

A NooJ Grammar of the French Nucleus Verb Phrase

François Trouilleux – Clermont Université, Université Blaise-Pascal, EA 999, LRL – November 2012

This document provides basic information about a NooJ grammar of the French nucleus verb phrase: the license, how to use the grammar with NooJ and a series of screen captures of the graphical NooJ grammar. The reference article for this grammar is:

François Trouilleux, 2013. “A Description of the French Nucleus VP Using Co-occurrence Constraints”, in *Formalising Natural Languages with NooJ. Selected Papers from the NooJ 2012 International Conference*. A. Donabedian, V. Khurshudian, M. Silberztein, eds. Cambridge Scholars Publishing, 2013.

License

The grammar is distributed under the terms of the GNU General Public License, cf. <http://www.gnu.org/licenses/licenses.html#GPL>.

How to use the grammar

The grammar comes as a NooJ Project file named `French_Nucleus_VP.nop`. To open it, launch NooJ (www.nooj4nlp.net) and select the menu item **Project > Run Project**.

The project contains the following files (use the **NooJ Windows** menu to select windows):

- `French_Nucleus_VP.nog`, the grammar itself;
- `FNVP_special.dic`, a small dictionary of verbs to be used with a higher priority than the DM; verbs in this dictionary have information on transitivity which the DM has not (cf. <http://www.univ-bpclermont.fr/LABOS/lrl/spip.php?rubrique48>); grammaticality of strings for which this aspect is relevant (*i.e.* the passive voice) should only be tested with respect to the sample vocabulary in `FNVP_special.dic`;
- `FNVP_sample-text.not`, a small text to illustrate how the grammar may be used to annotate text;
- `properties.def`, a NooJ properties definition file; it specifies the vocabulary to be used in constraints.

To navigate the various graphs of the grammar, while the grammar window is selected, select **GRAMMAR > Show structure**.

To see the “contract” of the grammar, *i.e.* a set of grammatical and ungrammatical strings it should or should not parse: while the grammar window is selected, select **GRAMMAR > Show contract**. After clicking on the **Check** button, all string should come out green. If it is not the case, please report the problem to me, indicating the NooJ version you are using (**Info > About NooJ**).

To see the annotations of the sample text, check the **Show Text Annotation Structure** box in the text window. Alternatively, you may export the annotations as XML-like tagging: while the text window is selected, select **TEXT > Export annotated text as an XML document**. You should obtain a pseudo XML file named `FNVP_sample-text.xml.txt`. Note that while the grammar recognizes *elle n’oublia pas* and *il était engagé*, it annotates *elle n’oublia* and *il était* as NVs, *pas* as GR, and *engagé* as NVPP (EASY-style segmentation). The text contains significant examples with past participles:

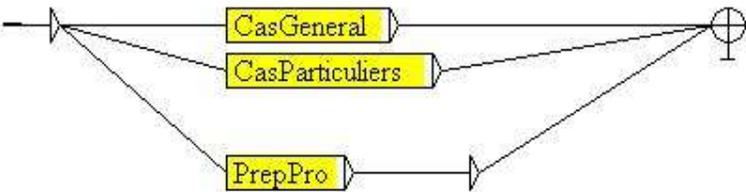
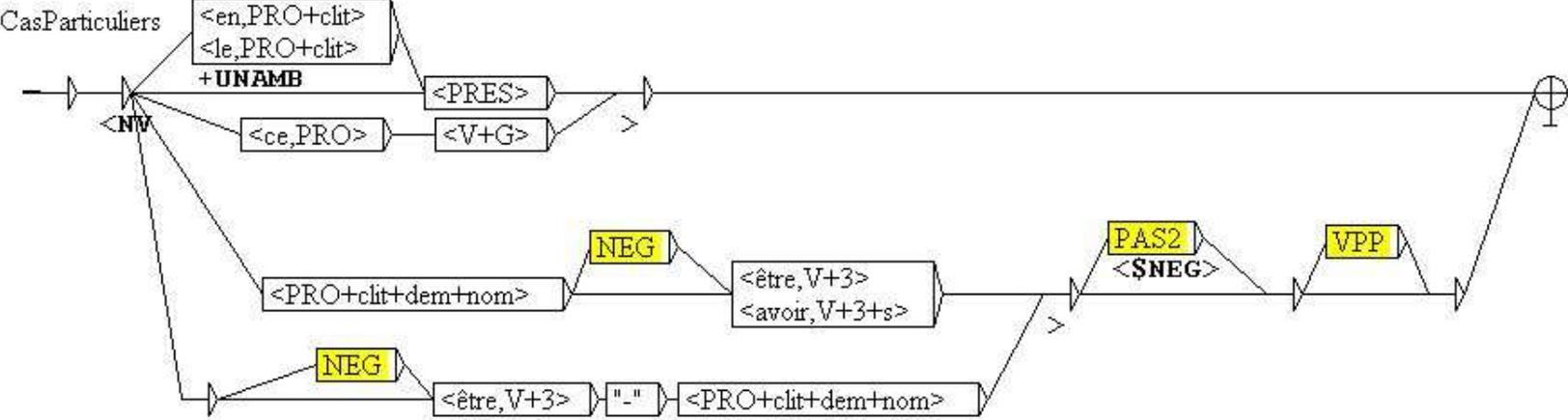
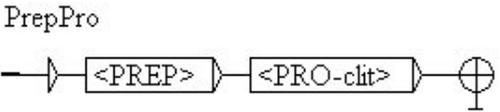
- in *Il serait, peut-être, deux ans parti*, the grammar fails to identify *parti* as a past participle NV; it handles only simpler sequences between the auxiliary verb and the past participle;
- in *N’avez-vous pas crainte*, the grammar adequately does *not* identify *crainte* as a past participle NV, because there is no feminine direct object before the verb;

