

Synpathy Syntax Editor

Manual



Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
This manual was updated on 29-11-2006

Table of contents (click the name of the chapter to skip to that section)

Notation conventions

- 1 Introduction
 - 1.1 Background
 - 1.2 Installing Synpathy
- 2 User interface
 - 2.1 Options toolbar
 - 2.1.1 File menu
 - 2.1.2 Edit menu
 - 2.1.3 Graph menu
 - 2.1.4 View menu
 - 2.1.5 Options menu
 - 2.1.6 Help menu
 - 2.1.7 About
- 3 Editing
- 4 Outlook
- 5 Concepts and definitions

Notation conventions



The following notation conventions are used:

- Screens, schemata and fields on web pages are written in *italic*.
- Information on troubleshooting starts as follows: 

1 Introduction



Synpathy is a tool for annotating, analyzing, and graphically editing the syntactical structure of sentences (e.g. linguistically annotated text corpora), developed at the [Max Planck Institute for Psycholinguistics](http://www.mpi.nl), Nijmegen, the Netherlands. The application is based on the SyntaxViewer from the TIGER search project developed by the IMS (Institute für Maschinelle Sprachverarbeitung, University of Stuttgart).

Since all (non) terminal node features values are user definable a wide range of linguistic descriptions like syntax trees, functional structures, dependency-style structures or predicate-argument structures can be accommodated. The annotated text together with its treebank graph information is stored separately from the list of labels used in the graph (features). Output formats are in persistent TIGER-XML. This facilitates the further processing of the data by other linguistic applications (like [ELAN](#) and [ANNEX](#)).

The Synpathy project webpage can be found on <http://www.mpi.nl/tools/synpathy.html>. Supported platforms are Microsoft Windows (9x, NT, 2000, XP), Mac OS X and Linux. Java Runtime Environment (at least 1.4) has to be present on the machine. All code is open source and free to use by non-commercial institutions.

1.1 Background



Synpathy's Java-based graphical syntax viewer presents syntactic trees in an interactive interface, allowing the user to view the syntactic structure of a sentence and to manually manipulate ("rebuilding" and "retagging") the relationships between the different elements of the text. Each element in the sentence can be represented by a node in the syntax graph or annotation field thus showing both form and function information.

The default output format of Synpathy is in the TIGER-XML format (.tig extension). This is the treebank exchange format agreed upon by the [Nordic Treebank Network](#), allowing free data exchange and the use of tools developed by the international [TIGER project](#) community. In the TIGER format, edge labels contain the original syntactic function tags, and the (non-terminal) cat category contains phrase and clause forms. A TIGER-XML file consists of a header and a body. The corpus header can contain meta-information about the corpus (such as corpus name, date, author, etc) and a declaration of the tags that are used in the morphology Part-of-Speech, non-terminal nodes and edges. In the second part of a TIGER-XML file the corpus body contains words, Part-of-Speech tags, morphological tags and lemmata which are listed as attributes of the element "terminal". Non-terminals are represented in an additional element called "nonterminal" referring to the corresponding terminal ID. This part of the XML file contains the encoding for secondary edges as well. (the [User interface](#) Chapter gives some more information about the terms used above).

A big advantage of using the xml format is its exchangeability and usability with a large range of other applications. For example the [TIGERsearch corpus query tool](#), and in the future multimedia annotation tools as [ELAN](#) and [ANNEX](#).

