

# **EUDICO Linguistic Annotator (ELAN) version 2.6**

## **Manual**

This manual was last updated on 2006-04-04

The latest version can be downloaded from [www.mpi.nl/tools](http://www.mpi.nl/tools)

Original Author: Birgit Hellwig

Updates for version 2.x: Dieter Van Uytvanck

## **Introduction**

---

ELAN (EUDICO Linguistic Annotator) is an annotation tool that allows you to create, edit, visualize and search annotations for video and audio data. It was developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, with the aim to provide a sound technological basis for the annotation and exploitation of multi-media recordings. ELAN is specifically designed for the analysis of language, sign language, and gesture, but it can be used by everybody who works with media corpora, i.e., with video and/or audio data, for purposes of annotation, analysis and documentation.

ELAN supports:

- display a speech and/or video signals, together with their annotations;
- time linking of annotations to media streams;
- linking of annotations to other annotations;
- unlimited number of annotation tiers as defined by the users;
- different character sets;
- export as tab-delimited text files;
- im- and export between ELAN and Shoebox;
- search options.

This manual helps you to understand and use the features of ELAN.

Part I explains the installation process.

Part II is the user's guide. It is organized around the following four topics:

- ELAN documents
- The screen display and the navigation through a document.
- Annotations.
- Search Options.

For each topic, basic information is given. Following that, the use of features is explained in a step by step way. It is recommended that you read the relevant chapter before starting to work with your own data.

Part III is the reference guide, i.e., it provides brief information on the following topics: mouse options, menu items, and shortcut keys. In addition, the key concepts of ELAN are defined and new features (compared to earlier versions) are listed.

## Notation Conventions

---

The following notation conventions are used:

- Menu items, icons and screen displays are written in the font MS Sans Serif.
- (SHORTCUT) KEYS ARE WRITTEN IN SMALL CAPS.
- Step by step instructions on how to use a specific feature start with a caption on a gray background.

Information on troubleshooting is printed like this

Some of the screenshots used for illustration purposes throughout this manual are taken from corpora collected and annotated by Birgit Hellwig (on Goemai, a West Chadic language of Nigeria) and Gunter Senft (on Kilivila, an Austronesian language of Papua New Guinea).

# **Table of Contents**

<b>PART I: PRELIMINARIES.....</b>	<b>7</b>
<b>1     INSTALLING ELAN.....</b>	<b>7</b>
<b>2     SELECTING A USER INTERFACE LANGUAGE .....</b>	<b>7</b>
<b>3     ELAN DOCUMENTS .....</b>	<b>8</b>
3.1    Basic Information: Media Files and Annotation Files .....	8
3.2    How to manage documents.....	9
3.2.1    Creating a new document .....	11
3.2.2    Synchronizing video files .....	12
3.2.3    Opening an existing document .....	14
3.2.4    Saving a document .....	15
3.2.5    Saving a selection as .eaf file.....	15
3.2.6    Merging transcriptions.....	16
3.2.7    Saving a template .....	17
3.2.8    Changing the links to media files .....	17
3.2.9    Creating automatic backups.....	18
3.2.10    Printing .....	18
3.2.11    Configuring the page settings (page setup) .....	19
3.2.12    Previewing the printed pages.....	19
3.2.13    Importing Transcriber files.....	22
3.2.14    Importing a CHAT file .....	22
3.2.15    Importing a document from Shoebox .....	23
3.2.16    Exporting a document to Shoebox.....	29
3.2.17    Exporting a document as a tab-delimited text file .....	32
3.2.18    Exporting CHAT files .....	34
3.2.19    Exporting traditional transcript files.....	35
3.2.20    Exporting a part of a clip .....	36
3.2.21    Exporting a SMIL clip.....	36
3.2.22    Exporting ELAN's document view .....	36
3.2.23    Exporting to interlinear text.....	37
3.2.24    Opening a wave file in Praat.....	38
3.2.25    Exiting ELAN.....	39
3.3    Screen display and navigation through a document .....	39
3.3.1    Basic Information: The ELAN window.....	40
3.3.2    The Video Viewer .....	41
3.3.3    The Annotation Density Viewer.....	41
3.3.4    The Waveform Viewer .....	42
3.3.5    The Subtitle Viewer.....	43
3.3.6    The Grid Viewer.....	44
3.3.7    The Text Viewer.....	45
3.3.8    The Timeline Viewer and the Interlinear Viewer .....	46
3.3.9    The Timeseries Viewer.....	49
3.3.10    The Menu options.....	50
3.3.11    The Media Player options.....	51
3.3.12    The color coding.....	53
3.3.13    Increasing/decreasing the size of the ELAN window .....	54
3.3.14    Switching Viewers on/off.....	54
3.3.15    Increasing/decreasing the size of Viewers.....	55
3.3.16    Switching tiers on/off .....	56

3.3.17	Rearranging the order of tiers .....	56
3.3.18	Sorting tiers .....	57
3.3.19	Changing the time zoom.....	58
3.3.20	Changing the font size .....	58
<b>3.4</b>	<b>How to navigate through a document.....</b>	<b>59</b>
3.4.1	Accessing points in time.....	60
3.4.2	Going to the borders of a selection .....	61
3.4.3	Jumping step by step through a document.....	61
3.4.4	Viewing a list of annotations within one tier (Grid Viewer) .....	63
3.4.5	Jumping from annotation to annotation (Interlinear Viewer).....	65
<b>3.5</b>	<b>How to play a document .....</b>	<b>66</b>
3.5.1	Playing a document .....	66
3.5.2	Playing a selection.....	67
3.5.3	Playing around a selection.....	67
3.5.4	Playing in slow motion .....	68
3.5.5	Changing playback rate and volume via the keyboard .....	68
<b>4</b>	<b>ANNOTATIONS.....</b>	<b>69</b>
<b>4.1</b>	<b>Basic Information: Annotations, tiers and linguistic types.....</b>	<b>69</b>
<b>4.2</b>	<b>How to define a linguistic type .....</b>	<b>73</b>
4.2.1	Adding new linguistic types .....	73
4.2.2	Creating linguistic types .....	73
4.2.3	Changing linguistic types .....	74
4.2.4	Deleting linguistic types .....	74
<b>4.3</b>	<b>How to define a tier and its attributes .....</b>	<b>75</b>
4.3.1	Adding new tiers .....	75
4.3.2	Importing tiers .....	76
4.3.3	Changing tier attributes .....	77
4.3.4	Deleting tiers .....	78
4.3.5	Changing the parent of a tier .....	78
<b>4.4</b>	<b>Associating Controlled Vocabularies with linguistic types.....</b>	<b>82</b>
<b>4.5</b>	<b>How to make a selection.....</b>	<b>82</b>
4.5.1	Making a selection on an independent tier .....	82
4.5.2	Using the selection controls.....	83
4.5.3	Creating annotation units without gaps .....	85
4.5.4	Making and saving a selection on a referring tier.....	87
4.5.5	Deselecting a selection .....	87
4.5.6	Changing the boundaries of an existing selection .....	88
4.5.7	Activating and deactivating the Bulldozer mode or Shift mode .....	89
<b>4.6</b>	<b>How to enter annotations.....</b>	<b>91</b>
4.6.1	Entering annotations (into the Inline Edit box).....	92
4.6.2	Entering annotations (into the Edit Annotation box) .....	94
4.6.3	Entering annotations before/after other annotations.....	96
4.6.4	On-the-fly segmentation.....	97
4.6.5	Modifying the content of annotations.....	98
4.6.6	Deleting annotations.....	98
4.6.7	Entering annotations in different character sets.....	99
<b>4.7</b>	<b>How to create annotation units on dependent tiers .....</b>	<b>103</b>
<b>4.8</b>	<b>How to filter and copy tiers .....</b>	<b>104</b>

