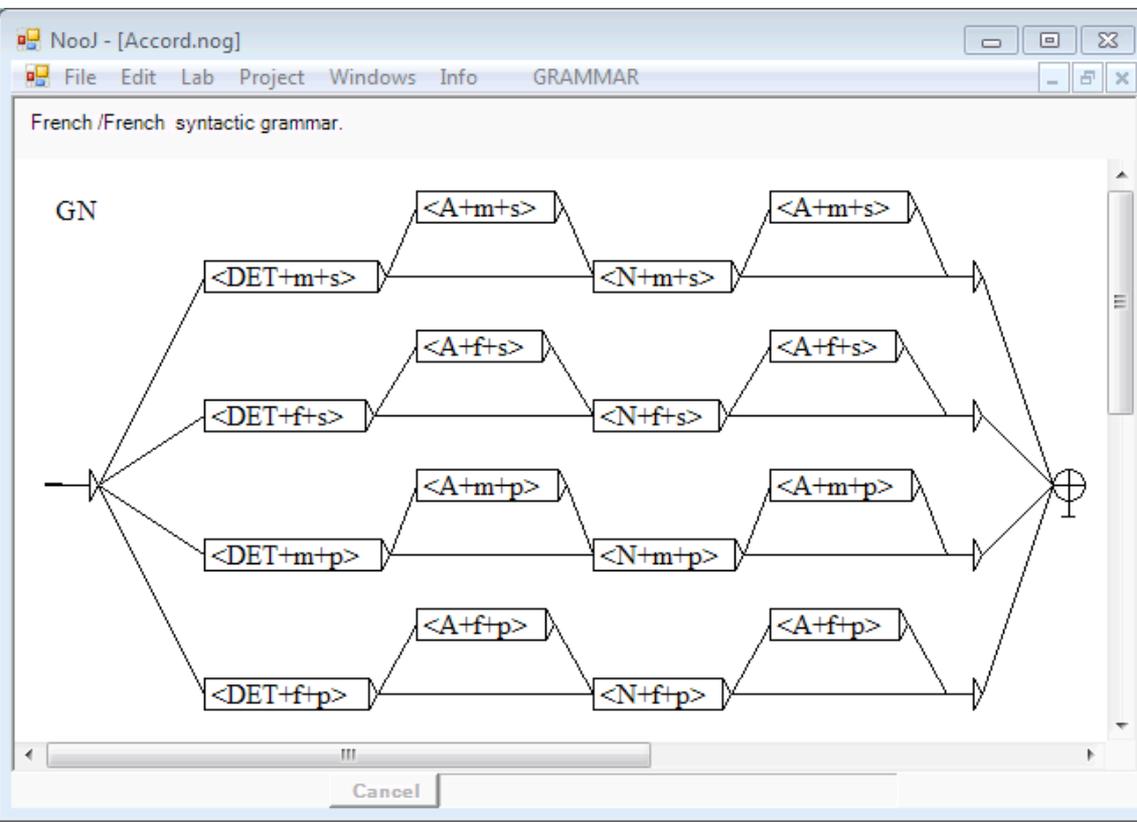
The screenshot shows the NooJ software interface with the following components:

- Window Title:** NooJ
- Menu Bar:** File Edit Lab Project Windows Info GRAMMAR
- File Name:** Réductions syntaxiques.nog
- Text:** French /French syntactic grammar.
- Diagram:** A syntactic tree diagram representing the grammar. It starts with a root node (a triangle with a horizontal line) leading to a red 'X'. This is followed by a red '(' parenthesis, a box containing '<WF>' with a self-loop arrow above it, and a red ')' parenthesis. This is followed by a red 'Y', another red '(' parenthesis, another box containing '<WF>' with a self-loop arrow above it, and another red ')' parenthesis. Finally, there is a triangle node leading to a circle node containing '<\$X=\$Y>', which then leads to a final triangle node.
- Contract for: Réductions syntaxiques**
 - Check** button
 - #** symbol
 - Bye Bye** (text in green)
- Cancel** button

Context-Sensitive Grammars

- Example of a grammar that describes the agreement in number and in gender in some French noun phrases:

$\langle \text{DET}+\text{m}+\text{s} \rangle (\langle \text{A}+\text{m}+\text{s} \rangle | \langle \text{E} \rangle) \langle \text{N}+\text{m}+\text{s} \rangle (\langle \text{A}+\text{m}+\text{s} \rangle | \langle \text{E} \rangle) |$
 $\langle \text{DET}+\text{f}+\text{s} \rangle (\langle \text{A}+\text{f}+\text{s} \rangle | \langle \text{E} \rangle) \langle \text{N}+\text{f}+\text{s} \rangle (\langle \text{A}+\text{f}+\text{s} \rangle | \langle \text{E} \rangle) |$
 $\langle \text{DET}+\text{m}+\text{p} \rangle (\langle \text{A}+\text{m}+\text{p} \rangle | \langle \text{E} \rangle) \langle \text{N}+\text{m}+\text{p} \rangle (\langle \text{A}+\text{m}+\text{p} \rangle | \langle \text{E} \rangle) |$
 $\langle \text{DET}+\text{f}+\text{p} \rangle (\langle \text{A}+\text{f}+\text{p} \rangle | \langle \text{E} \rangle) \langle \text{N}+\text{f}+\text{p} \rangle (\langle \text{A}+\text{f}+\text{p} \rangle | \langle \text{E} \rangle)$



Four quasi-identical terms for French...