1.2 Installing Synpathy

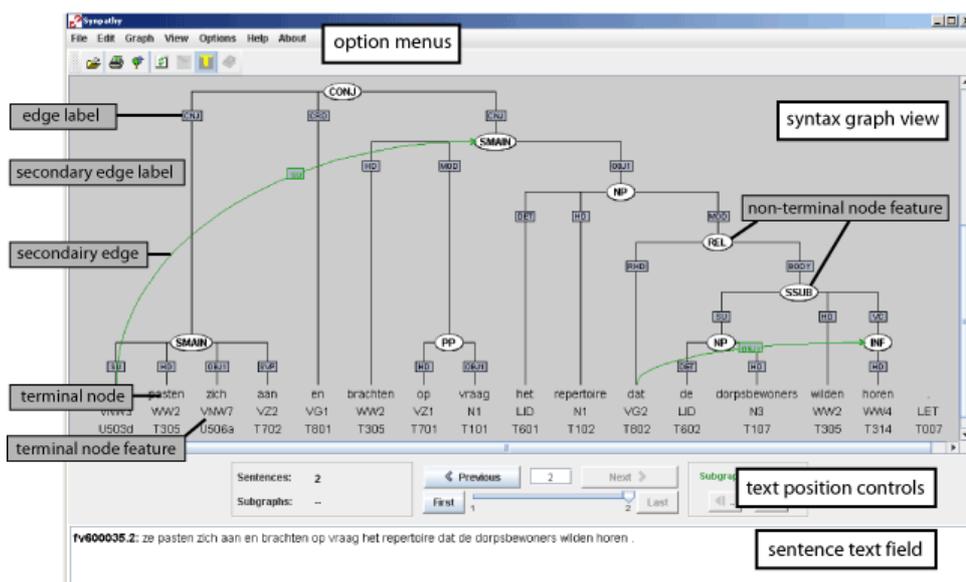


Go to <http://www.mpi.nl/tools/synpathy.html> in your browser and click *choose installer*. A new window will present the recommended version of Synpathy to match your operating system (Windows, OS-X or Linux). Click the *installer* button to start the download and follow the instructions of the installer. Optionally you can download Synpathy without Java Virtual Machine. In this case at least Java VM 1.1.8 has to be present on your system to be able to run Synpathy.

2 User interface



The next picture shows the main Synpathy window:



The *option menus* which can be selected from the top toolbar offer various functions for importing/exporting, editing features/labels- and preferences.

The *syntax graph view* shows the connections between the different morphemes/words from the input sentence (which is displayed as well in the sentence text field) by means of line graphics between related units. Each line is tagged (non-terminal node feature, edge label). The parts which make up the annotated sentence are called terminal nodes, and can contain Part of Speech, glosses or other tags. The collection of all possible active tags (terminal and non-terminal) is called a "feature". and is stored in a separate "feature" XML file separate of the text corpus.

The buttons/field in the *text position controls* area of Synpathy are used to select the current active sentence (shown in the *sentence text field* and in the *syntax graph view*) from a larger text or corpus. Enter the number of the desired sentence in the field to directly skip to that part of the text.

The next chapters of this manual will explain the functions found in these three parts of the main Synpathy window. If the mouse pointer is placed over a node (token or inner node), a tooltip window appears which comprises the annotation of all features.

2.1 Options toolbar

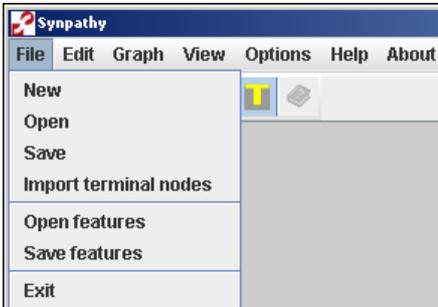


The functions found in the menus of the options toolbar are explained in the next chapters.

2.1.1 File menu



The following options are accessible using the File menu from the options toolbar:



- **New:** clears all data except current features.
- **Open:** opens a search window for a TIGER-XML file containing the annotated text (terminal) and associated graphs (non-terminal).
- **Save:** saves the current text corpus and associated graphs in the TIGER-XML format.
- **Import terminal nodes:** opens a search window for (annotated) text without associated graph. The default file format is .txt (plain text file). It should have the following block structure for each sentence, followed by an empty line:

```
you go out of the institue
pro v adv prep art n
```

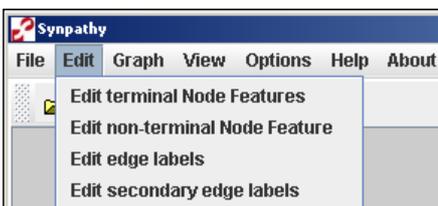
Apart from the words each block can contain an arbitrary number of additional lines, representing further terminal node features like "Part-of-Speech", etc.