4.8.1	Filtering tiers .....	104
4.8.2	Copying tiers .....	105
<b>4.9</b>	<b>How to undo an action .....</b>	<b>106</b>
<b>5</b>	<b>SEARCHING.....</b>	<b>107</b>
<b>5.1</b>	<b>Searching in a single annotation file .....</b>	<b>108</b>
5.1.1	Advanced searching: an example .....	113
<b>5.2</b>	<b>Displaying search results .....</b>	<b>114</b>
5.2.1	The search history .....	114
5.2.2	Refining a search .....	115
5.2.3	Saving and loading queries.....	115
5.2.4	Exporting the search results.....	116
5.2.5	Search and replace.....	117
<b>5.3</b>	<b>Jumping from the search results to annotations.....</b>	<b>118</b>
<b>5.4</b>	<b>Searching through multiple annotation files.....</b>	<b>118</b>
	<b><u>PART III: REFERENCE GUIDE .....</u></b>	<b><u>121</u></b>
<b>5.5</b>	<b>The Mouse Options .....</b>	<b>121</b>
5.5.1	Resizing Viewers.....	121
5.5.2	Rearranging the order of tiers.....	121
5.5.3	Make a tier the active tier.....	121
5.5.4	Accessing points in time.....	121
5.5.5	Playing a document .....	121
5.5.6	Making a selection.....	121
5.5.7	Adding/modifying an annotation.....	121
<b>5.6</b>	<b>The Menu Items.....</b>	<b>122</b>
5.6.1	File menu.....	122
5.6.2	Edit menu .....	123
5.6.3	Search menu .....	123
5.6.4	View menu .....	124
5.6.5	Options menu .....	124
5.6.6	Help menu .....	124
<b>5.7</b>	<b>The Shortcut Keys.....</b>	<b>124</b>
5.7.1	File options.....	124
5.7.2	Playing a document & making selections.....	125
5.7.3	Working with annotations .....	125
5.7.4	Working with selections .....	126
5.7.5	Searching.....	126
5.7.6	General .....	126
<b>5.8</b>	<b>Key concepts of ELAN.....</b>	<b>126</b>
5.8.1	Annotation file (*.eaf) .....	126
5.8.2	Media file (*.mpg/*.mov, *.wav).....	126
5.8.3	Annotation .....	126
5.8.4	Tier .....	126
	<b><u>APPENDIX: REGULAR EXPRESSION SEARCH.....</u></b>	<b><u>128</u></b>
	Character Classes .....	131
	Line terminators .....	131

Groups and capturing .....	131
Comparison to Perl 5 .....	132

## **PART I: PRELIMINARIES**

---

### **1 Installing ELAN**

To install ELAN on your computer, go to <http://www.mpi.nl/tools> and follow the instructions. On this webpage, you will also find information about software and hardware requirements.

### **2 Selecting a user interface language**

Once ELAN has been started, you can change the language of the user interface at any time. To achieve this, choose **Options > Language** and select one of the available languages.

The selected Language does not influence the content of the produced or edited \*.eaf files in any way.

At present Catalan, Dutch, English, French ,German, Portuguese, Spanish and Swedish language modules are available. However, new languages can be easily added. If you want to provide a translation for a different language, please contact the ELAN development team.

## PART II: USER'S GUIDE

This part of the manual contains the user's guide. It is organized as follows:

- ELAN documents (see section 3)
- Screen display and navigation through a document (see section 3.3)
- Annotations (see section 4)
- Search Options (see section 4.7)

It is recommended that you read the relevant section before starting to work with your own data.

### **3 ELAN documents**

#### **3.1 Basic Information: Media Files and Annotation Files**

Every ELAN project consists of at least two files: one (or more) media file(s), and one annotation file.

**(1) One (or more) media file(s):**

- 0 up to 4 video file(s) (e.g. \*.mpg, \*.mov)
- and/or one audio file (\*.wav)

The video file allows you to view the video and listen to the sound. If you want to view the waveform as well, you need to create an additional \*.wav file through a conversion program that converts the audio data from one \*.mp(e)g file into a \*.wav format. In the case of multiple video files, the audio of the first selected video file is played. The kind and number of supported video formats depend upon the media framework you are using. If your media framework supports a format, the same goes for ELAN.

The following media frameworks are known to work:

- Windows (in order of preference, DirectX being the best solution):
  - DirectX/DirectShow (Windows Media Player)
  - QuickTime
  - JMF (Java Media Framework)
- MacOS: QuickTime
- Linux: JMF

For \*.mov files (i.e., Cinepak-Quicktime-Movies) it is important that these are self-contained files, i.e., the video information needs to be contained within the \*.mov file itself. If this is not the case, ELAN will not be able to display the file.

**(2) One annotation file:**

- an annotation file created by ELAN (\*.eaf, "EUDICO Annotation Format")



- or an imported annotation file. The following formats can be imported:
  - Shoebox/Toolbox files (with the extension **\*.txt**)
  - CHAT files (**\*.cha**)
  - Transcriber files (**\*.trs**)

All information (e.g., the tier setup, the time alignment, the annotations) is saved to the annotation file only – never to the media file(s).

Take care when editing a media file. Afterwards you probably will want to resynchronize its alignment with the corresponding the annotations, as described in section 3.2.2.

Although it's not compulsory it is a good practice to use a common name for media files and the annotation file. So, it is recommended to use *a.eaf* next to *a.mpg* and *a.wav*.

Imported Shoebox files (**\*.txt**) also do not need to have the same name as their media files, and they can be located in different directories. All imported files can alternatively be saved as ELAN files (**\*.eaf**).

All annotation files (**\*.eaf**) can be exported as text, Shoebox/Toolbox and CHAT files.

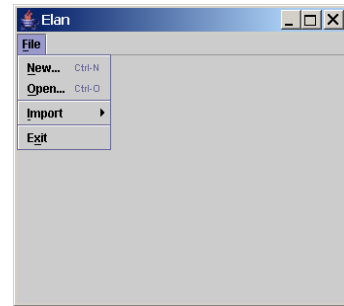
## 3.2 How to manage documents

To start ELAN, do the following:

1. Double-click on the ELAN icon (on your desktop, or contained within the Java Web Start icon on your desktop).  
The start dialog window is displayed, asking you what kind of document you want to open.
2. Click on the **File** menu

The File menu supports the following options:

- New (section 3.2.1)
- Open (section 3.2.3)
- Import (section 3.2.13, 3.2.14 and 3.2.15)
- Exit (section 3.2.23)



3. Click on:

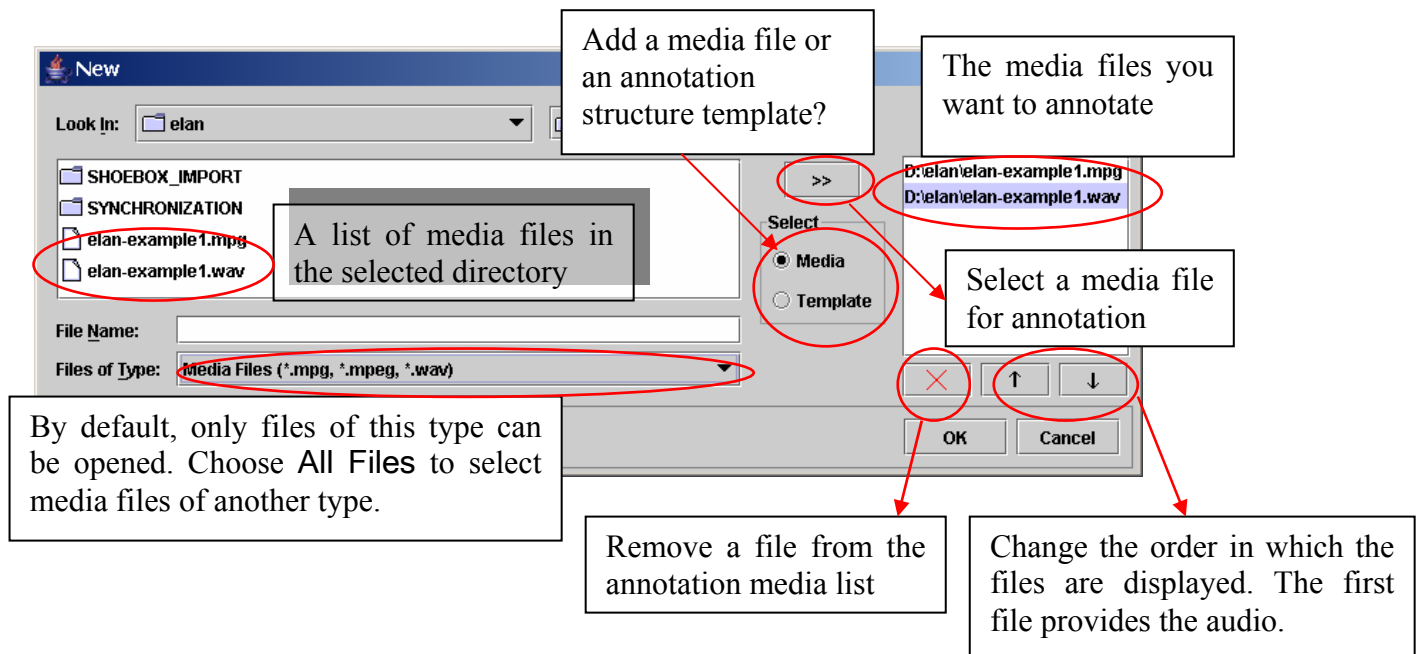
- Open... in case there is an ELAN file (\*.eaf) (see section 3.2.3)
- New... in case there is a media file (e.g. \*.mpg, \*.wav), but no annotation file (\*.eaf, \*.txt) (see section 3.2.1).
- Import > Shoebox file... (see section 3.2.15)
- Import > CHAT file... (see section 3.2.14)
- Import > Transcriber file... (see section 3.2.13)

Different dialog windows appear and prompt you to enter the names and locations of the different files. Then the ELAN window appears and displays the selected files.

