Annotating Text With Local Grammars

Practical Uses

Éric Laporte

Annotating Text
With Local Grammars

Unitex/GramLab provides tools for inserting and modifying annotations in text

How to insert annotations?

How to copy information into annotations?

Where to place annotations?

How to delete annotations?

Outline

Insert annotations

Copy from text to annotation

Copy from a dictionary

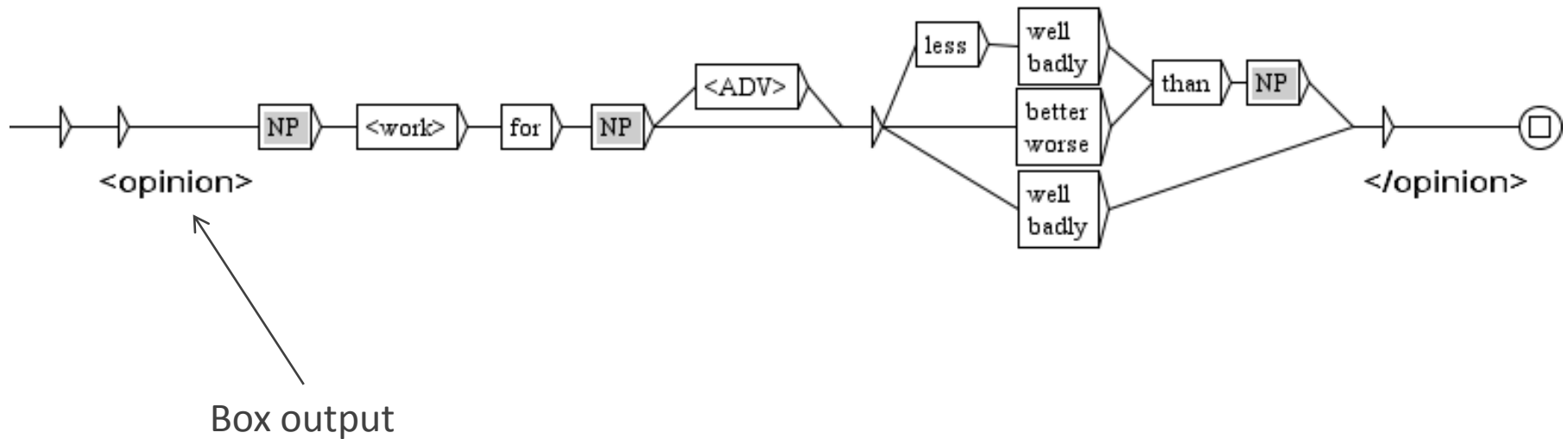Annotate at the right place

Delete annotations

Inserting annotations

It worked for my mental health better than any other drug

↓

<opinion>It worked for my mental health better than any other drug</opinion>

Box output

Inserting annotations

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

It worked for my mental health better than any other drug

↓

<opinion>It worked for my mental health better than any other drug</opinion>

Check the 'Merge with input text' option of Text > Locate Pattern

By default, box outputs are ignored

Inserting annotations

It worked for my mental health better than any other drug

↓

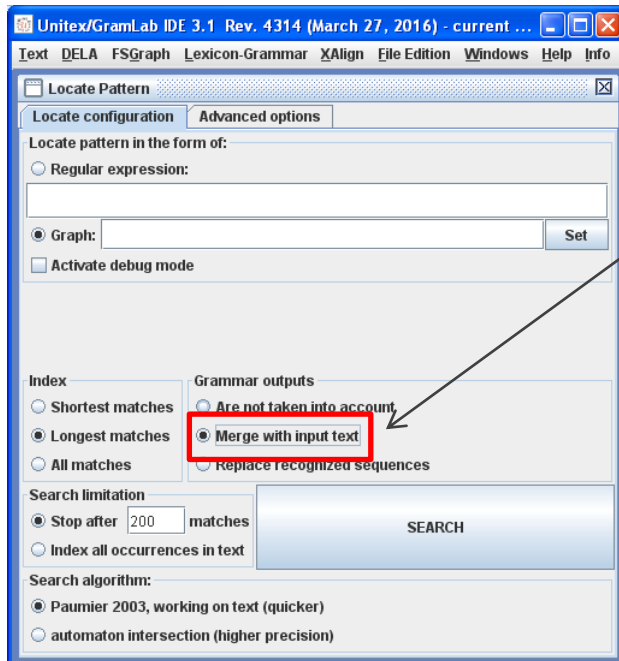<opinion>*It worked for my mental health better than any other drug*</opinion>

Clearly distinguish annotations from the original corpus

XML tags are a good solution for that

They are delimited by < > which are rarely used in normal text

Occurrences of '<' in the original text may be replaced by `<`

Outline

Copy from text to annotation

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<comp phrase='better'/></opinion>

Copy from text to annotation

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<comp phrase='better'/></opinion>

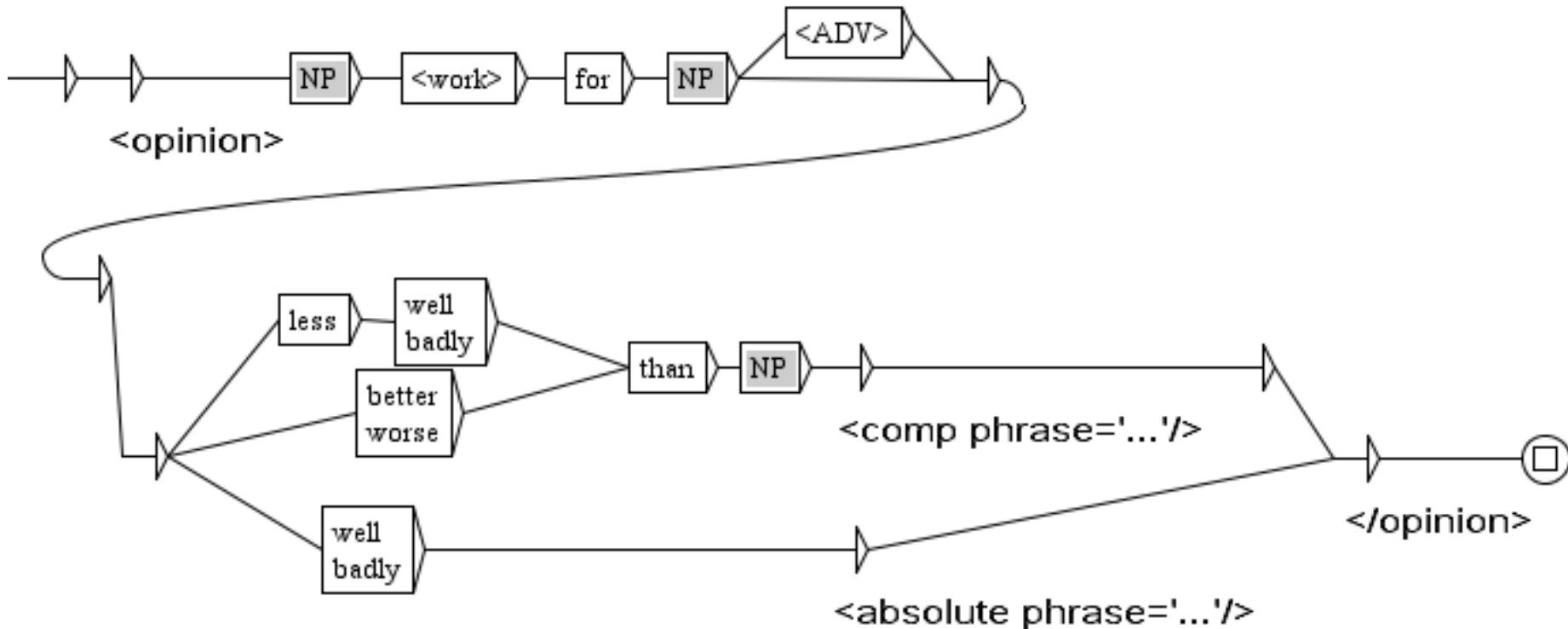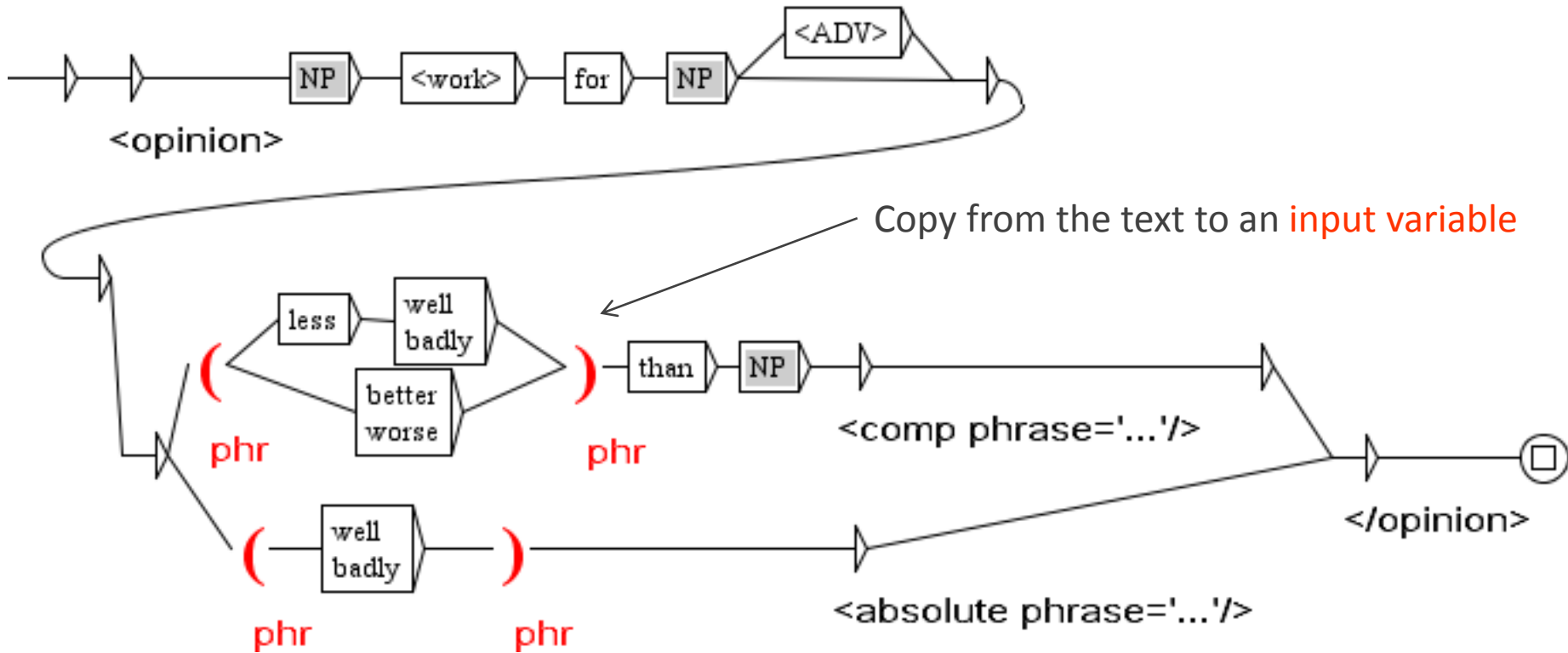Copy from the text to an input variable