But 30 quasi-identical terms for some slavic languages (2 numbers, 3 genders, 5 cases).

Context-Sensitive Grammars

- Example of a Context-Free Grammar that only describes the structure of some French noun phrases:

$GN = \langle \text{DET} \rangle (\langle \text{A} \rangle | \langle \text{E} \rangle) \langle \text{N} \rangle (\langle \text{A} \rangle | \langle \text{E} \rangle)$

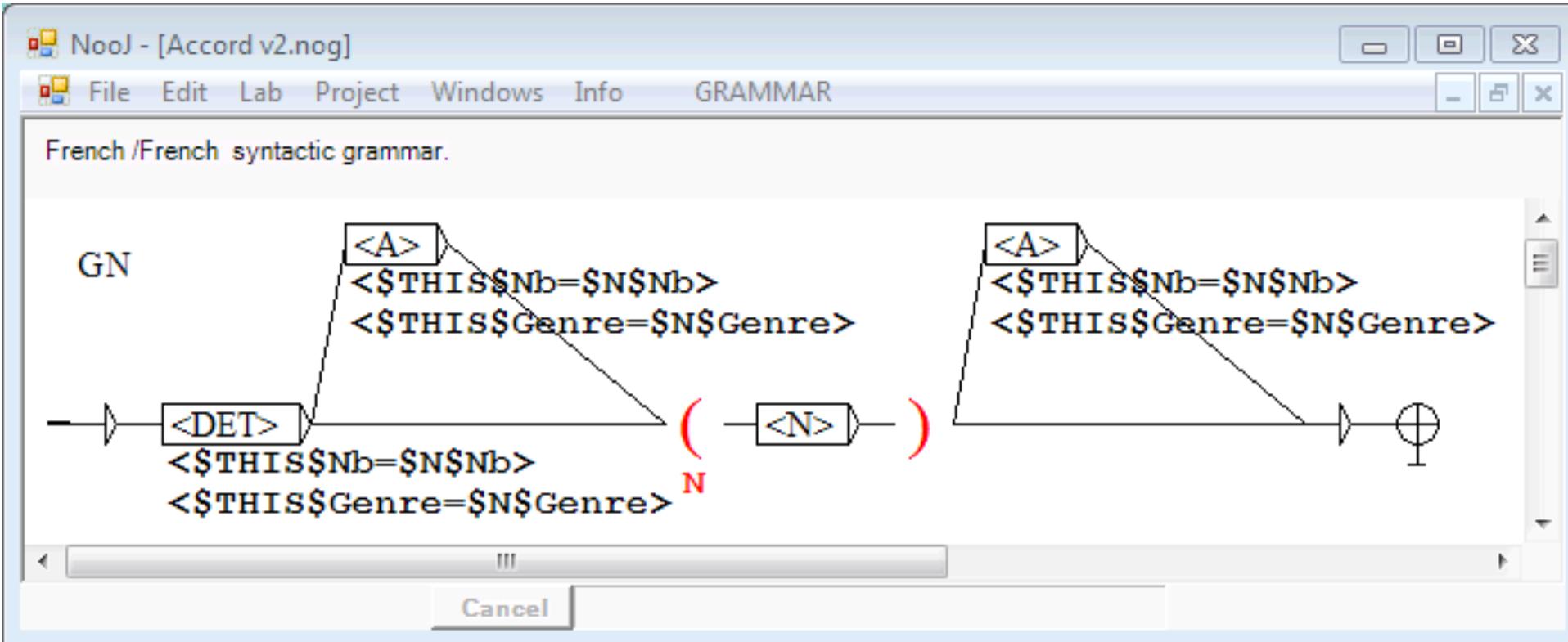
- We then add agreement constraints for gender and number:

$\langle \$THIS\$Gender = \$N\$Gender \rangle$

$\langle \$THIS\$Number = \$N\$Number \rangle$

- Variable $\$THIS$ refers to the current ALU; variable $\$N$ refers to the noun.

