

# Syntactic Structures of Sentences from Large Corpora

Emmanuel Giguet and Jacques Vergne  
GREYC — CNRS UPRESA 6072 — Université de Caen  
14032 Caen cedex — France  
{Emmanuel.Giguet,Jacques.Vergne}@info.unicaen.fr

## 1 The Demonstration

The demonstration will consist in displaying the syntactic structure of sentences from novels, scientific texts, newspapers, ... The syntactic structures are computed by our syntactic parser and the output shows in a human-friendly graphic interface (1) word features (as computed by POS tagger) (2) non-recursive phrases (as computed by shallow parser) and (3) their relations (the functional structure). It will be a closed demo. The outputs will be split into sets in order to focus on the resolution of different kinds of problems (i.e. coordination, long distance subject-verb relation, preposition/determiner resolution). A large scale corpus will be available to prove that the results are not hand-made and open demos will be possible during informal demo sessions.

## 2 Demo Sessions

Schedule of the intended demo sessions:

Tuesday April 1	12.30-13.30 16.30-17.30
Wednesday April 2	10.00-11.00 15.30-17.00
Thursday April 3	10.30-12.00 15.00-16.00 16.30-17.00

Other informal sessions can be scheduled in order to allow people to parse their own sentences or texts.

## 3 The Graphic Output

Our viewer, developed with Java, allows to display in a graphical way (see Fig. 1) the dependency tree between non-recursive phrases and the tags of words.

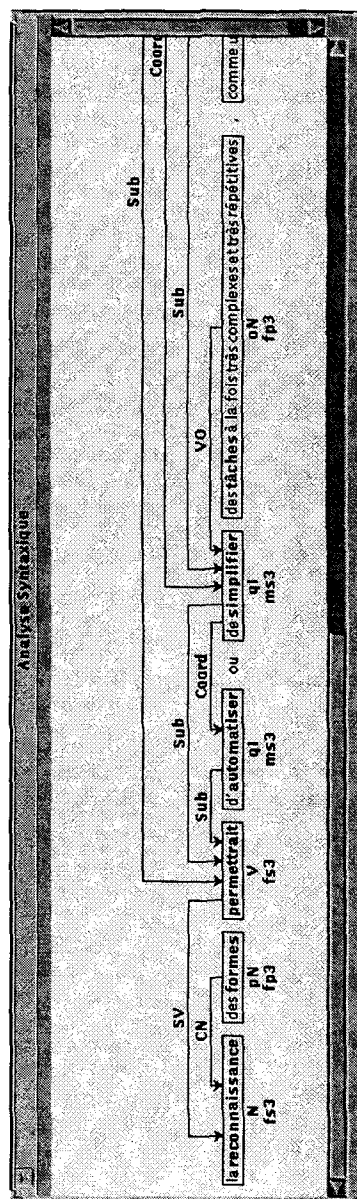


Figure 1: A GRAPHIC OUTPUT

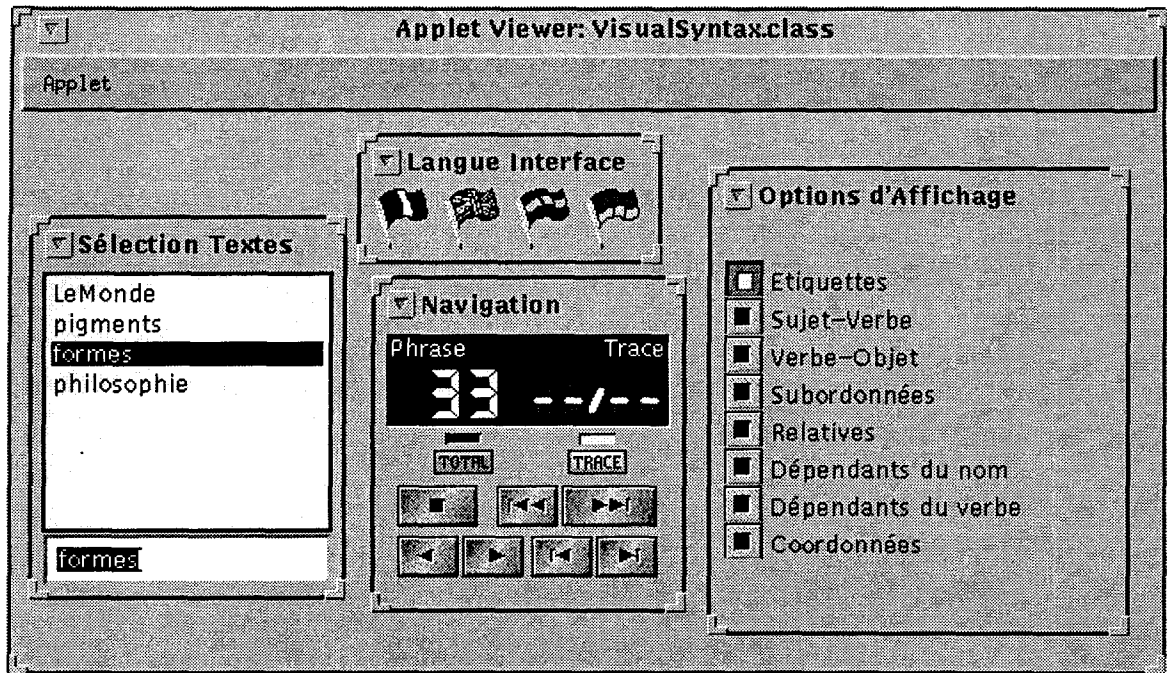


Figure 2: THE CONTROL PANEL

#### 4 The Control Panel

The control panel (see Fig. 2) still in progress allows the choice of the corpus, the sentence to be displayed. Filters can be applied in order to show or to hide specified dependency and coordination relations. The tags of the words can also be shown or hidden. This principle allows to concentrate on word-level or on dependency-level and allows to check precisely the behavior of our parser on specific relation types. A trace mode enables the step by step display of the construction of the analysis.

In a near future, a graphic comparison between two outputs will be available to control the modification of the rule base and to point out easily the differences. This comparison between two outputs can also be used to compare an output and an expected output.

#### 5 Web Site Opening

After this conference, the same demonstration will be available on Internet at the following URL.

<http://www.info.unicaen.fr/~giguet>

This website will required a browser running Java and is platform independent (It has been tested with Netscape on Linux, Solaris2.5, MacOS, Windows NT and Windows 95).

#### 6 Getting Annotated Corpora

People who are interested in getting annotated corpora are invited to contact the authors.

#### 7 References

No references are available yet.