Once you have started ELAN, use the File menu to open, create or import a second document.

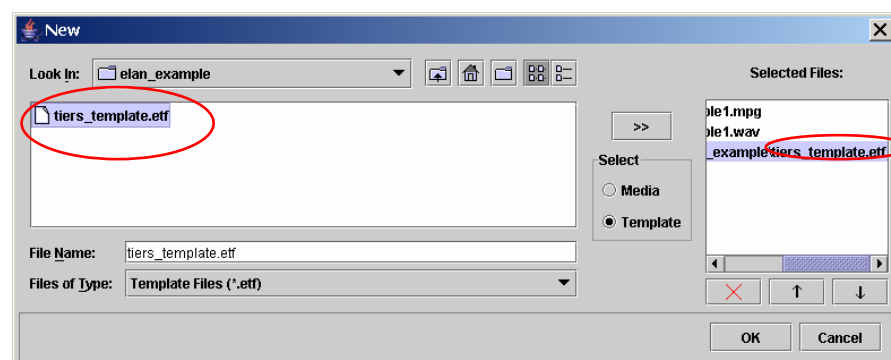
### 3.2.1 Creating a new document

In case there is a media file but no annotation file (\*.eaf, \*.txt, \*.trs), click on New in the File menu. The New Transcription dialog window will be displayed, e.g.:



Do the following:

1. Click on the **Look in** pull down box and browse to the directory that contains the media files
2. If you want to use media files of another type (e.g. QuickTime \*.mov) then select **All Files** in the **Files of type** dropdown menu. If a media type is supported depends on your software configuration.
3. Double-click on the media file (\*.mpg/\*.mov, \*.wav, etc.) to select it. It appears now in the rightmost box. Alternatively, you can click on the media file name and click afterwards on the >> button.
4. If you want to use a predefined set of tiers (a template), select the **Template** radio button and choose the template (i.e. \*.etf) to be used:



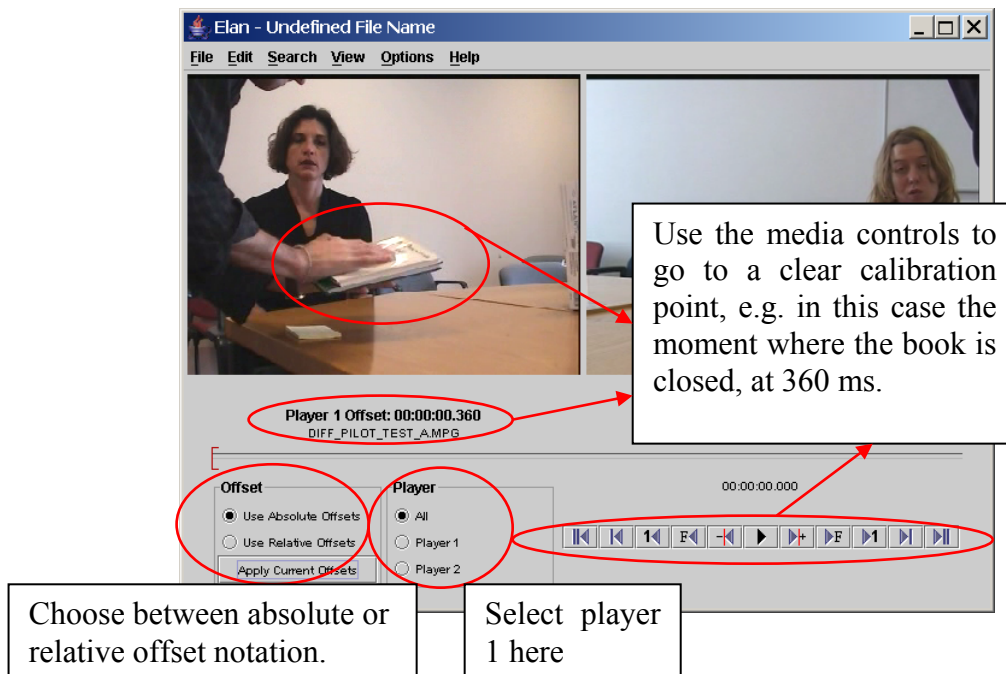
5. Click OK to open the new annotation document; otherwise click Cancel to exit the dialog window without creating a new file.

An ELAN window containing the new document appears.

### 3.2.2 Synchronizing video files

Sometimes one has two or more (up to 4) video files of the same recorded scene, e.g. when 2 different cameras were used. In that case it may happen that both recordings don't start exactly at the same moment. In order to fix this, one should synchronize the videos. This can be done as follows:

1. Open the new document with the 2 (or more) video files.
2. Select the pull down menu **Options > Media Synchronization Mode**
3. Make a choice about how the time codes should be displayed:
  - a. Absolute offsets: for every video its own timing is being shown.
  - b. Relative offsets: the video of player 1 is appointed to be the "master", i.e. the time position of the other videos will be expressed as to the starting point of this file, which starts at 00:00:00.000.
4. Select the radio button *player 1*. You can now choose a moment in the video which is easy to calibrate (some clear anchor point, in both of the videos). For instructions how to navigate through the video file, see section 3.4.



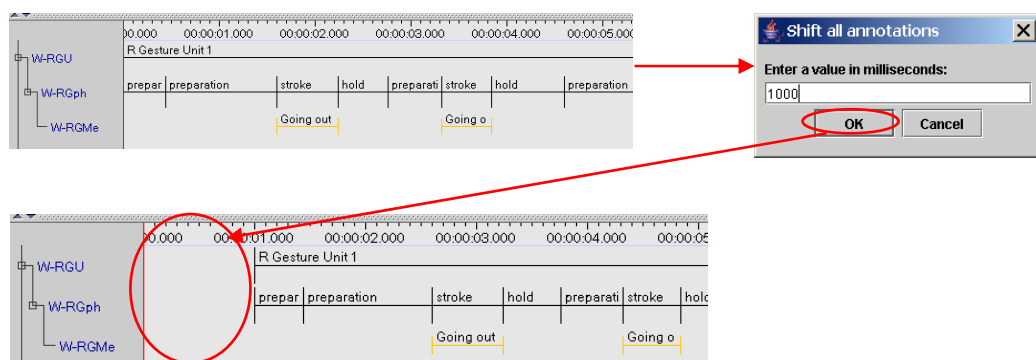
5. Now select *player 2* and go to the same calibration point.

See section 3.2.8 for changing the order of the videos, i.e. the order of appearing in Player 1, Player 2, etc.



6. Finally, choose **Apply current offset**. By selecting the play button both videos will be played together now, so you can check if the synchronization between them is correct. If not, please repeat step 3-5 until the result is satisfactory.
7. Leave the synchronization mode by selecting **Options > Annotation Mode**. Now you are ready to start entering annotations.
8. By double clicking on a video, it will be placed in the leftmost video window (which is also the biggest one in case there are 3 videos).

If you changed the media file synchronization of a file that already is annotated, you might want to move the annotation units all together to the right (later, positive value) or to the left (earlier, negative value) on the time axis. This can be done using the **Edit > Shift all annotations ...** menu:

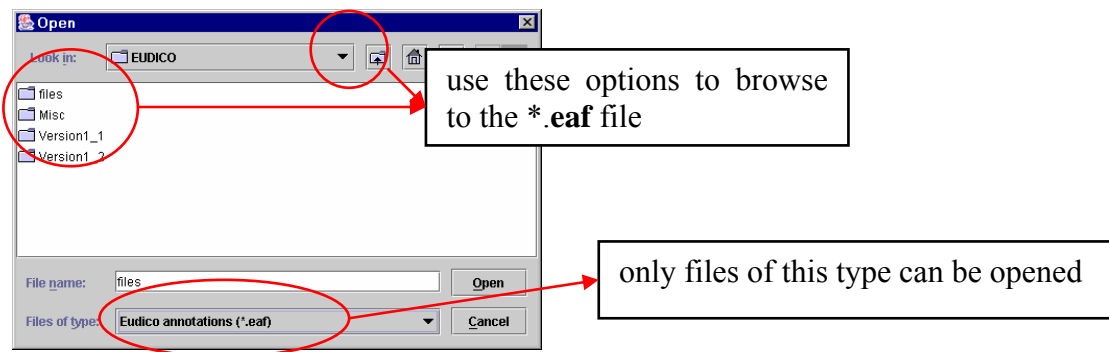


This process won't delete any annotation. If the annotations are shifted to the left, the maximum shift will be restricted by the leftmost annotation unit.