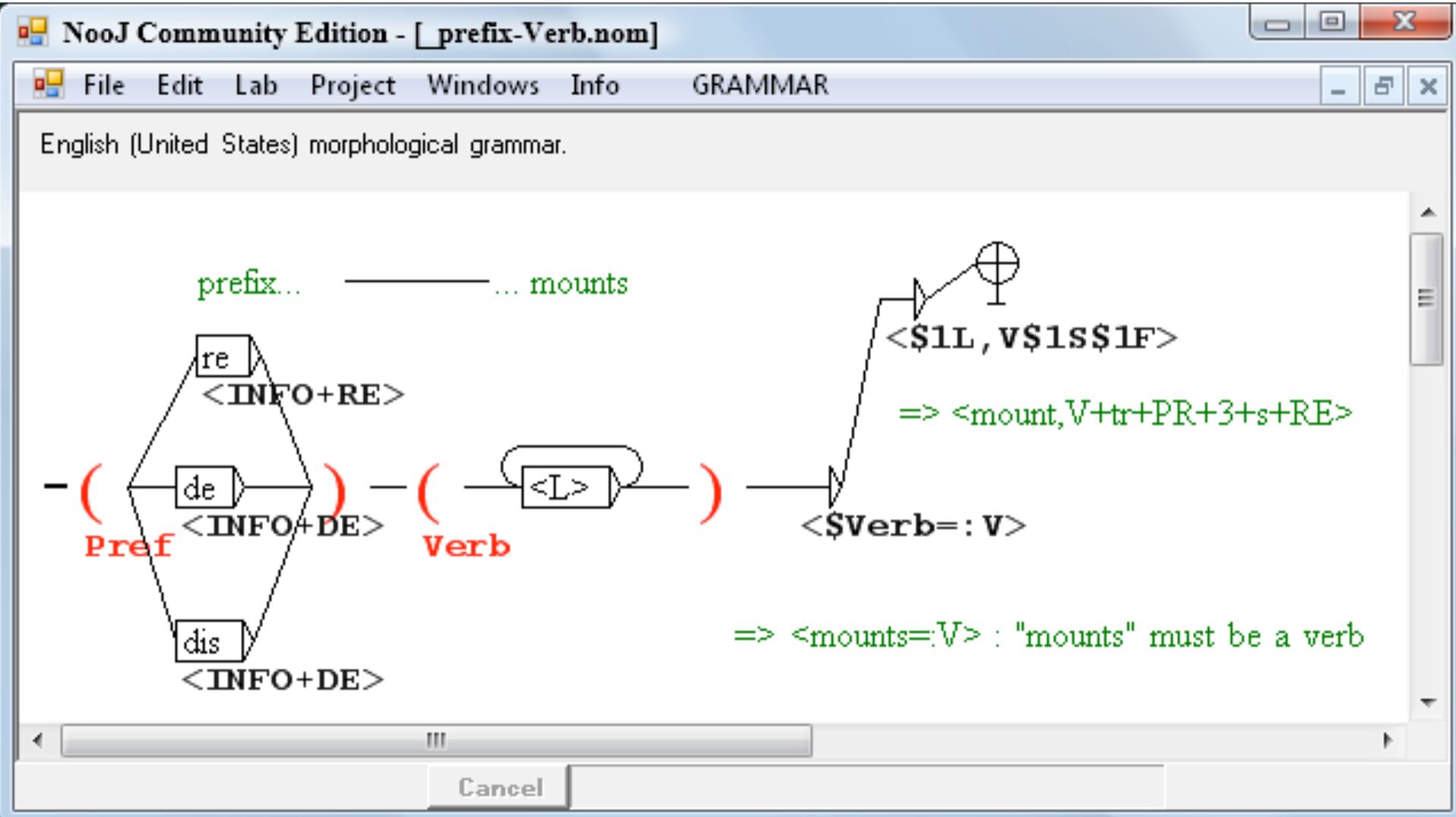
Context-Sensitive Grammars



- The noun **<N>** is stored in variable **\$N**
- Special variable **\$THIS** refers to the current ALU
- Variable **\$N\$Nb** refers to the “NB” property of the noun; **\$THIS\$Nb** refers to the “NB” property of the current ALU;
- Variable **\$N\$Genre** refers to the “Genre” property of the noun; **\$THIS\$Genre** refers to the “Genre” property of the current ALU.

A morphological context-sensitive grammar

A grammar that recognizes wordforms that start with “re”, “de” or “dis”
The constraint $\langle \$Verb = :V \rangle$ checks that the following letters constitute a valid verb.



There are three types of constraints

- The operator “=” tests that the two sides are identical strings, *e.g.*:

<\$X=\$Y> tests that ALU \$X is identical to ALU \$Y

<\$N\$Nb="plural"> tests that ALU \$N number is “plural”

<\$DET\$Nb=\$N\$Nb> tests that ALUs \$DET and \$N agree in number

- Matching:

<\$X=:N> tests that \$X is a noun (<N>)

<\$Y=:have> tests that \$Y is a form of have (<have>)

<\$V=:V+Trans+F+3+s> test that \$V is a transitive verb conjugated in the Present, third person, singular (<V+Trans+F+3+s>)

- Existence:

<\$NE> tests that variable \$NE has a non-empty value, *i.e.*, the corresponding ALU occurs in the matching sequence.



CONGRATULATIONS



You know how to construct context-sensitive grammars to describe phrase structures and agreements between their components