Copy from text to annotation

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<comp phrase='better'/></opinion>*

Copy from the input variable into an annotation

It worked for me better than any other drug

*<opinion>It worked for me better than any other drug**<content polarity='good'/></opinion>*

It worked for me better than any other drug

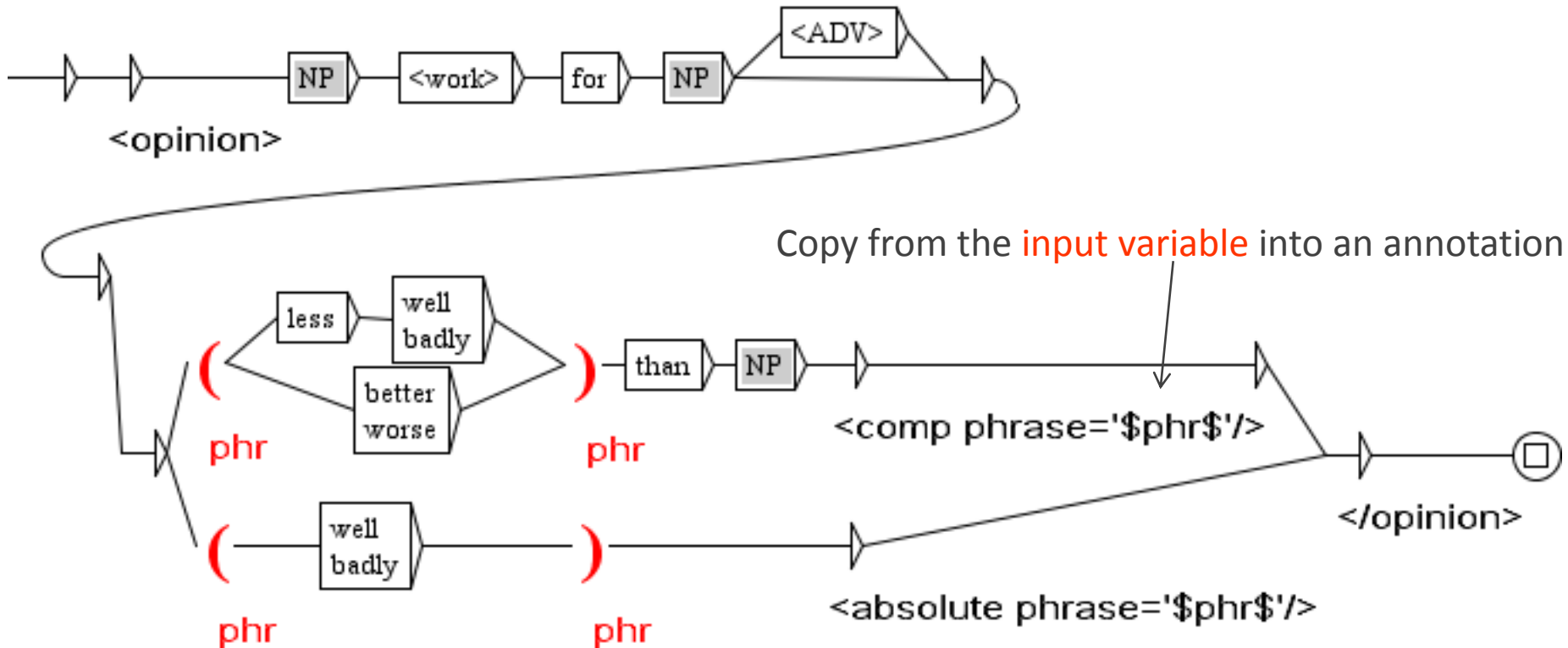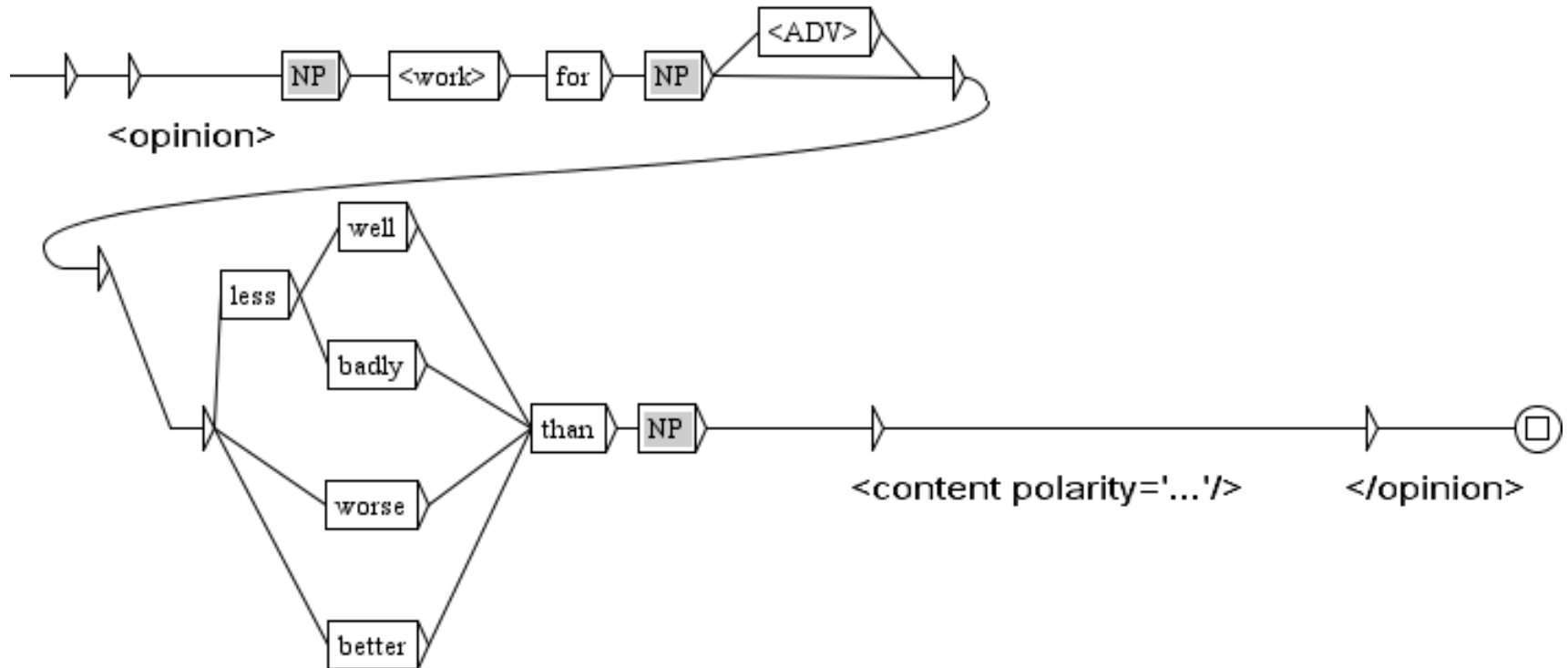*<opinion>It worked for me better than any other drug<content polarity='good'/></opinion>*