- **Open features:** opens a search window for a xml file containing definitions of possible tags for terminal as well as non-terminal nodes. Examples of possible labels are: the [Dutch CGN tagset](#), and the [Penn Project Treebank tagset](#).
- **Save features:** saves the current active tagset in a TIGER-XML formatted file.
- **Exit:** shuts down Synpathy.

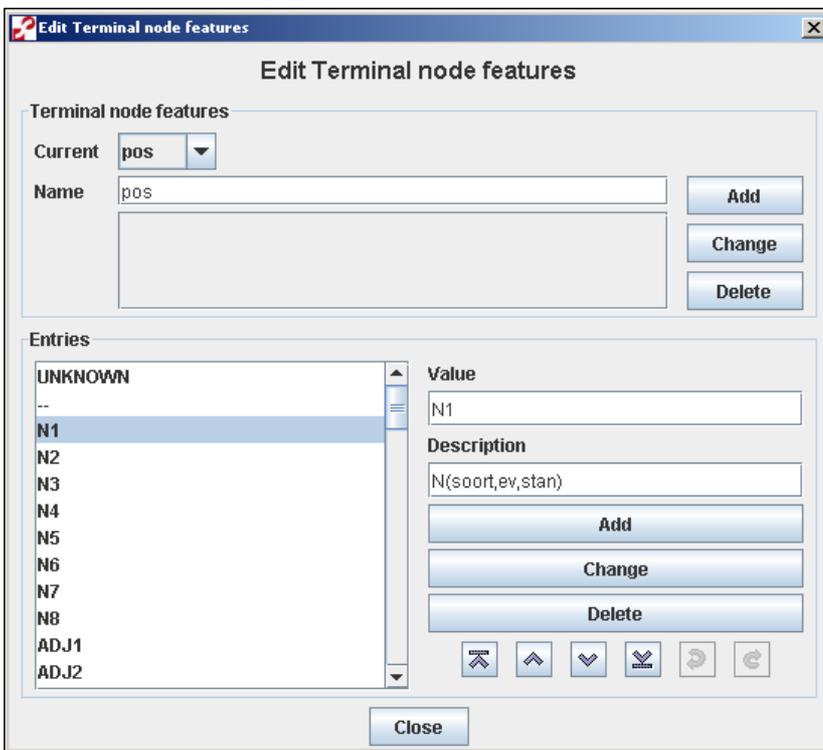
2.1.2 Edit menu



The Edit menu is used to manage the contents of terminal- and non-terminal features and labels.



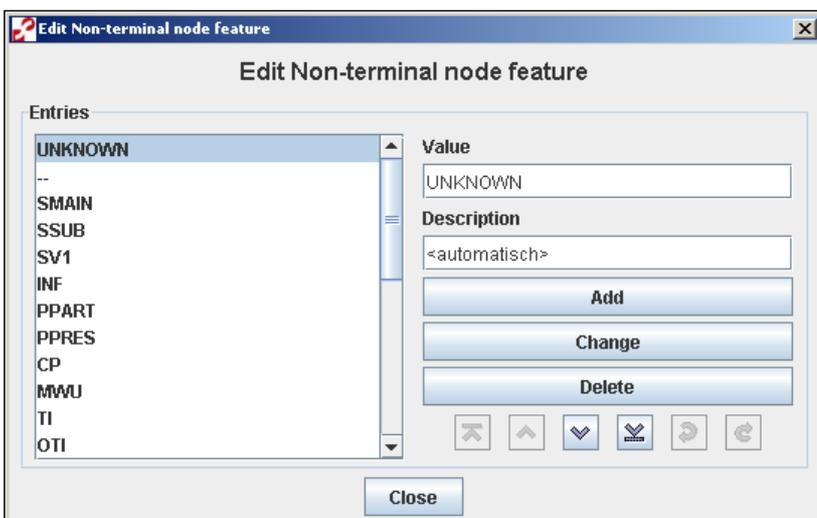
- **Edit terminal Node Features:** opens a window for managing terminal node tags. These are the tags used in the annotation of the text. The list of terminal node features is opened/saved by the corresponding option from the File menu.