### 3.2.3 Opening an existing document

In case there is a media file (\*.mpg/\*.mov, \*.wav) and an ELAN file (\*.eaf), click Open in the File menu.

The Open dialog window is displayed:

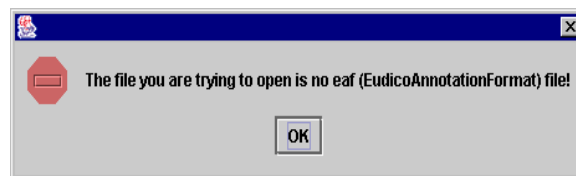


Do the following:

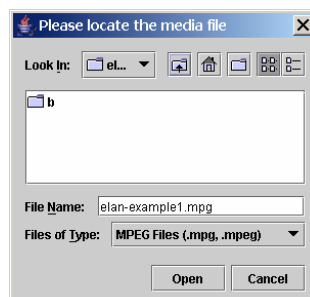
1. Browse to the directory that contains the ELAN file (\*.eaf).
2. Double-click on the annotation file to open it.

An ELAN window containing the document appears.

You can only open files of the EUDICO annotation format (\*.eaf). If you try to open a file of a different format, the following error message appears:



If ELAN cannot find the associated media files (\*.mpg, \*.mpeg, \*.mov, \*.wav), it will check if these files exist in the directory of the EAF-file. If they are still not found there, it will ask you where the media files are located:



### 3.2.4 Saving a document

All documents can be saved as ELAN files (\*.eaf, “EUDICO Annotation Format”). This includes documents that were created by ELAN itself (see section 3.2.1) as well as documents that were imported into ELAN from Shoebox/Toolbox, CHAT or Transcriber (see section 3.2.15, 3.2.14 or 3.2.13). To save a document as an ELAN file:

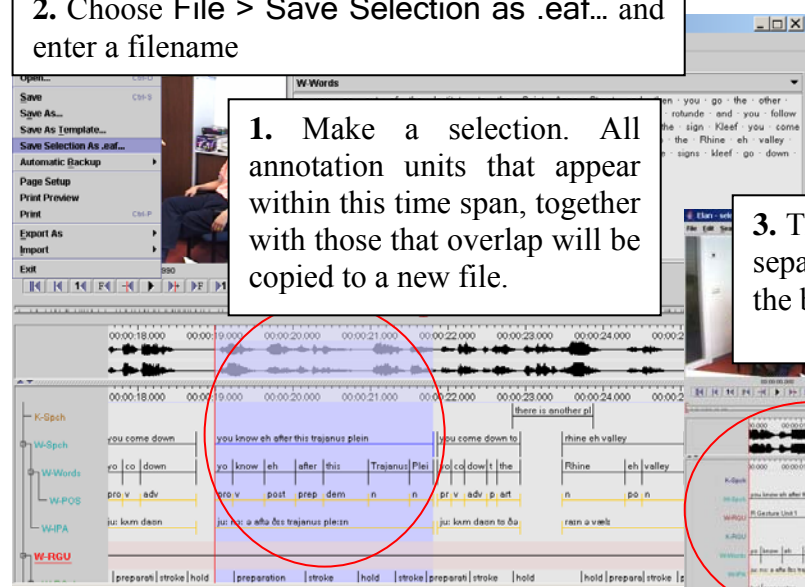
- (a) Either use the menu options:
  1. Click on **File** menu.
  2. Click on **Save** or **Save as**.
- (b) Or use the shortcut key CTRL+S.

Apart from the \*.eaf file, a \*.pfs file will be written as well. This file contains user and document specific settings like the font size used to display text. The \*.pfs file can however be safely removed as it does not contain any annotation data.

### 3.2.5 Saving a selection as .eaf file

Apart from saving a whole document you can also store the contents of a certain time span to a .eaf file, using the following steps:

**2. Choose File > Save Selection as .eaf... and enter a filename**



**1. Make a selection. All annotation units that appear within this time span, together with those that overlap will be copied to a new file.**

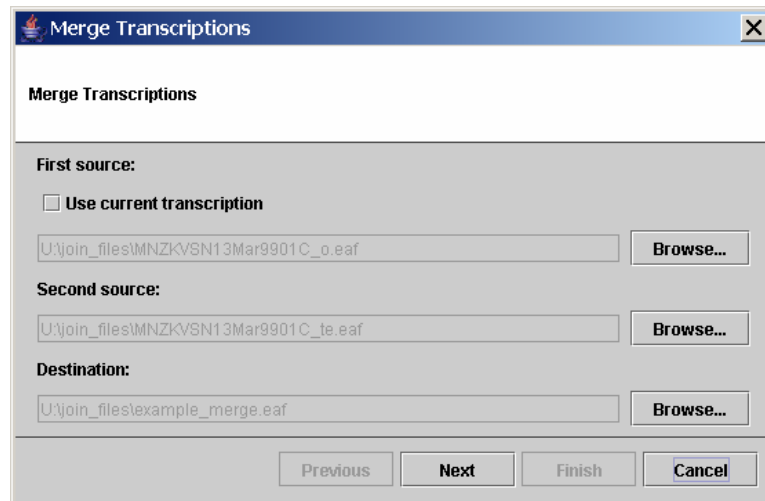
**3. The selection is now saved as a separate file. It will be placed at the begin of the timeline.**

If annotation units overlap with the selection, they will be shrunk until they fit within the selected interval.

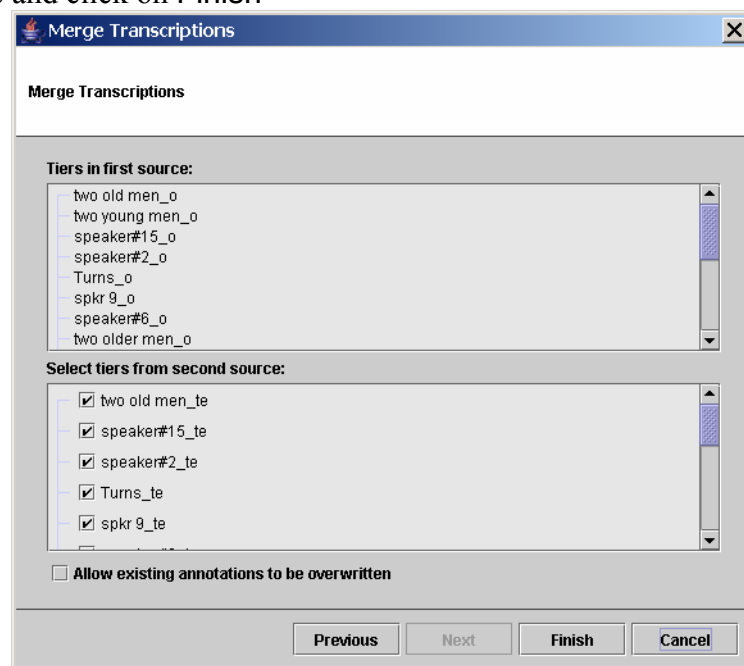
### 3.2.6 Merging transcriptions

Under some circumstances it might be useful to combine the contents of two separate transcription files into a single one. To achieve this, follow these steps:

1. Choose **File > Merge Transcriptions**
2. A dialog window appears:



3. If one of the files to be merged is currently opened, select **Use current transcription**. Otherwise choose **Browse...** and select the first eaf-file
4. Choose the second file. If there are common tiers in both files and you want those of the second file to overwrite those of the first, make sure **Allow existing annotations to be overwritten** is checked.
5. Enter a file name for the result of the merge operation.
6. Click on **Next**
7. Select the tiers of the second source file that you want to merge with the first file and click on **Finish**





8. When the merge procedure has been finished you can choose whether to open the result immediately in new ELAN window:



### 3.2.7 Saving a template

A template offers the possibility to reuse the same document setup for more than one media file. This includes:

- linguistic types (see section 4.2)
- controlled vocabulary (see section 4.4)
- tiers (see section 4.3)

Saving a template is done as follows:

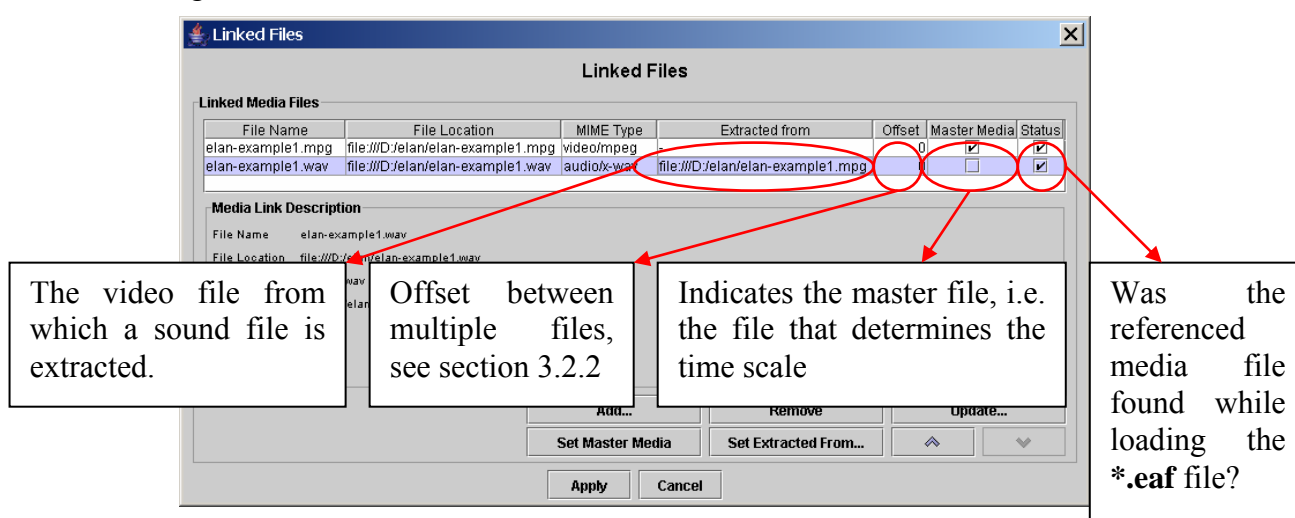
1. Click on **File** menu.
2. Click on **Save as Template...**
3. Choose a file name ending in **.etf**
4. Click on **Save**

See section 3.2.1 for a instructions on using a template.

### 3.2.8 Changing the links to media files


As from ELAN version 2.4, the possibility exists to explicitly change the links to media files that are linked from an **\*.eaf** file. This option can be handy if e.g. you have moved media files to another location after the last time you edited an ELAN file.

Activating the Linked Files dialog (via **Edit > Linked files...**) will get you the following screen:



These options are available:

- **Add...** : add a link to a new media file to the current **\*.eaf** file

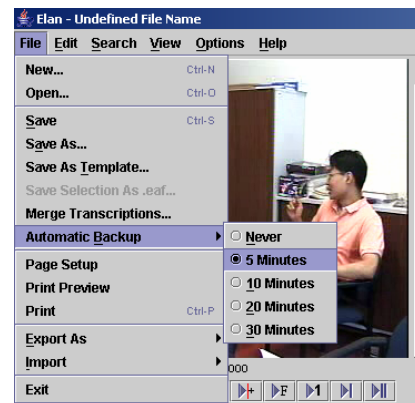
- **Remove:** remove the selected media file
- **Update...** : specify a new location of the selected file. Especially useful if the checkbox *Status* is not marked. The latter indicates the media file could not be found while the ELAN file was opened (e.g. because the media files was moved).
- **Set Master Media:** make the selected media file the Master Media
- **Set Extracted from...** : indicate that a sound file has been extracted from a video file
- : moves a file up/down in the linked file list. The file on top automatically becomes the Master Media file. The audio file on the highest location is displayed in the Waveform Viewer.

### 3.2.9 Creating automatic backups

ELAN allows you to create automatic backup copies. To create backups, do the following:

1. Click on File menu.
2. Go to Automatic backup.
3. Click on the time interval after which ELAN should create the backup, e.g., after every 10 minutes.

A checkmark appears next to the selected time interval. From now on ELAN will automatically create a backup copy, saving it with the extension **\*.eaf.001**. Before opening such a file, rename its extension to *.eaf* instead of *.eaf.001*

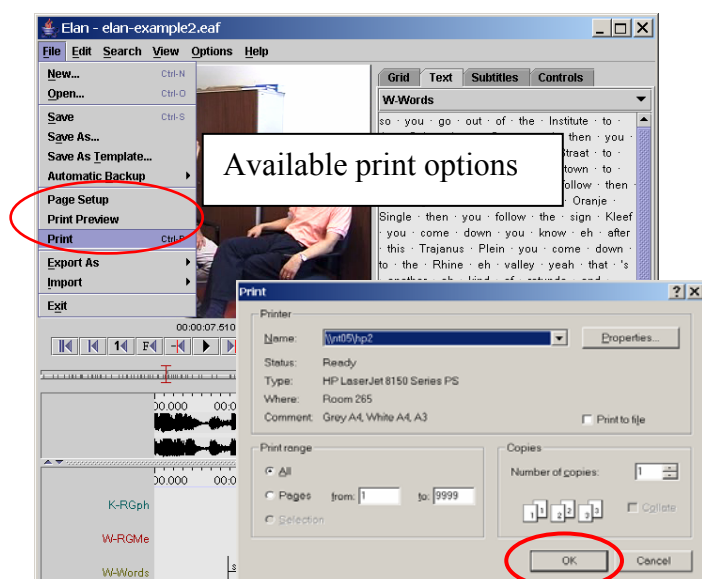


**Automatic backups can only be made *after* a file has been saved!**

### 3.2.10 Printing

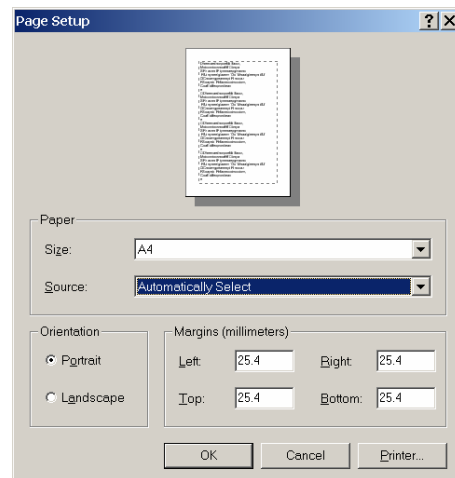
1. Printing from within ELAN can be achieved by selecting the File > Print menu.

2. Then a standard print dialog is shown, choose OK to start printing.



### 3.2.11 Configuring the page settings (page setup)

Through **File > Page setup** you can alter the paper size and other settings of the pages to be printed.





### 3.2.12 Previewing the printed pages

The fine tuning of the print result can be done by opening the Print Preview window, which is accessible via the **File > Print Preview** menu.

**Annotations:**

- the actual print preview of the tiers and their labels** (points to the left preview pane)
- tiers to print, their font size and order** (points to the Tiers list in the right pane)
- extra print options** (points to the 'What' section in the right pane)
- layout settings** (points to the 'How' section in the right pane)
- After changing a print setting, click here to update the preview.** (points to the 'Apply Changes' button)
- Open the Page Setup dialog.** (points to the 'Page Setup' button)
- Go to the print window.** (points to the 'Print' button)

## Tiers settings

- Put a checkmark in front of all the tiers that should be printed.
- Arrange the order of the tiers with the   buttons.
- The font size of the tiers can be adapted by clicking on the **Font Sizes** button. A new window will appear:

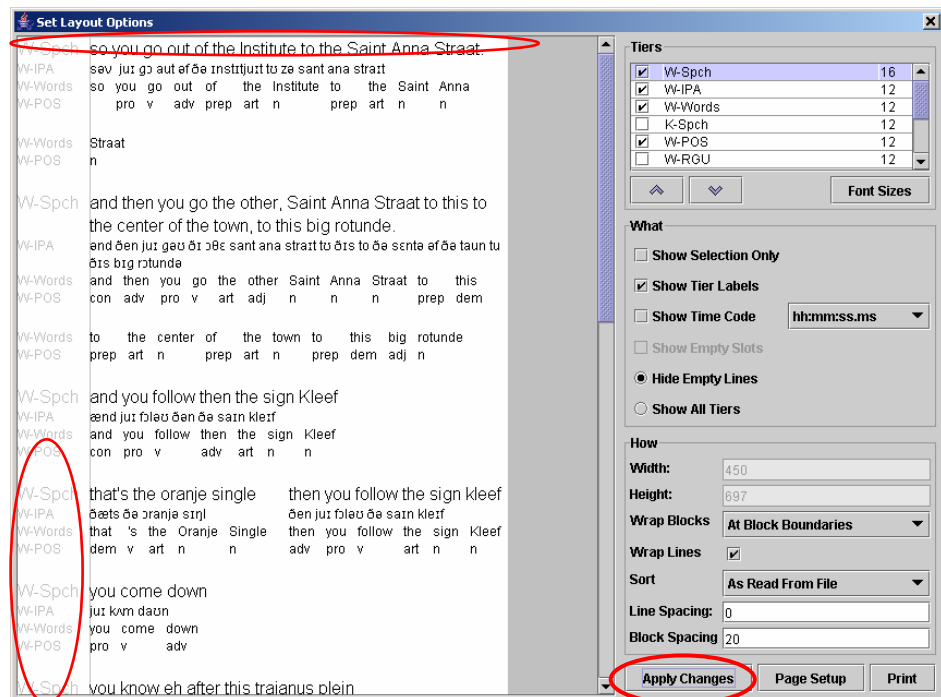
Change this tier's font size



New font size

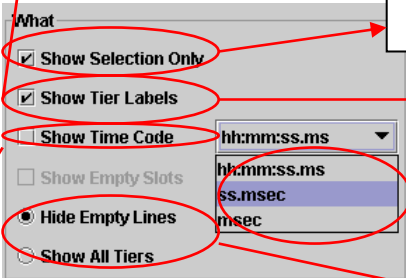
Set the font size of *all* tiers at once

After choosing the desired font size, click on the **Apply Changes** button in the Print Preview window. After that, the changes will appear:



The image shows the 'Set Layout Options' dialog box. The 'Tiers' section on the right lists various tiers with checkboxes and font sizes. The 'What' section has checkboxes for 'Show Selection Only', 'Show Tier Labels', 'Show Time Code', 'Show Empty Slots', 'Hide Empty Lines', and 'Show All Tiers'. The 'How' section has fields for 'Width', 'Height', 'Wrap Blocks', 'Wrap Lines', 'Sort', 'Line Spacing', and 'Block Spacing'. The 'Apply Changes' button is circled in red.

## “What” settings



Only display the selected part of the annotation

toggles the visibility of the tier labels

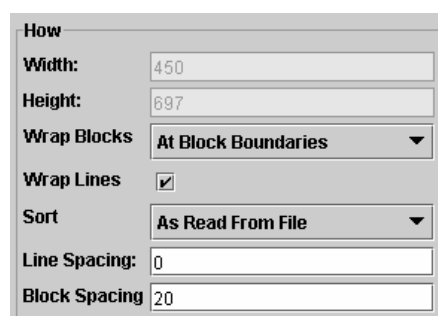
select the time format for the TC lines

Choose between displaying all lines or hiding the empty ones.

displays the duration of the annotation units (or its parent if it is a dependent tier) on the “TC” (Time Code) line

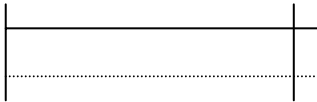
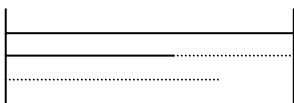
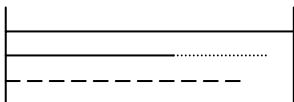
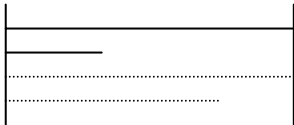
The option *Show empty slots* is not yet implemented.

## “How” settings



The 'How' settings dialog box includes the following controls:

- Width:** 450
- Height:** 697
- Wrap Blocks:** At Block Boundaries
- Wrap Lines:** ☒
- Sort:** As Read From File
- Line Spacing:** 0
- Block Spacing:** 20

- **Width:** specify the width of the printed area (in pixels). This value can only be changed by selecting a paper format in the **Page setup** dialog (see section 3.2.11).
- **Height:** enter the height of the printed area (in pixels). If you leave this empty, the default height will depend upon the selected paper size.
- **Wrap Blocks:**
  - *No wrapping:* use 1 line for each tier, only usable for files that contain a small amount of annotations. 
  - *Within block:* wrap blocks, and continue with a new block on the same line if there is space left. 
  - *At block boundaries:* wrap blocks, and continue with a new block on the same line if there is space left and if the new block fits on that line. 
  - *Each block:* wrap blocks, and start on a new line if a block ends. 
- **Sort:** specify in which order the blocks will appear. This is similar to the tier sorting function (see section 3.3.18).
- **Line spacing:** amount of whitespace between the lines (default: 0 pixels)
- **Block spacing:** amount of whitespace between the blocks (default: 20 pixels)

### 3.2.13 Importing Transcriber files

The feature to import Transcriber annotation files into ELAN works as follows:

1. Choose **File > Import > Transcriber File ...**
2. Select the transcriber file (\*.trs) and click on **Open**
3. If the associated sound file cannot be found, a dialog will be shown asking you to locate it. When this request is cancelled, one can choose to open the annotation file without the sound, or to stop the whole import process.

The transcriber tiers will be mapped on the ELAN equivalents:

- *Section* becomes a root tier and *turns* becomes a dependent time subdivision of *section*.
- Events are embedded into the annotation text.

### 3.2.14 Importing a CHAT file

It is possible to import CHAT files (used in e.g. the Childe project) in ELAN:

1. Select **File > Import > CHAT File ...**
2. Select the Chat file
3. Click on **Open**

Some remarks about this import feature:

- supported are old CHAT files and CHAT-UTF8, not XML CHAT
- existing media alignment in %snd tiers is maintained in ELAN:
  - when no media alignment is present at all, each CHAT utterance gets a default interval of 1 second assigned
  - when partial media alignment is present, the time interval is equally distributed over preceding unaligned utterances
  - overlapping utterances of the same participant are corrected as good as possible

1 CHAT dependent tier names are mapped to ELAN Linguistic Types

2 ELAN tier names are either CHAT participant labels or CHAT tier names,

3 followed by '@participantName'

Remaining issues:

- '<' and '>' characters in CHAT cause parsing errors when the imported file is saved as EAF file

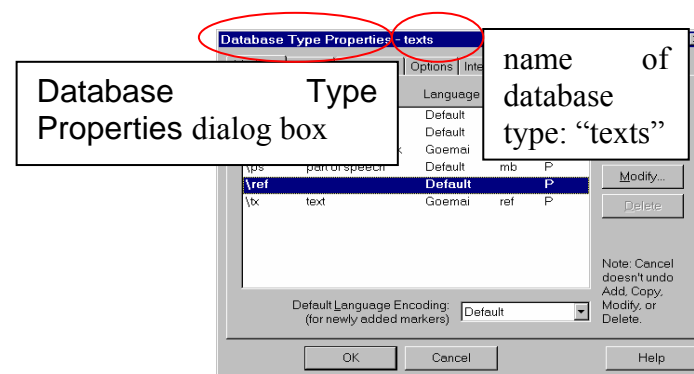
### 3.2.15 Importing a document from Shoebox<sup>1</sup>

ELAN supports the import of documents from Shoebox, thereby allowing you to link transcribed and/or interlinearized documents to the time axis of media files. In order to import from Shoebox, you need at least the following two files:

- the Shoebox file (\*.txt);
- the media file(s) (\*.mpg/\*.mov, \*.wav);
- Optionally the corresponding Shoebox database type file (\*.typ). If this is not available, one has to provide a list with field markers (= tier names).

If you do not know the Shoebox database type file, do the following:

1. Open the Shoebox \*.txt file in Shoebox. Make sure it is the active window (click on it to activate it).
2. Click on Database menu.
3. Click on Properties .... The Database Type Properties dialog box appears. The name of the database type is displayed in the header, e.g.:



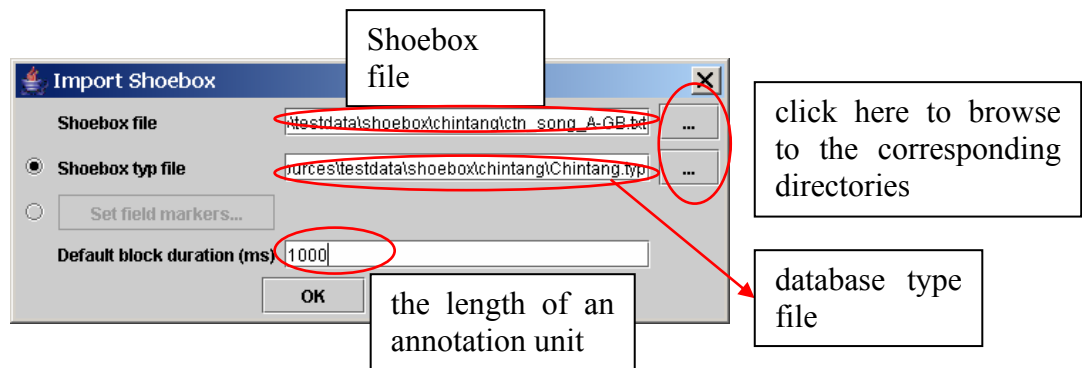
4. Locate the directory of the database type file (e.g., “texts.typ” in the above illustration). It is probably located in the directory “My Shoebox Settings”.