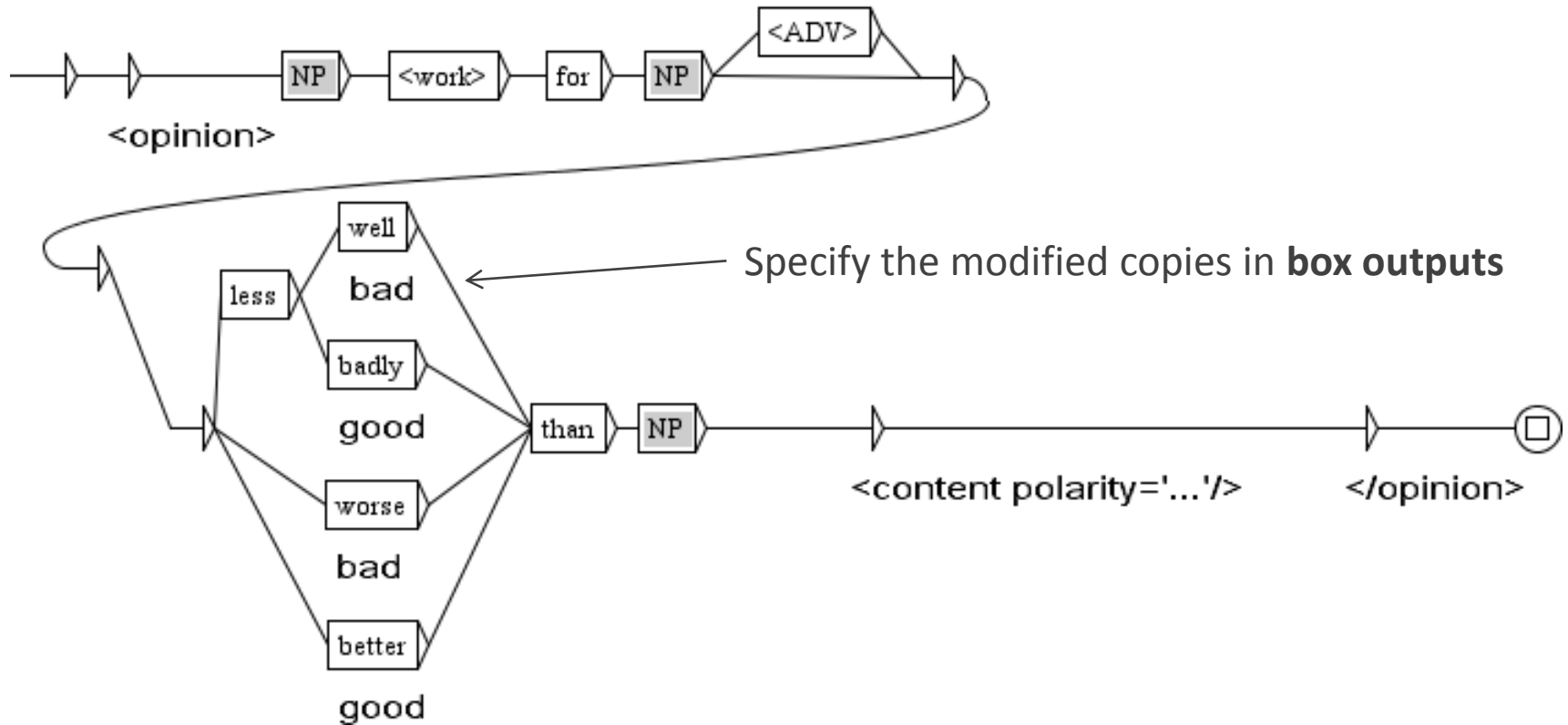Specify the modified copies in **box outputs**

Insert a modified copy

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<content polarity='good'/></opinion>

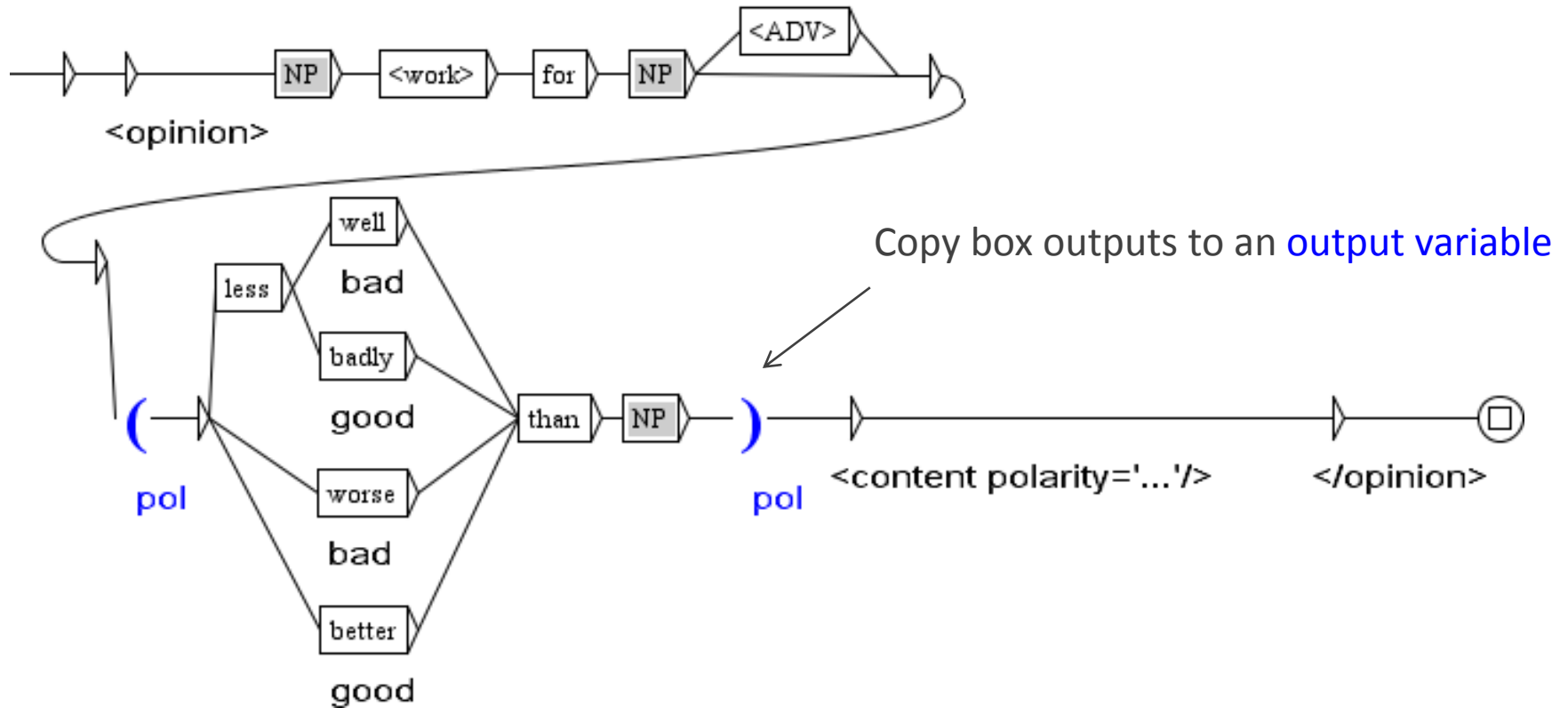Copy box outputs to an output variable

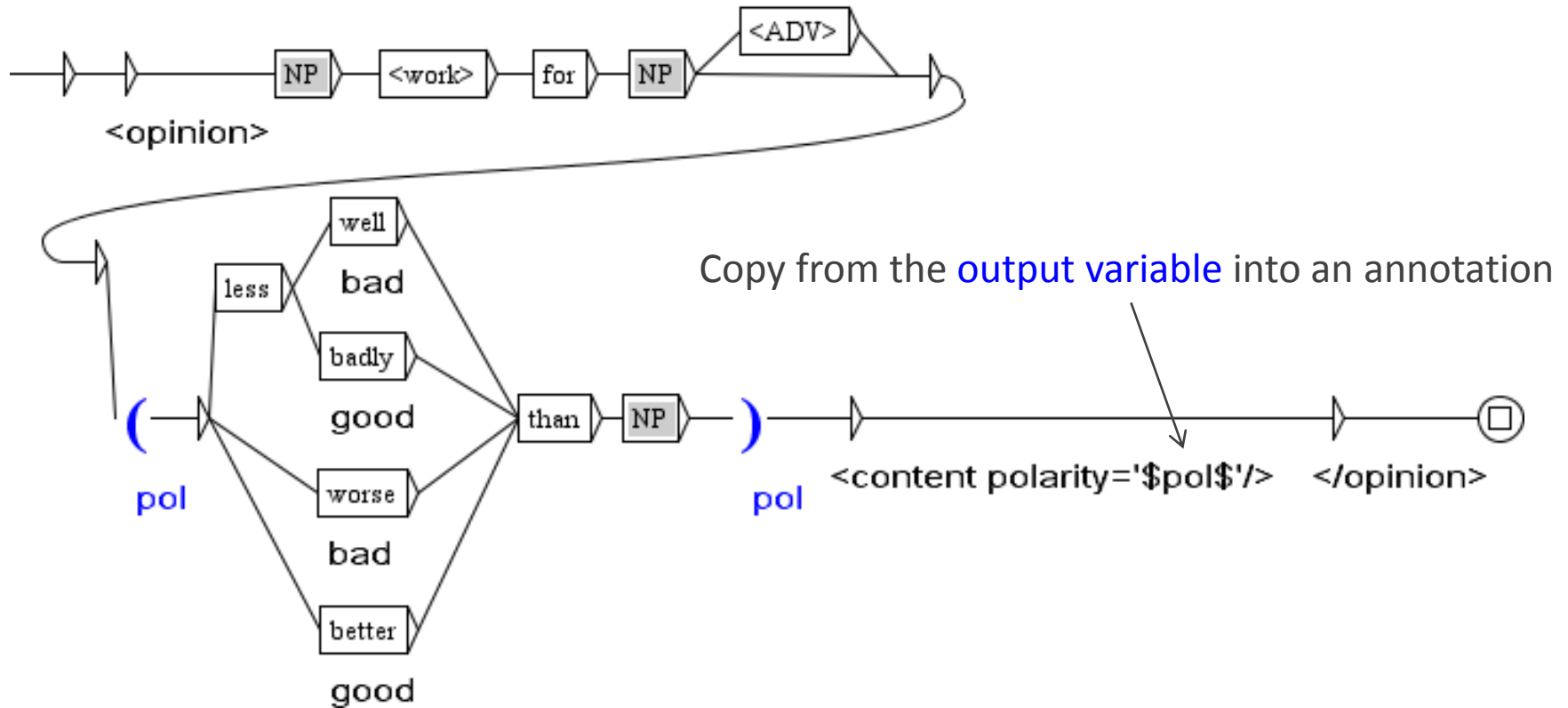# Insert a modified copy

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<content polarity='good'/></opinion>