The top part of this window is used to select the tier in the annotation (in the example above POS, "Part of Speech") from the pulldown list. Click the  button and choose the tier whose entries needs to be managed. To add/rename or delete a new terminal node feature click the corresponding button on the right after entering the name in the Name field. Each tier may contain different entries. These can be edited in the lower part of the window. Provide Value name and optional Description. Click the Add button to append the new entry to the list. To rename an entry click the Change button after typing the new name in the Value field. The Delete button removes the selected entry from the list. The following square buttons are used to navigate though the list of entries:

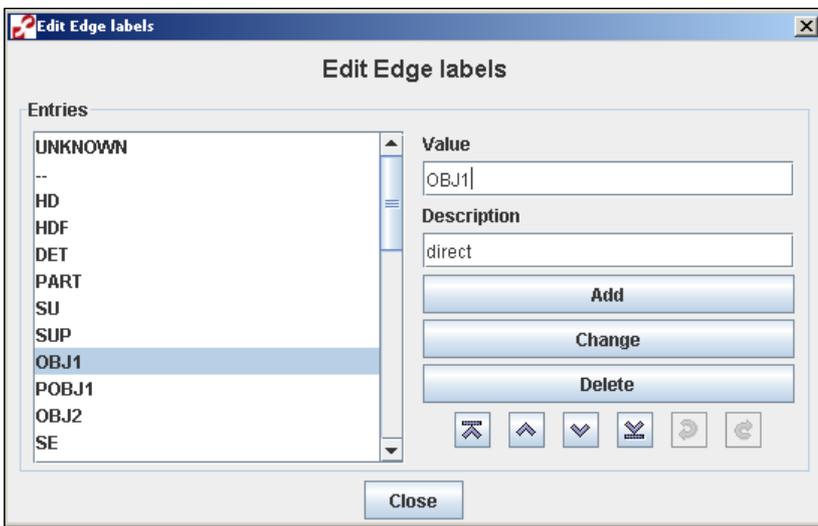
-   Skip to the top/bottom of the list.
-   Go up/down one entry.
-   Undo/redo previous action. Only your last action can be revoked.

- Edit non-terminal Node Feature: opens a window for managing non-terminal node tags. These appear in the treebank graph in the form of for example  or . They are used to label the edges which define the relation between the different terminal nodes which together make up the sentence. The list of these tags (non-terminal node features) is opened/saved by the corresponding option from the File menu.



To add/rename a new non-terminal node feature click the corresponding button on the right after entering it's Value name and Description in the two corresponding fields. The Delete button removes the selected entry from the list. The same square buttons as mentioned above are used to navigate though the list of entries.

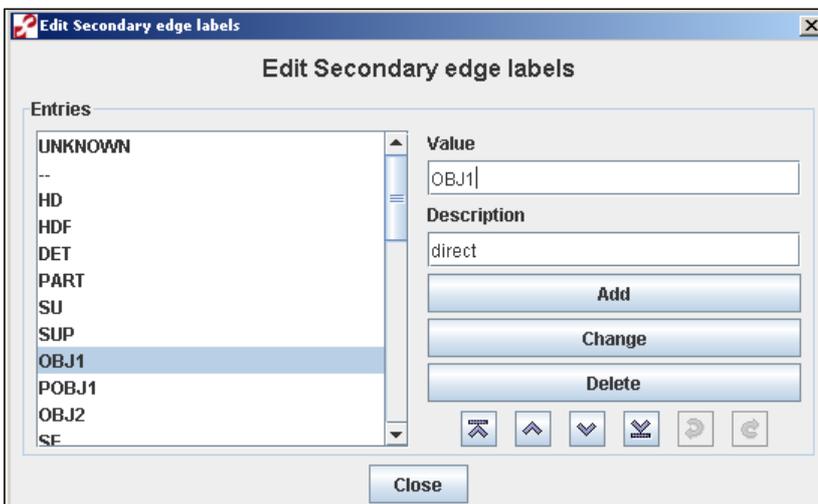
- Edit edge labels: opens a window for managing edge labels. These appear in the treebank graph in the form of for example  or . The tags are used to describe the relationship between different parts of the sentence. Entries are edited in the following window:



To add/rename a new edge label click the corresponding button on the right after entering it's name and description in the two corresponding fields. The "Delete" button removes the selected entry from the list. The same square buttons as mentioned above are used to navigate through the list of entries.