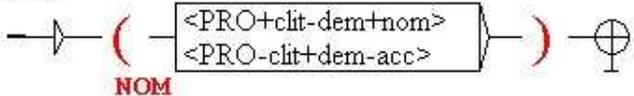
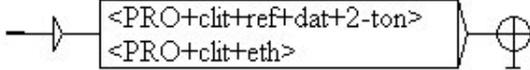
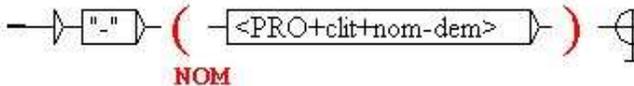
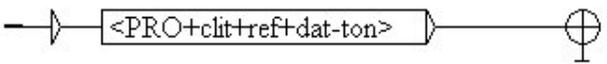
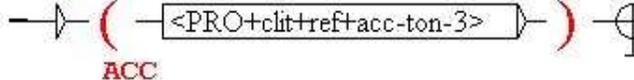
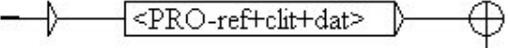
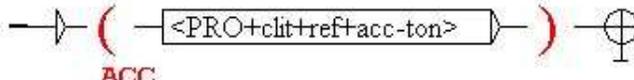
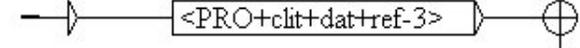
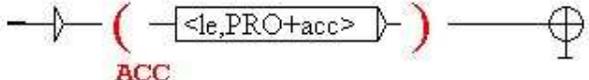
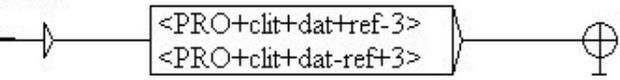
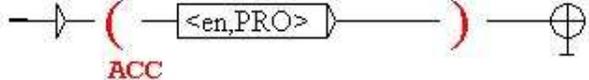
- in *leurs grands-parents défunts que les Liébard avaient connus*, the grammar fails to identify *connus* as a past participle NV, because it fails to identify the direct object *que* before the verb; such cases are out of scope of the grammar; only clitic pronoun direct objects are accessible to grammar.

See the NooJ manual for other uses (parsing your own text, editing the grammar, etc.).

Commented screen captures

The following screen captures of the graphical NooJ grammar serve two goals: (1) they allow us to document a few details which are not discussed in the paper, (2) they provide people with an overview of the grammar without having to download and use the NooJ software.

 <p>The diagram shows a root node on the left with three outgoing arrows. The top arrow points to a yellow box labeled 'CasGeneral'. The middle arrow points to a yellow box labeled 'CasParticuliers'. The bottom arrow points to a yellow box labeled 'PrepPro'. All three boxes have arrows pointing to a final node on the right, which is a circle with a cross inside.</p>	<p>Main graph The subgraph CasGeneral is the one documented in the article.</p>
 <p>The diagram shows a root node on the left with several outgoing arrows. The top arrow points to a box containing '<en, PRO+clit>' and '<le, PRO+clit>' with '+UNAMB' below them. The second arrow points to '<PRES>'. The third arrow points to '<ce, PRO>' and '<V+G>'. The fourth arrow points to '<PRO+clit+dem+nom>', which then points to a yellow box 'NEG', which points to a box containing '<être, V+3>' and '<avoir, V+3+s>'. The fifth arrow points to another yellow box 'NEG', which points to a box containing '<être, V+3>', '<-"', and '<PRO+clit+dem+nom>'. The box with '<être, V+3>' and '<avoir, V+3+s>' points to a yellow box 'PAS2' containing '<\$NEG>', which points to a yellow box 'VPP'. All paths eventually lead to a final node on the right, a circle with a cross inside.</p>	<p>Three particular cases:</p> <ul style="list-style-type: none"> – presentatives : <i>en voici,</i> <i>le voilà ;</i> – sequences with <i>ce</i> as object: <i>ce faisant ;</i> – sequences with <i>ce</i> as subject: <i>c'est,</i> <i>n'est-ce pas,</i> <i>c'aurait.</i>
 <p>The diagram shows a root node on the left with one outgoing arrow pointing to a box containing '<PREP>', which then points to a box containing '<PRO-clit>', which finally points to a final node on the right, a circle with a cross inside.</p>	<p>The PrepPro graph is not part of the FNVP definition. It is included to prevent some pronouns to be analysed as clitics in text analysis (e.g. in <i>au dedans de lui grondait un orage</i>).</p>

<p>NOM</p> 	<p>Demonstrative pronouns are included in the NV as subjects.</p>	<p>DAT0</p> 	<p>In our view, ethical dative pronouns are always 2nd person. The DM has <i>te me</i>, <i>te nous</i> and <i>te vous</i> as atomic ethical datives (+eth)</p>
<p>NOMi</p> 		<p>DAT1</p>  <p>DAT1r</p> 	
<p>ACC1</p> 	<p>Slot 1 pronouns before the verb may not be marked +ton (<i>moi</i>, <i>toi</i>). Non reflexives are only first or second person.</p>	<p>DAT2</p> 	
<p>ACC1r</p> 		<p>DATir</p> 	
<p>ACC2</p> 		<p>DATi</p> 	
<p>ACC3</p> 			
<p>ACCI</p> 