Copy from the output variable into an annotation

Outline

Copy from a dictionary

It worked for my mental health quite poorly

<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Copy from a dictionary

It worked for my mental health quite poorly

<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Copy from the entries to a **dictionary-entry variable**

Copy from a dictionary

It worked for my mental health quite poorly
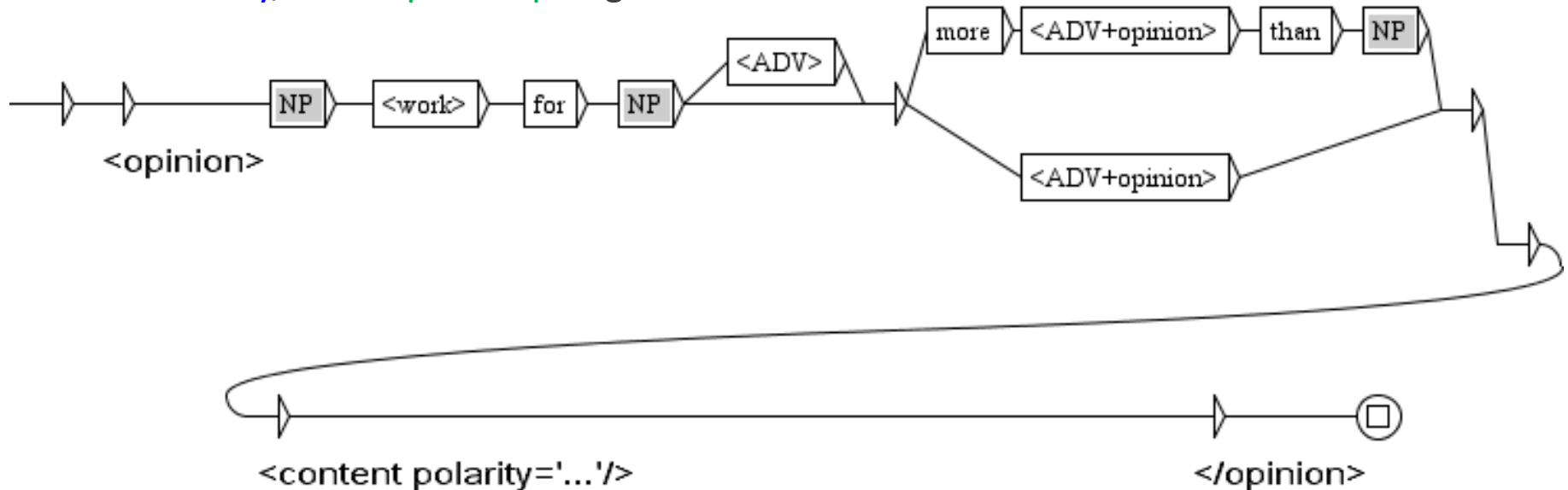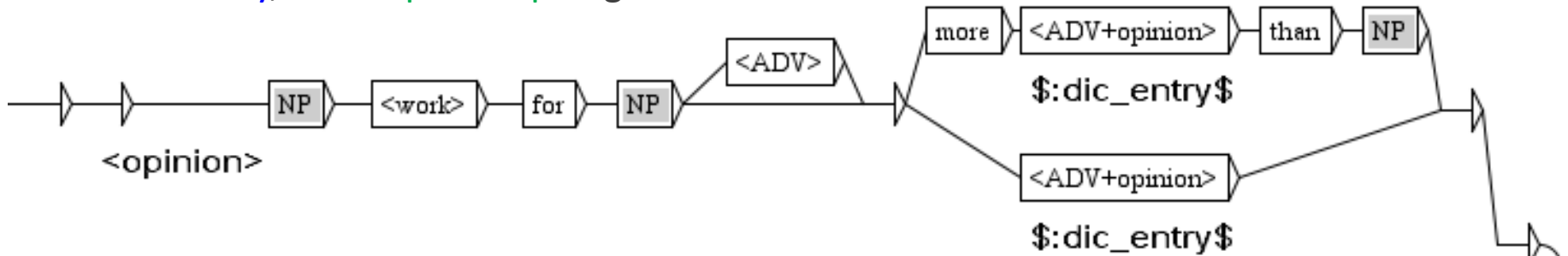
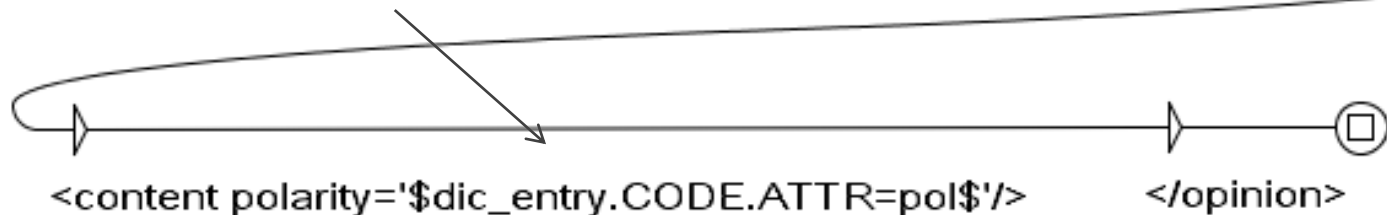*<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>*

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
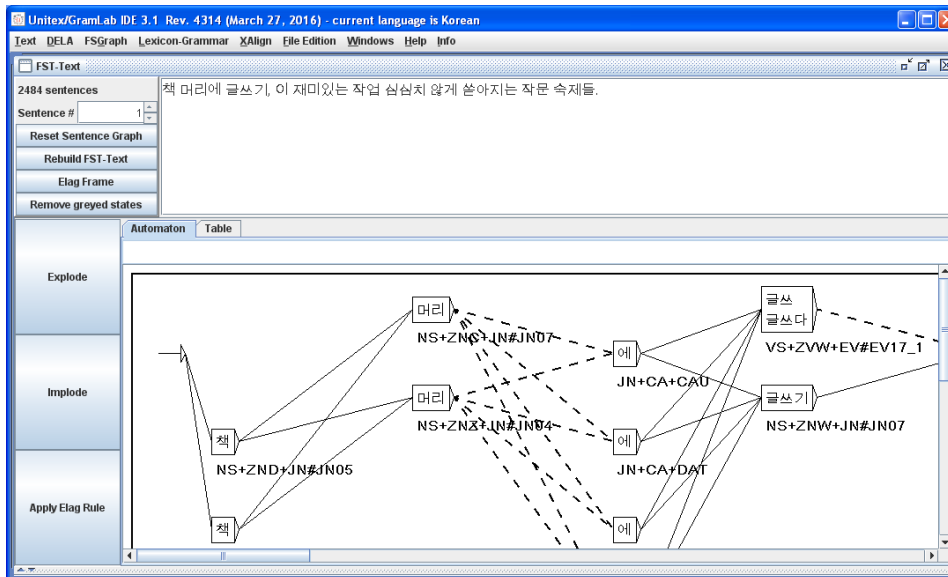satisfactorily,.ADV+opinion+pol=good

Copy from the **dictionary-entry variable** into an annotation

Dictionary-entry variables

Graphs with dictionary-entry variables must be applied:

- to the text automaton
- with the 'automaton intersection' option of Locate Pattern

Dictionary-entry variables
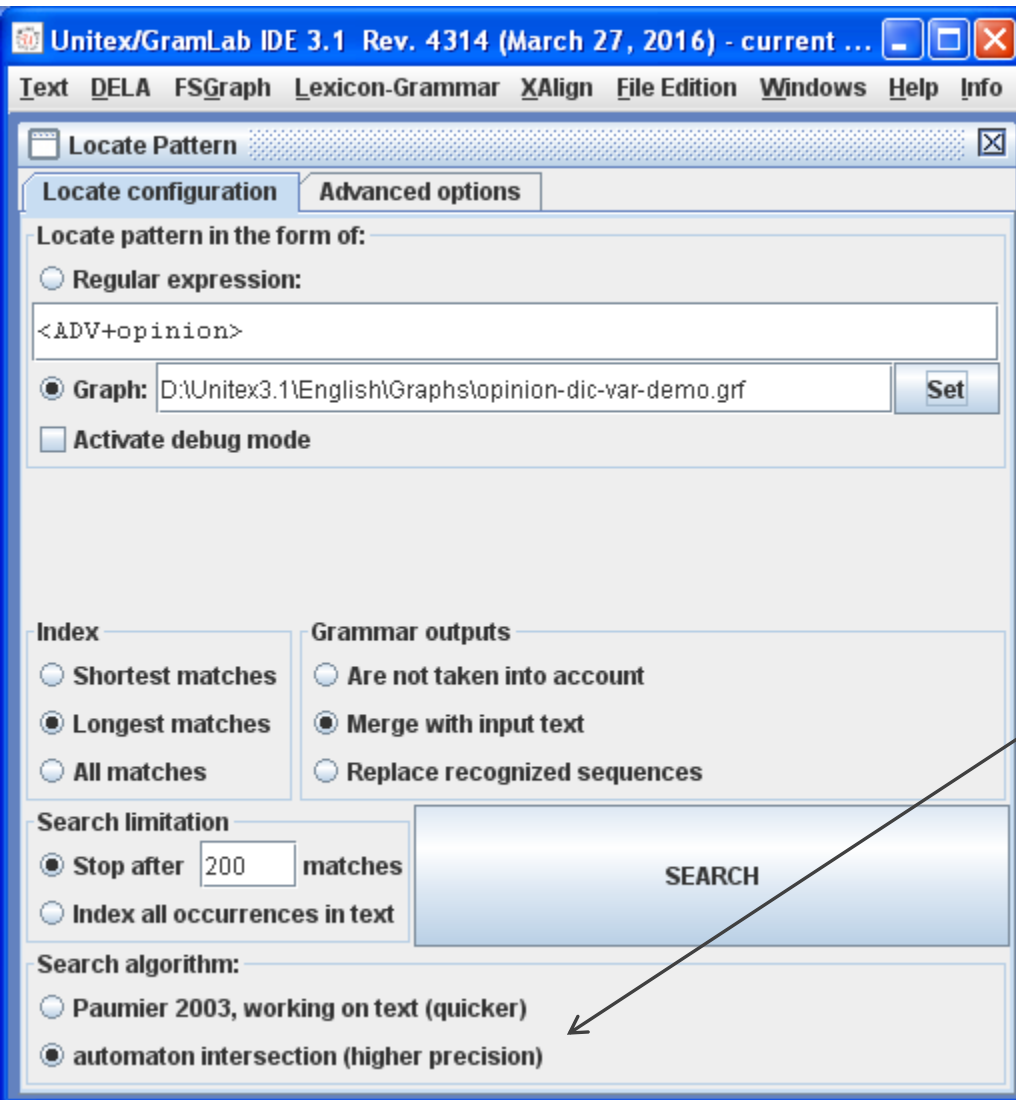

Graphs with dictionary-entry variables must be applied:

- to the text automaton

- with the 'automaton intersection' option of Locate Pattern

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Dictionary-entry variables
in the morphological mode

Graphs with dictionary-entry variables can also be applied without constructing the text automaton, but with other conventions

Dictionary-entry variables in the morphological mode

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

It worked for my mental health quite poorly
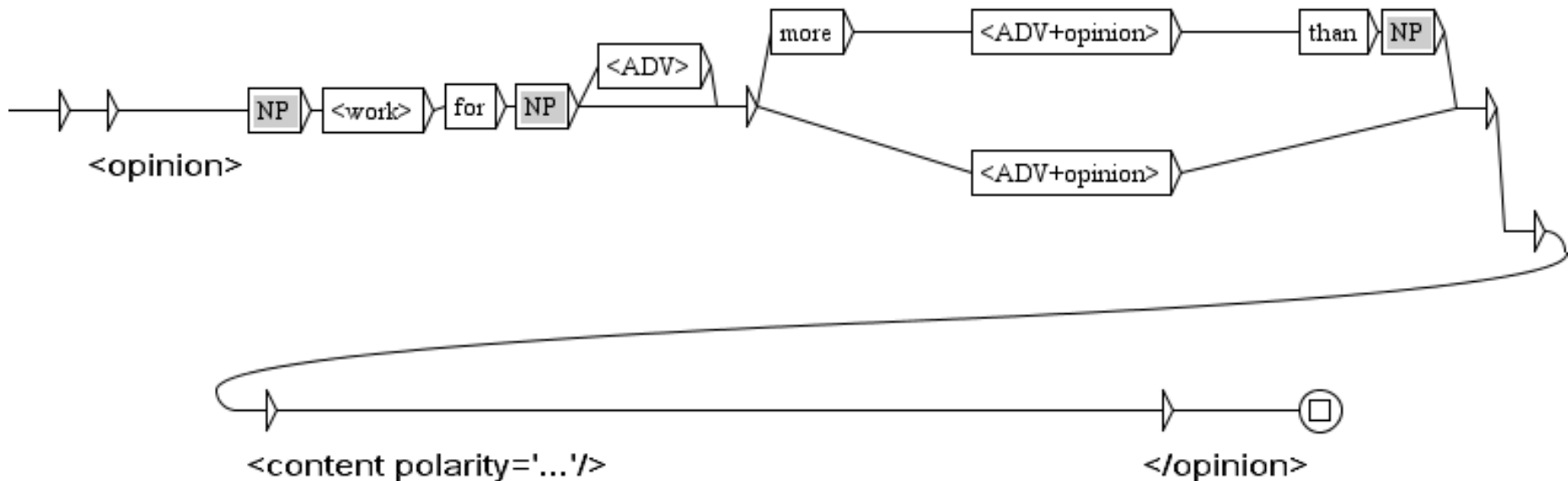
<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good
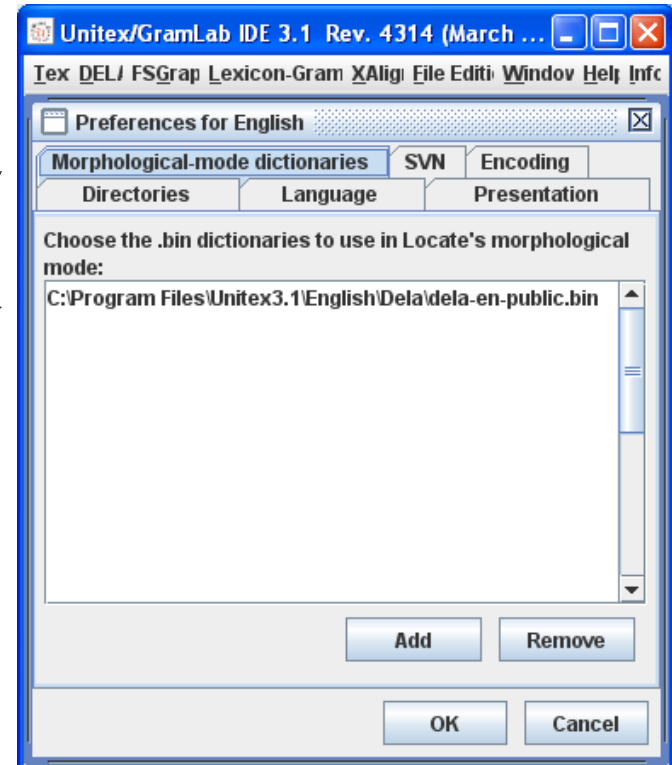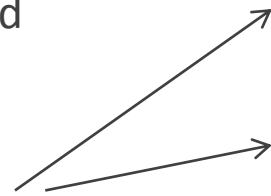
Dictionary-entry variables in the morphological mode

It worked for my mental health quite poorly

<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Declare the dictionary among the
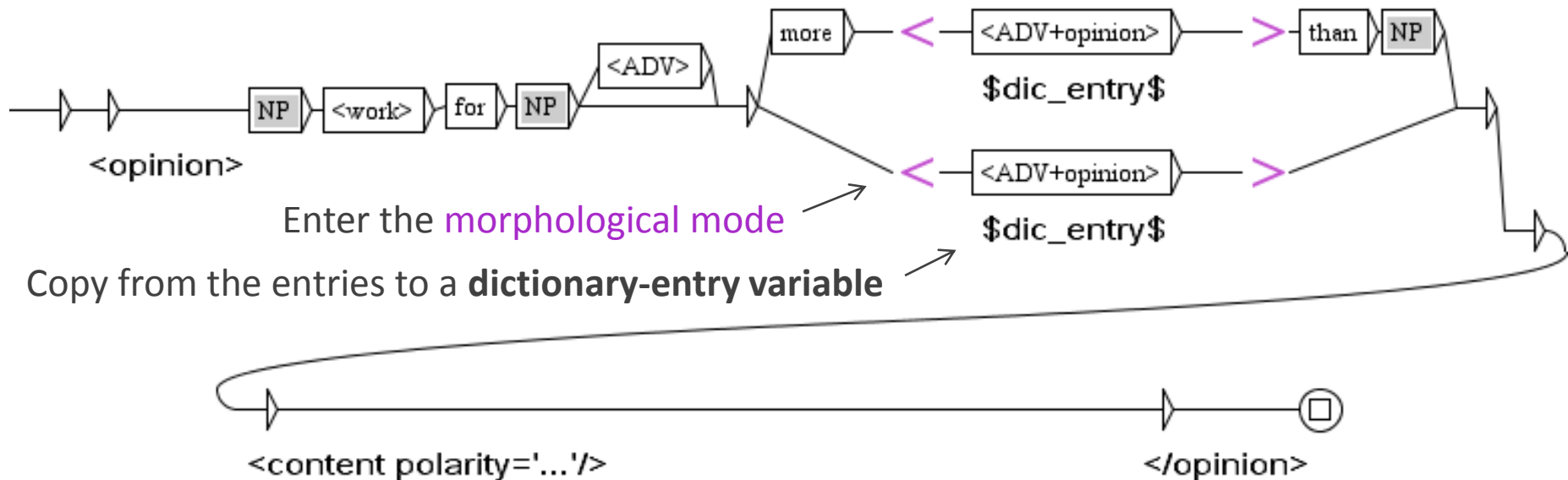   morphological-mode dictionaries in the
    Info > Preferences menu

Unitex/GramLab IDE 3.1 Rev. 4314 (March ...