- Edit secondary edge labels: opens a window for managing edge labels. These appear in the treebank graph in the form of for example `OBJ1` or `SU`. Secondary edges are used as an extra layer in order to allow for structure sharing in syntax graphs. They define an additional parent node for a given node.

Entries are edited in the following window:

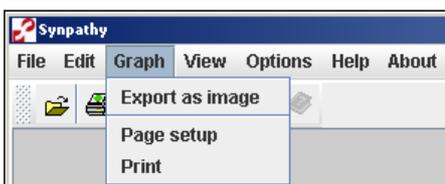


To add/rename a new secondary edge label click the corresponding button on the right after entering it's name and description in the fields. The Delete button removes the selected entry from the list. The same square buttons as mentioned above are used to navigate through the list of entries.

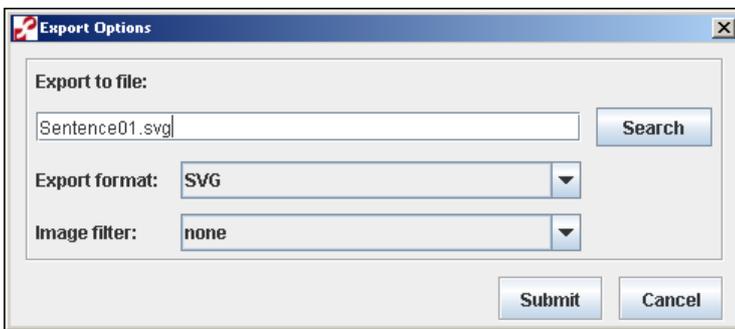
2.1.3 Graph menu



The Graph menu offers options to export the current treebank graph as an image or get a hardcopy on paper.



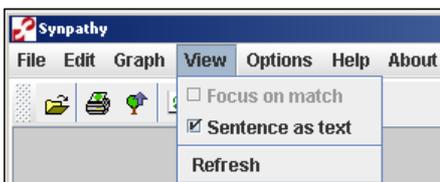
- Export as image: save the current treebank view as an image file. This feature can also be activated by clicking the  icon in the toolbar.



Enter the name of the new image in the specified field, or use the Search button to select an existing image that needs to be replaced. The Export format pulldown menu offers the choice between SVG, TIF or PNG image formats. The SVG format which is an XML-based vector graphics format. In contrast to binary formats such as JPG, images encoded using SVG can be manipulated (e.g. scaled) without loss of quality. Thus, you can change a token or a syntactic category within the exported SVG file. If you selected a non-SVG image, first an SVG representation is generated and afterwards converted into the preferred format. The SVG output has been designed in such a way that the different node types (inner/outer node, matching node, imploded node etc.) can be identified by additional XML attributes. Thus, this information can be used to modify the generated SVG image, e.g. by Cascading Stylesheets (CSS). The image filters you can choose (cf. screenshot above) are realized by inserting predefined CSS stylesheets into the SVG document. The embedding of stylesheets in SVG documents is explained in the SVG specifications located at the [W3C web site](#). In the current version of Synpathy the export functionality is not yet properly implemented.

- Page setup: gives some general print options like paper size, orientation and margins.
- Print: shows the print dialogue.

2.1.4 View menu

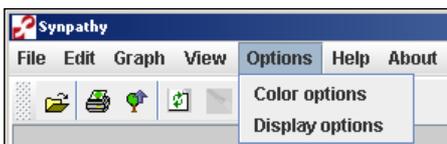


- Sentence as text: turns the Sentence text field on the bottom of the window on or off.
- Refresh: redraws the treebank graph.