<sup>1</sup> From here on, every appearance of Shoebox can also be read as Toolbox, i.e. the newer version of what was formerly known as Shoebox.

## Importing shoebox files with a TYP file

To import a Shoebox file into ELAN, do the following:

1. Click on File > Import > Shoebox File. The Import Shoebox dialog box appears.
2. Specify the name and directory of the two files, e.g.:



Like \*.eaf documents, the Shoebox file and the media file(s) do not necessarily need to have the same name, and they do not need to be in the same directory (see section 3.1).

If the shoebox file contains both aligned (i.e. containing time information) and non-aligned records, the aligned ones will maintain the timing, whereas the location of the non-aligned records will be interpolated automatically.

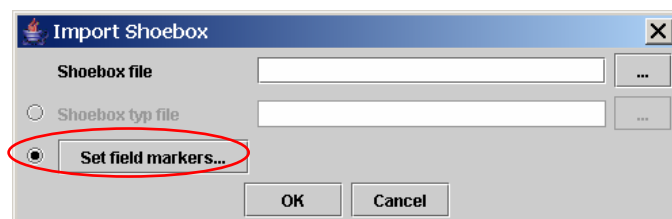
3. Click OK to import the file; otherwise click Cancel to exit the dialog box without importing the file.

An ELAN window containing the imported Shoebox file appears.

## Importing shoebox files without a TYP file

Instead of using a shoebox \*.typ file, there also is an option in ELAN to define the field markers yourself when importing a shoebox file.

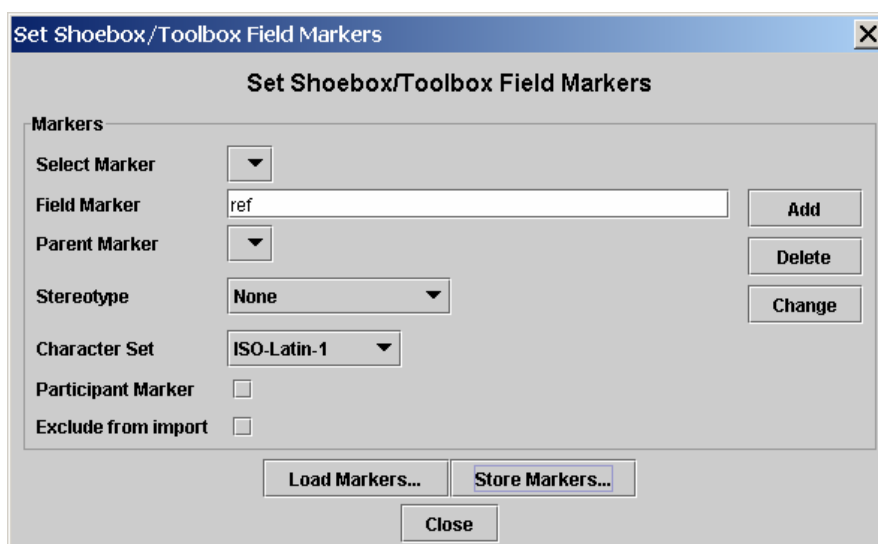
1. select the Set field markers and click on the button in the import dialog:



2. Now fill in a field marker as used in the shoebox \*.txt file
3. Optionally select a parent marker,
4. Optionally select a stereotype (symbolic subdivision or association)



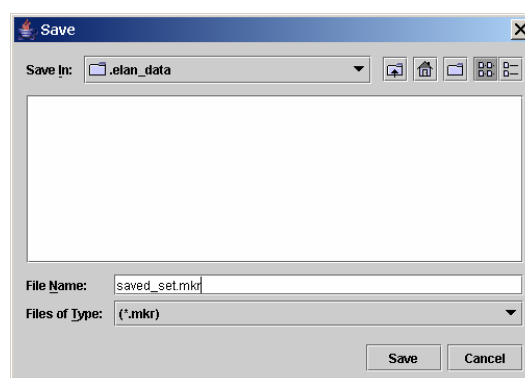
5. Choose a character set (Latin-1, SIL IPA or UTF-8) for the tier
6. Click on Add.
7. Repeat step 2-6 for all field markers.
8. If the selected marker designates a participant, check the **Participant Marker** checkbox. If you don't want the selected marker to be imported, tick **Exclude from import**.
9. finally choose Close and click on OK in the import shoebox file dialog



## Loading and storing Markers

Once you have manually created a set of field makers, you might want to reuse them later on. ELAN provides support for this:

- To save a set of field markers, select the **Store Markers...** button. This will display a save dialog. Enter a filename, and press save.
- The same way you can open a stored field marker set by clicking on **Load Markers...**



## Connecting the transcription to a media file

Once the import has succeeded, you can add a reference to a media file via the **Edit > Linked Files...** menu, as described in section 3.2.8. If the imported shoebox file was exported from ELAN before, you won't need to establish the link to the media file(s) again, as in that case the location information is stored in the file.

## About the import process

ELAN imports Shoebox files according to the following conventions:

- (1) The Shoebox field markers are imported as ELAN tiers. The tier label is identical to that of the field marker, except for the added extension '@Speaker-ID'.

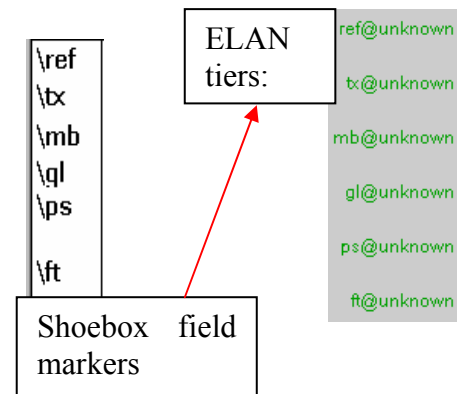
This addition is necessary because ELAN and Shoebox differ in how they code information about multiple speakers:

- In ELAN, each speaker is coded on a separate tier.
- In Shoebox, all speakers are coded using the same field, and their identity is specified in a separate field.

When importing texts by multiple speakers, ELAN splits each Shoebox field into several ELAN tiers (one for each speaker) and adds the speaker-ID to the tier label.

If speaker information is not specified in the Shoebox file, the extension @unknown is added.

The following screenshot illustrates how ELAN treats texts by multiple speakers:



The screenshot shows two windows of the ELAN software. The top window displays a tier for speaker 'Fix' with annotations for 'Born and raised in Indonesia'. The bottom window displays a tier for speaker 'Foxi' with annotations for 'BORN AND RAISED IN INDONESIA TOO'. Both windows show the 'Toolbox - [test]' interface with various fields like \name, \ref, \ELANBegin, \ELANEnd, \ELANParticipant, \tx, \mb, \gl, \ps, and \ft. Red circles and arrows highlight the speaker-specific annotations and the corresponding tier labels.

Annotations for speaker Fix:

- ft@Fix
- gl@Fix
- mb@Fix
- ps@Fix
- ref@Fix
- tx@Fix

Annotations for speaker Foxi:

- ft@Foxi
- gl@Foxi
- mb@Foxi

Note that ELAN can only read speaker information if:

- A marker is defined as a **Participant** marker in the Set field marker dialog (see *Importing shoebox files without a TYP file* above), or if:

- Or if it is coded in a Shoebox field labeled `\EUDICOp` or `\ELANParticipant` (see illustration above). If this field is not present, or if speaker information is coded in a different field, ELAN will assume that there is only one speaker. I.e., if you have multiple speakers and if you want ELAN to assign them to separate tiers, do the following:
  1. For every Shoebox record, add the field marker `\EUDICOp`.
  2. For every Shoebox record, enter the relevant speaker-ID into this field.