Tex DEL/ FSGrap Lexicon-Gram XAlign File Editi Window Help Info

Preferences for English

| Morphological-mode dictionaries | SVN | Encoding |
| Directories | Language | Presentation |

Choose the .bin dictionaries to use in Locate's morphological mode:

C:\Program Files\Unitex3.1\English\Dela\dela-en-public.bin

Add   Remove

OK   Cancel

Dictionary-entry variables
in the morphological mode

It worked for my mental health quite poorly
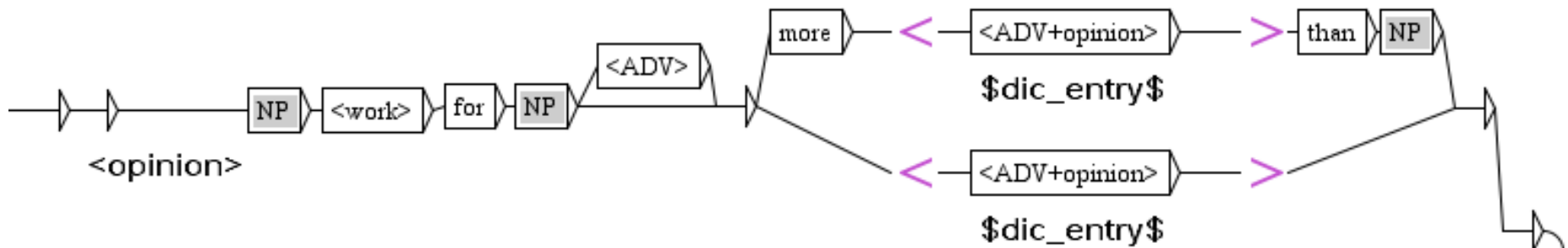
<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Enter the morphological mode

Copy from the entries to a **dictionary-entry variable**

Dictionary-entry variables in the morphological mode

It worked for my mental health quite poorly

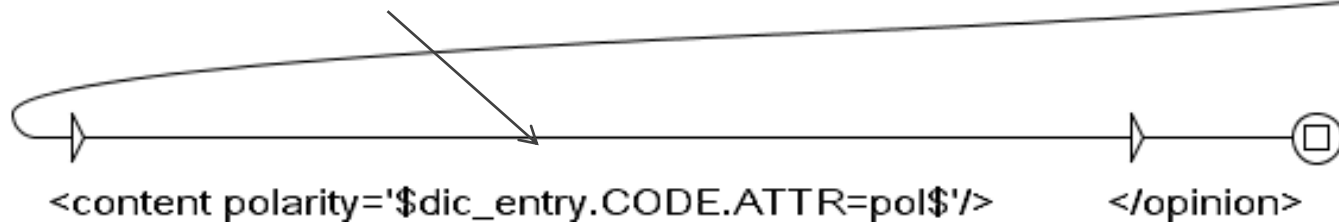*<opinion>It worked for my mental health quite poorly<content polarity='bad'/></opinion>*

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Copy from the **dictionary-entry variable** into an annotation

<content polarity='$dic_entry.CODE.ATTR=pol$'/>   </opinion>

Outline

Insert annotations

Copy from text to annotation

Copy from a dictionary

Annotate at the right place

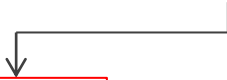Delete annotations

Annotate at the right place

It worked for me better than any other drug

<opinion>It worked for me better than any other drug<comp phrase='better'/></opinion>

It worked for me better than any other drug

<opinion phrase='better'>It worked for me better than any other drug</opinion>

start tag   attribute
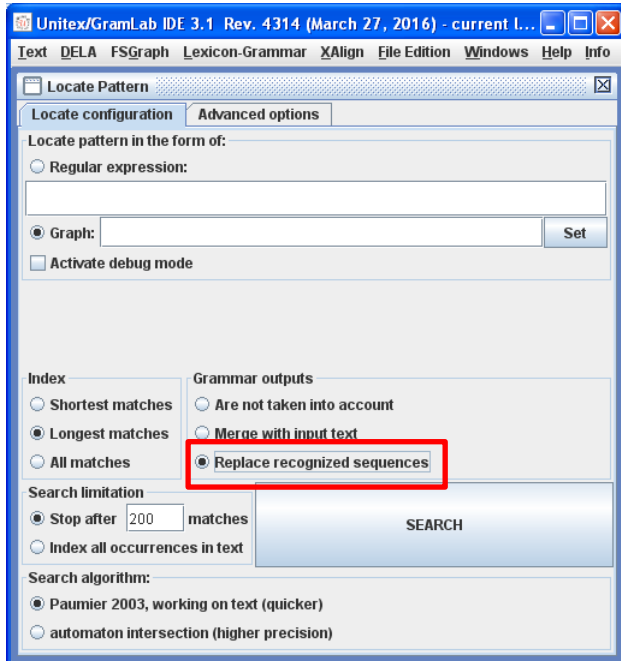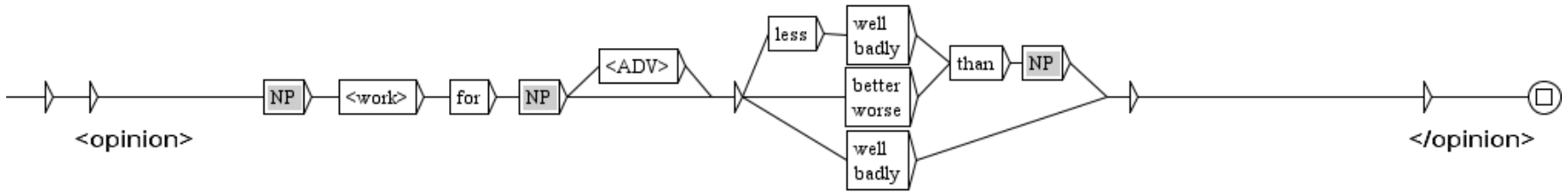
In XML, attributes are obligatorily placed in start tags

This implies we copy information to output in a different order than in input

The 'Merge with input text' mode

It worked for me badly and I stopped after three months

<opinion>It worked for me badly</opinion> and I stopped after three months

All the preceding examples were with the Merge mode

In this mode, Unitex tacitly copies the text it recognizes

This is simple but does not copy in a different order

We want to make copy explicit

The 'Replace recognized sequences' mode

It worked for me badly *and I stopped after three months*

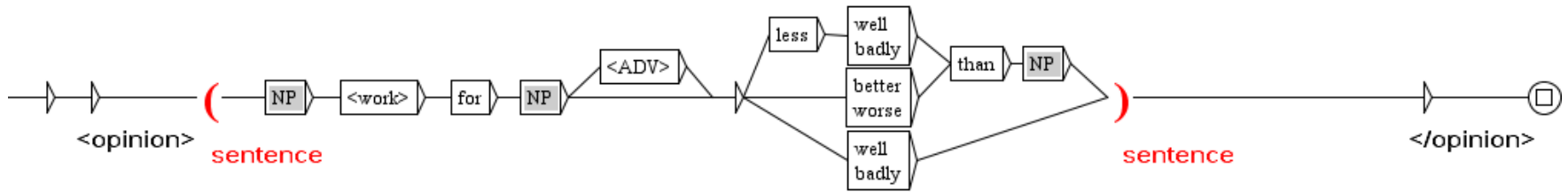*<opinion>It worked for me badly</opinion> and I stopped after three months*

Recognized sequences are replaced by outputs

The 'Replace recognized sequences' mode

It worked for me badly *and I stopped after three months*

<opinion>It worked for me badly</opinion> and I stopped after three months
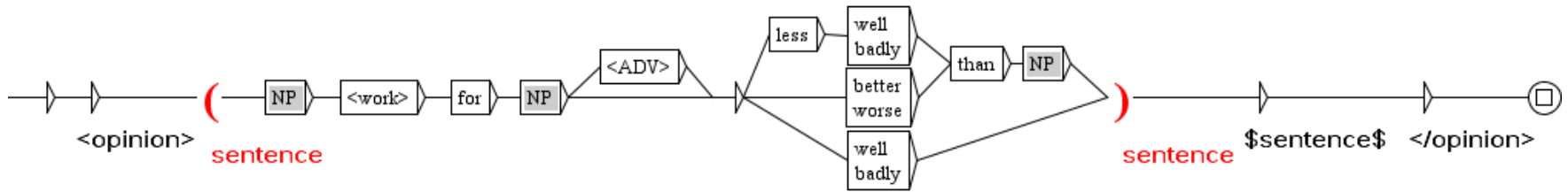

Explicitly copy from the text to an input variable

The 'Replace recognized sequences' mode

It worked for me badly *and I stopped after three months*

<opinion>It worked for me badly</opinion> and I stopped after three months
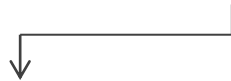

Explicitly copy from the input variable to the output text
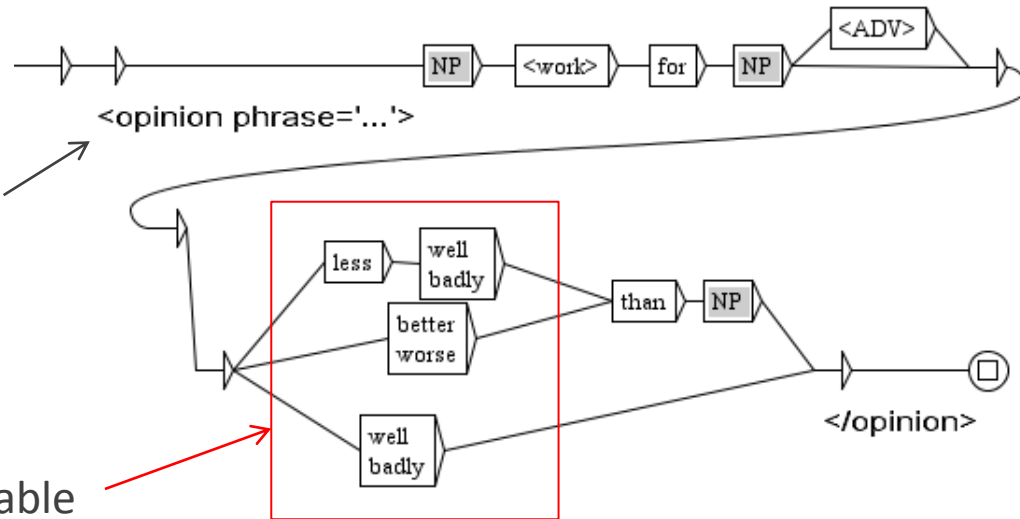
Copy into an annotation placed before