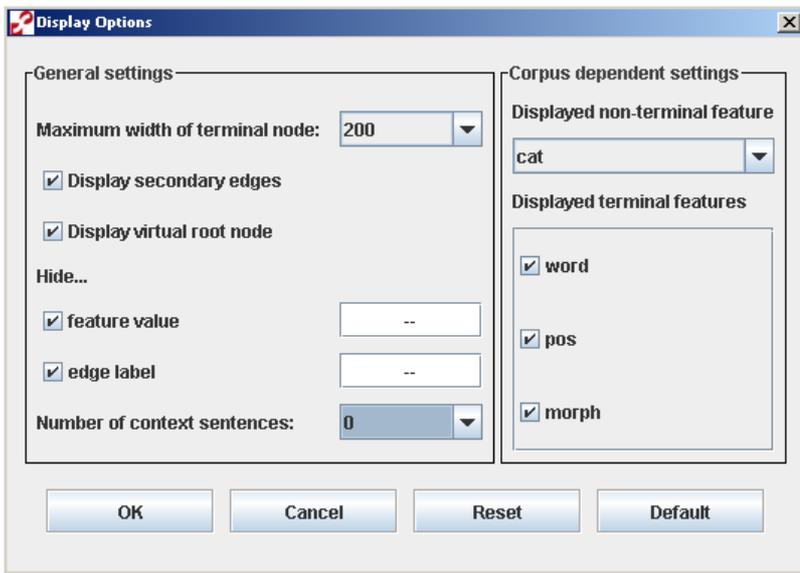
2.1.5 Options menu



The Options menu contains the following two options:



- Color options: determines the colors used in Synpathy for the different elements. To change a color click the corresponding Choose button. The color selection window will appear, you can now determine the new color using one of the three different tabs. Each tab offers a different method to define the new color.
- Display options: manages display parameters.



The following general parameters can be modified:

- Maximum width of terminal nodes: The maximum width of the terminal node visualization is limited. Thus, some feature values (especially word forms) may be truncated. If this occurs too often, just specify a higher width. from the pulldown menu.
- Display secondary edges: In order to turn on/off the display of secondary edges check/uncheck this box.
- Display virtual root node: Some corpus formats such as the Negra format do not integrate special types of tokens (e.g. punctuations marks) into the graph annotation. Unfortunately, the TIGERSearch query processing algorithm only works on connected graphs. Hence, during the indexing process a virtual root node is inserted in any graph that comprises terminal nodes (e.g. punctuations marks) which are not integrated into the corresponding syntax graph. In order to turn on/off displaying of virtual root nodes check/uncheck this box.
- Hide...: Sometimes annotating a token feature does not make sense for features such as morphological feature or case. In this case, a qualified symbol such as -- is used instead. We recommend you to use --. Since these symbols do not represent annotation information, it makes sense to hide them in the visualization. The two boxes must be checked if you want to hide a special feature value or edge label. Please specify the symbol which should be suppressed in the respective text fields.
- Number of context sentences: One or more context sentences can be displayed before/after the textual representation of the current sentence in the lower part of the window.

The following corpus-depending parameters can be modified:

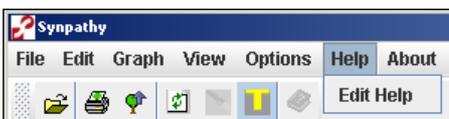
- Displayed non-terminal feature: For non-terminal nodes, only one feature can be displayed in the treebank graph. If there are two or more features defined for non-terminal nodes, the displayed feature can be changed in this dialog. By default, the non-terminal feature defined first in the TIGER-XML encoding is selected.
- Displayed terminal features: By default, all features of the terminal nodes are displayed. If you like to reduce the number of displayed features, just deselect the features you are not interested in. They will no longer appear in the treebank graph window.

To restore your configuration (that has been valid before opening the display options window), just click the Reset button. To restore the original system defaults that have been pre-installed by Synapsis click the Default button. To leave the window without changing the configuration press the Cancel button. To submit your changes press the OK button.

2.1.6 Help menu



The Help menu presents the user a small userguide on how to edit the treebank graph. see also chapter 3.



2.1.7 About menu



The about this tool option from the about menu shows the version of the Synpathy tool. It also give a [link](#) to other linguistic applications developed at the Max Planck Institute for Psycholinguistics.