It worked for me better than any other drug

<opinion phrase='better'>It worked for me better than any other drug</opinion>

Place to use the variable

Place to set the variable

A local grammar must set a variable before it uses it
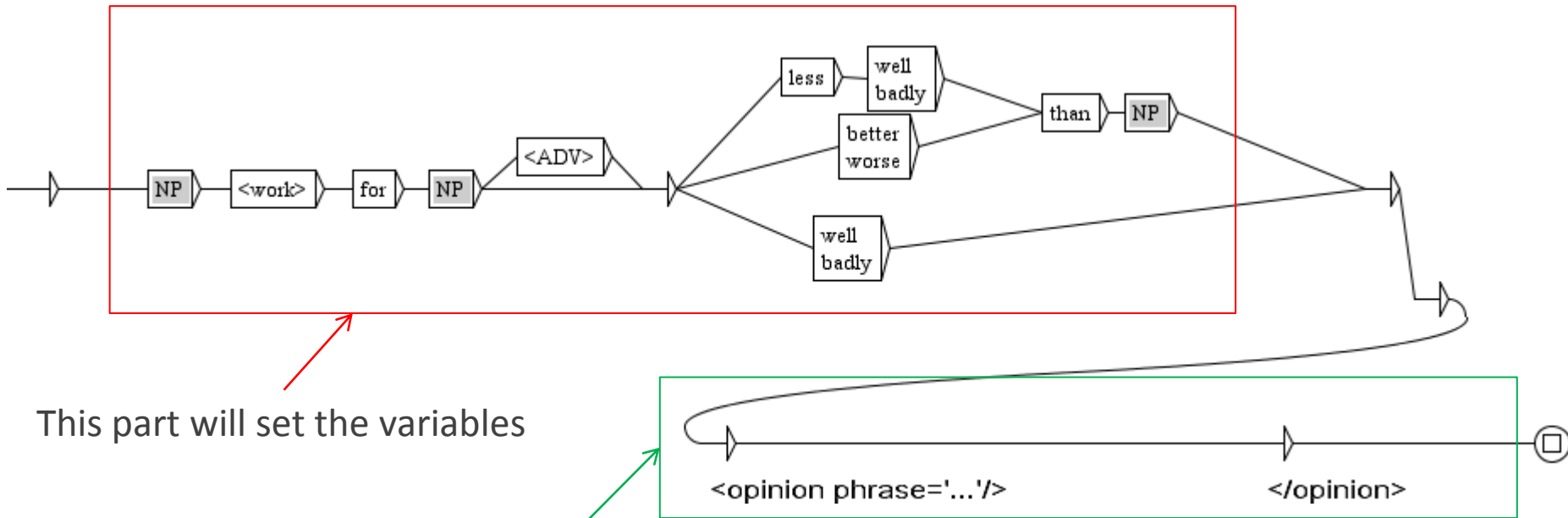
This solution will not work

Copy into an annotation placed before

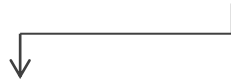*It worked for me better than any other drug*

<opinion phrase='better'>It worked for me better than any other drug</opinion>

This part will set the variables

This part will use them

Copy into an annotation placed before

It worked for me better than any other drug

<opinion phrase='better'>It worked for me better than any other drug</opinion>
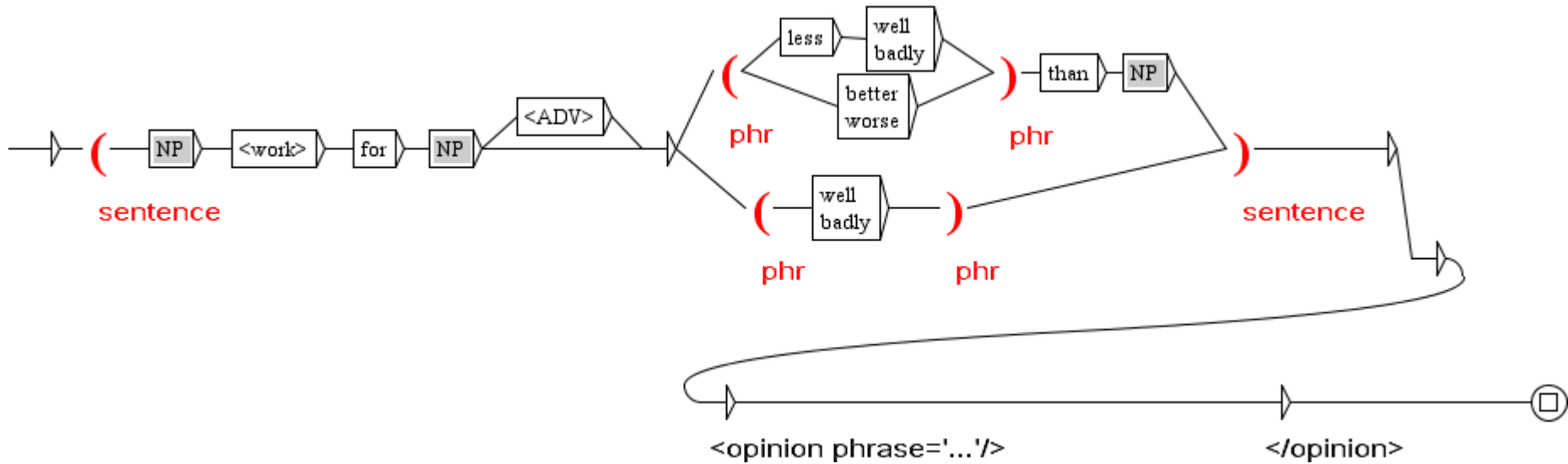
Copy into an annotation placed before

It worked for me better than any other drug

<opinion phrase='better'>It worked for me better than any other drug</opinion>
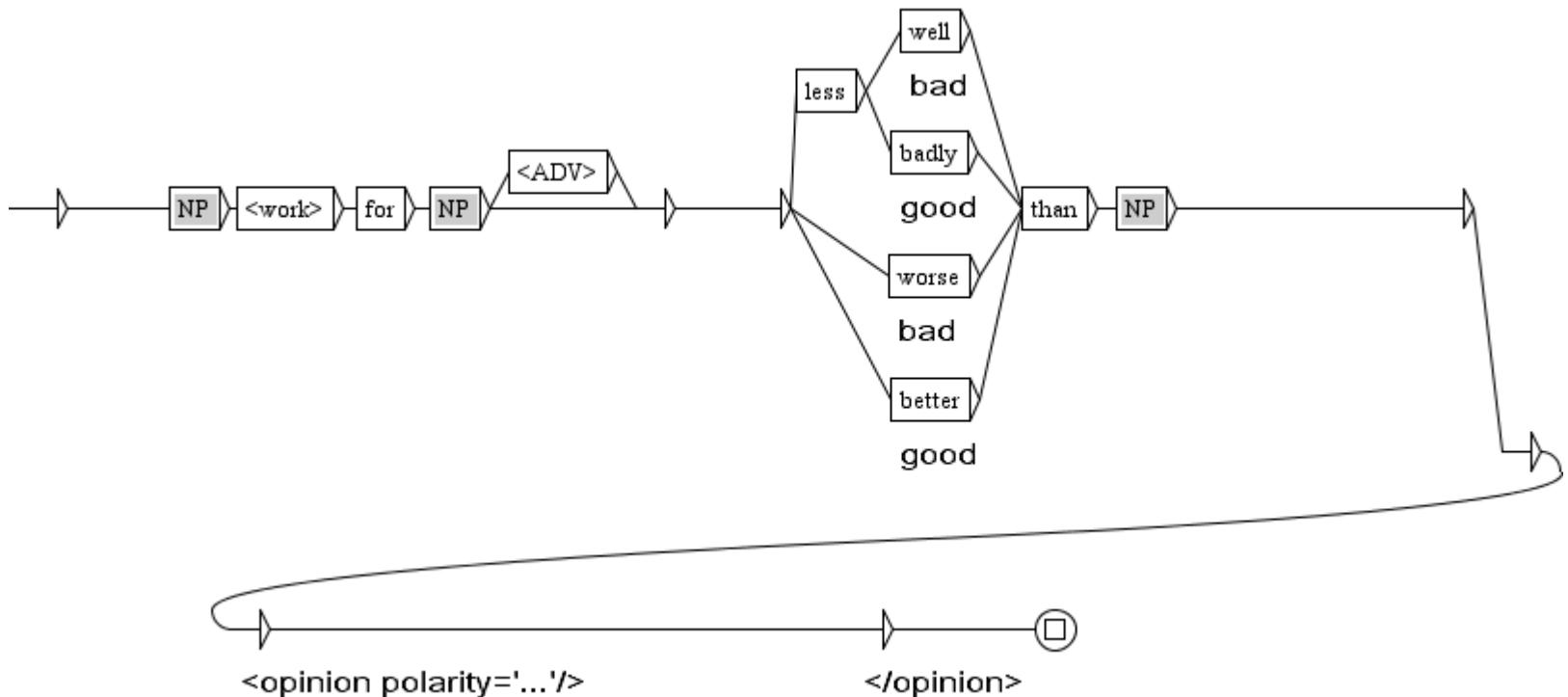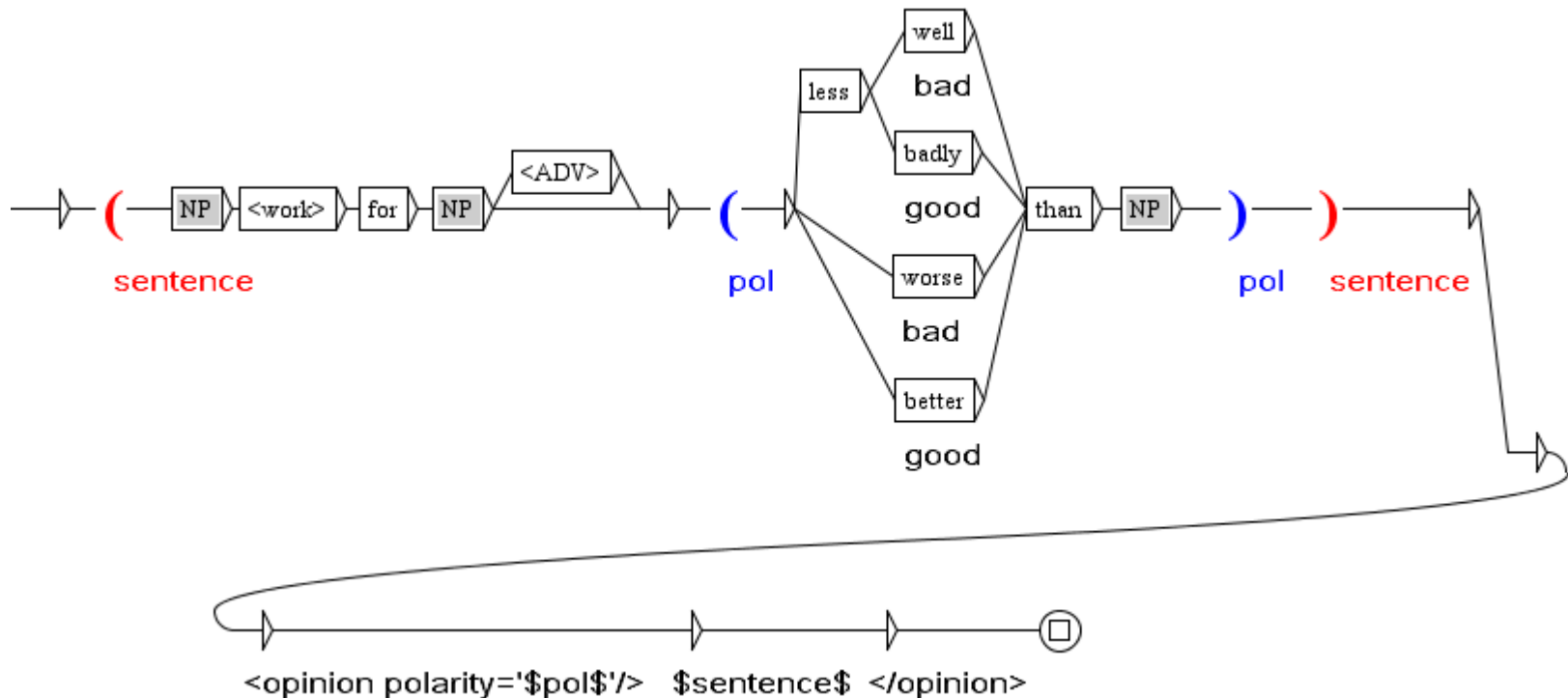
Insert a modified copy
into an annotation placed before