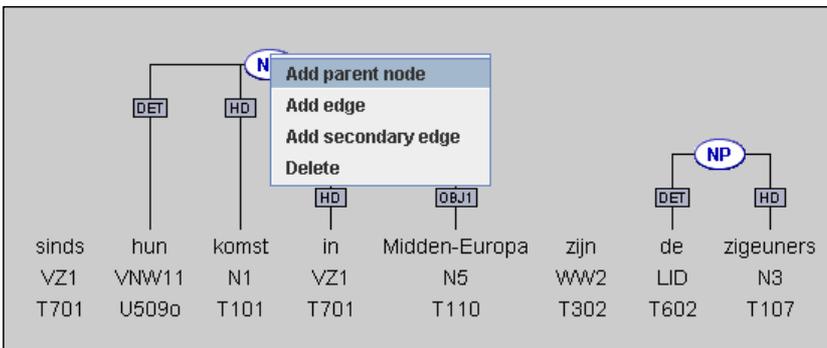


3 Editing

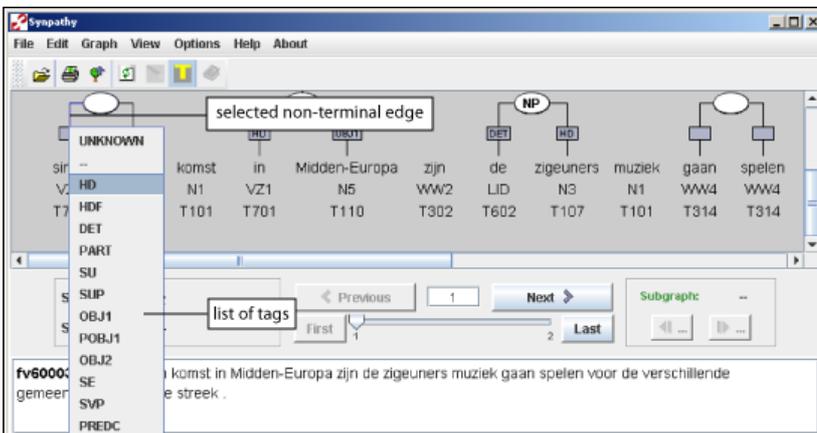


This chapter explains how to edit the treebank graph. It is assumed a (annotated) text is loaded into Synpathy.

First select at least one element from the syntax graph or sentence with the left mouse button. This part will now change color (the default color for selected elements is blue). You can mark further nodes by keeping the CTRL-key pressed while clicking other nodes. It is not possible to mark nodes beneath an already selected one. With the selection marked in blue press the right mouse button to activate a context sensitive pop-up menu from which you can select an editing operation.



For adding a new edge first select the "higher" node (where the edge comes from), than an arbitrary number of "lower" nodes (where the edge goes to). Secondary edges can point in the opposite direction though. In order to select labels, you first need to open an existing tagset (open feature from the [file](#) menu or to define this collection of tags on your own from scratch ([edit](#) menu)). To edit the label of a node, double-click on it's icon or edge in the syntax graph. A context sensitive menu will appear:



In case of non-terminal nodes (as shown in the picture above) choose the appropriate entry from the offered list. If the tag is not present on the current list, you can add it using the Edit -> Edit feature option. Do not forget to save the changes (File -> Save features). For terminal nodes you may modify the entries in the list but this will only affect the selected node. In order to change a value for a terminal node permanently you have to choose it again from the Edit menu.

4 Outlook



Synpathy will be further developed and supported by the Technical Department of the MPI. Plans for future extensions are extended editing of terminal nodes, scalable view, TeX export, synchronous visualization with [ELAN](#) and the inclusion of possible [TIGERSearch](#) in [ANNEX](#).

5 Concepts and definitions



- edge: labeled line in the treebank graph expressing relation between the connected parts.
- feature: TIGER-XML formatted file which contains label values used for the annotated sentence and for the tags in the treebank graph.
- Non-terminal edge: labeled element forming the treebank graph itself. It represents the syntactical structure of the sentence.
- Terminal node: unit in the sentence containing for instance the word, Part-of-Speech and other glosses.
- treebank: is a text corpus in which each sentence has been annotated with syntactic structure.
- TIGERsearch: specialized search engine for retrieving information from a treebank in TIGER-XML format. More info can be found [here](#).
- TIGER-XML: XML based format used to save information from Synpathy and TIGERsearch application. The extension is .tig.
- secondary edge: labeled line in the treebank graph which defines an additional parent node for a given node.
- SVG: An XML-based vector graphics format. This means that images can be scaled without loss of quality.