It worked for me better than any other drug

<opinion polarity='good'>It worked for me better than any other drug</opinion>

Insert a modified copy
into an annotation placed before

It worked for me better than any other drug

<opinion polarity='good'>It worked for me better than any other drug</opinion>

Copy from a dictionary
into an annotation placed before

It worked for my mental health quite poorly

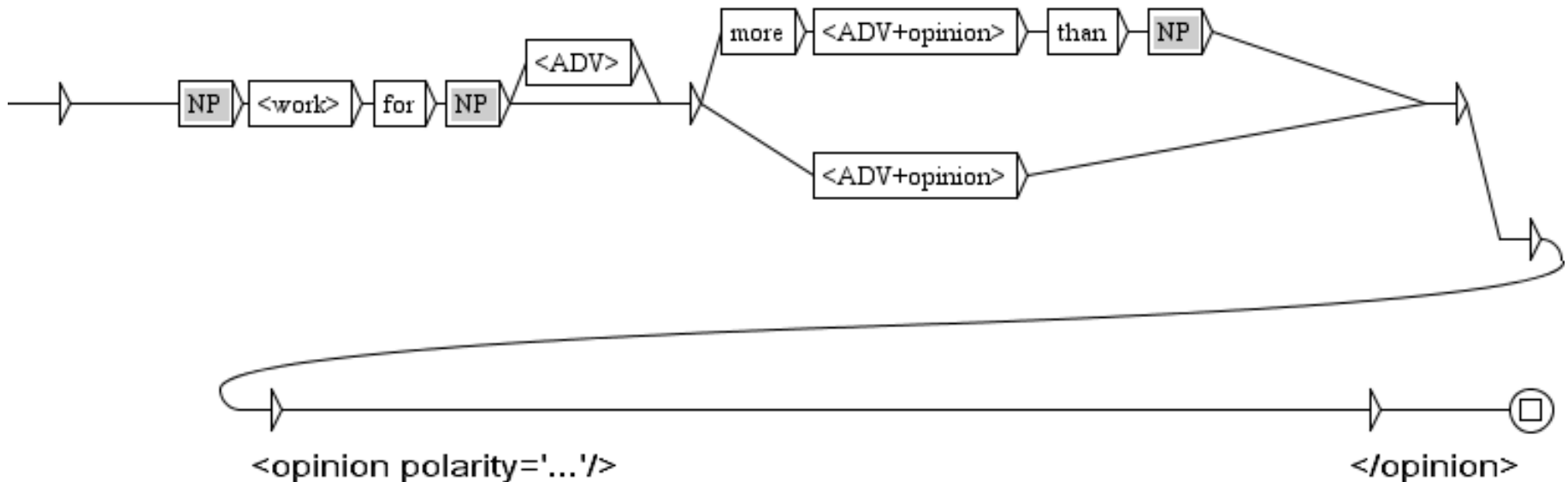*<opinion polarity='bad'>It worked for my mental health quite poorly</opinion>*

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

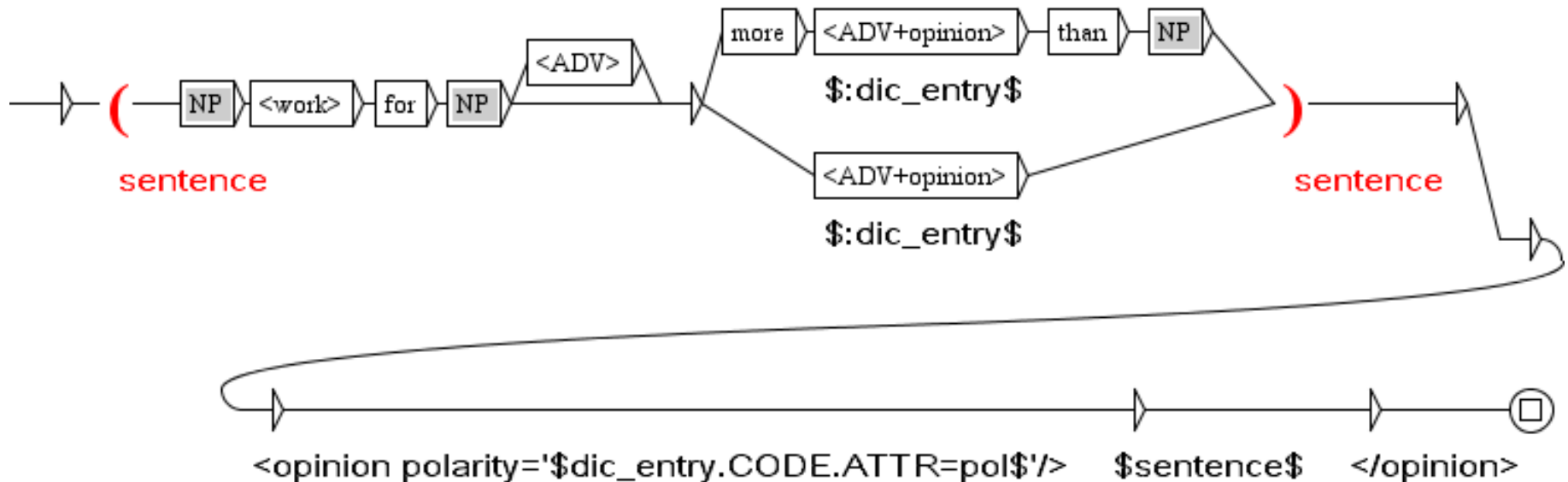*It worked for my mental health quite poorly*

<opinion polarity='bad'>It worked for my mental health quite poorly</opinion>

inefficiently,.ADV+opinion+pol=bad
poorly,.ADV+opinion+pol=bad
satisfactorily,.ADV+opinion+pol=good

Outline

Insert annotations

Copy from text to annotation

Copy from a dictionary

Annotate at the right place

Delete annotations

Delete annotations

<opinion polarity='good'>It worked for my mental health better than any other drug</opinion>

↓

It worked for my mental health better than any other drug

Clearly distinguish annotations from text

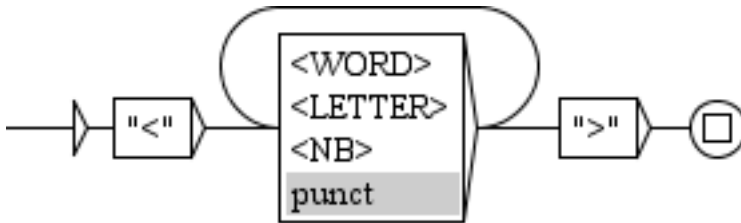They can be deleted so as to retrieve the original corpus

Delete annotations

<opinion polarity='good'>It worked for my mental health better than any other drug</opinion>

↓

It worked for my mental health better than any other drug

Use a graph that recognizes tags

Use it in the Replace mode

Recognized sequences are replaced by outputs

If there are no outputs, recognized sequences are skipped

Thanks

CONTACT

ÉRIC LAPORTE

00 +33 (0)1 60 95 75 52

ERIC.LAPORTE@UNIV-PARIS-EST.FR