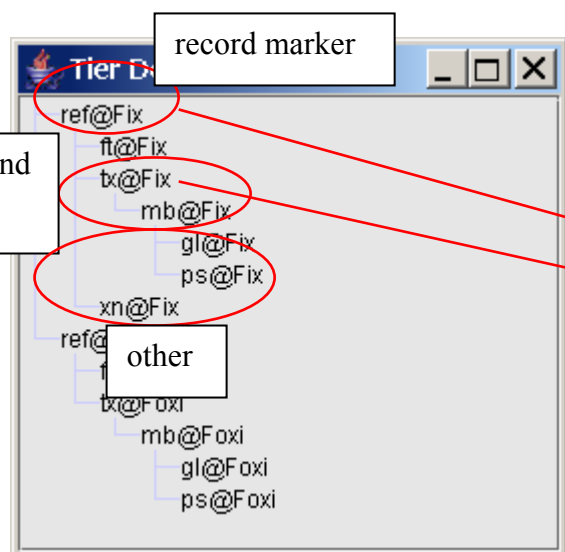
When the file is exported back to Shoebox (see section 3.2.16), the extension `@'Speaker-ID'` is automatically dropped from the field marker, and the Shoebox records are sorted according to their record marker (e.g., in the above illustration, “test 001” is sorted before “test 002” etc.)

(2) Based on the information contained in the Shoebox database type file, the tiers are brought into a hierarchical relationship and are assigned to linguistic types (see section 4.1 for details of tier hierarchies and linguistic types). For every tier name a corresponding linguistic type with the same name is created. All of these linguistic types are connected with a stereotype in such a way that it fits with the original shoebox structure.

- The Shoebox record marker is assigned to the stereotype **None**, i.e., it is an independent, time-alignable parent tier.
- The transcription and parsing fields of Shoebox are assigned to the stereotype **Symbolic Subdivision**, i.e., they are referring tiers that can be subdivided into smaller units.
- All other fields are assigned to the stereotype **Symbolic Association**, i.e., they are referring tiers that cannot be subdivided into smaller units.

If you define the markers yourself, then there also is the possibility to choose the **Time Subdivision** stereotype.

For example:



Change Tier Attributes

Tier Name	Parent Tier	Linguistic Type	Participant	Default Language
ft@Fix	ref@Fix	ft	Fix	us (EN)
ft@Foxi	ref@Foxi	ft	Foxi	us (EN)
gl@Fix	mb@Fix	gl	Fix	us (EN)
gl@Foxi	mb@Foxi	gl	Foxi	us (EN)
mb@Fix	tx@Fix	mb	Fix	us (EN)
mb@Foxi	tx@Foxi	mb	Foxi	us (EN)
ps@Fix	mb@Fix	ps	Fix	us (EN)
ps@Foxi	mb@Foxi	ps	Foxi	us (EN)
ref@Fix	-	ref	Fix	us (EN)
ref@Foxi	-	ref	Foxi	us (EN)
tx@Fix	ref@Fix	tx	Fix	us (EN)
tx@Foxi	ref@Foxi	tx	Foxi	us (EN)
xn@Fix	ref@Fix	xn	Fix	us (EN)

Add Change Delete Import

Select Tier: ft@Fix

Tier Name: ft@Fix

Participant: Fix

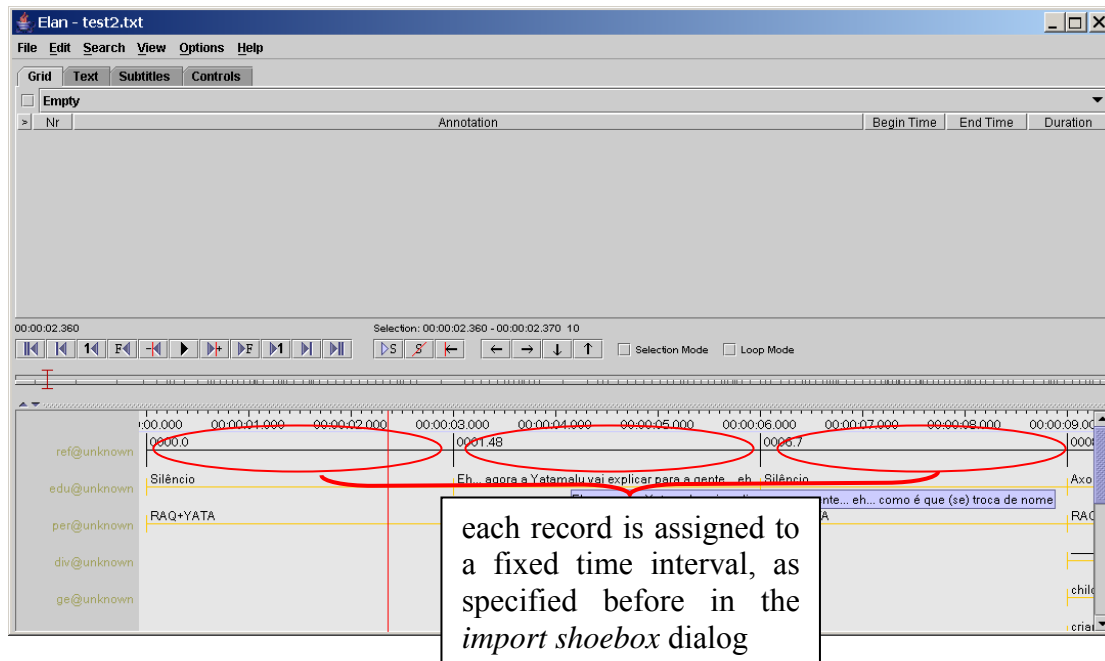
Parent Tier: ref@Fix

Linguistic Type: ft

Default Language: English (United Kingdom)

Change Close

- (3) All SIL IPA characters are converted into Unicode characters during import. If you export the file back into Shoebox (see section 3.2.16), the Unicode characters will be converted back into SIL IPA characters.
- (4) Initially, unless it had the time code information, the imported Shoebox file does not contain information about timing. Instead, ELAN automatically assigns each Shoebox record to a three second time interval, as in the following illustration:



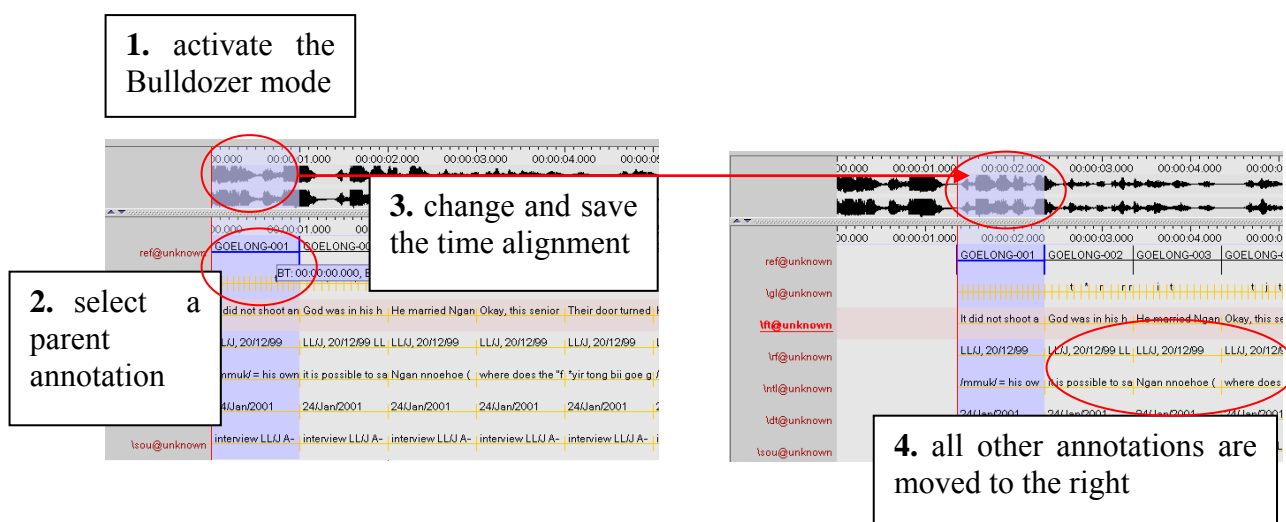
The time alignment has to be done manually for each Shoebox record. Do the following:

1. Activate the Bulldozer mode: Click on Options > Propagate Time Changes > Bulldozer Mode (see section 4.5.7 for the three available modes).

If you do not activate the Bulldozer mode, you will inadvertently overwrite and thereby delete existing annotations. Make sure that Bulldozer Mode is enabled in the Options > Propagate Time Changes menu.

2. Click on the first annotation on the parent tier (i.e., the first Shoebox record). It appears in a dark blue frame.
3. Modify the boundaries of that annotation, so that they are aligned with the correct time interval (see section 4.5.6 for ways of modifying boundaries).
4. Press CTRL+ENTER to apply the new time interval.  
The parent annotation (together with all its referring annotations) is assigned to the new time interval. All other parent annotations are moved to the right.
5. Repeat steps 2 to 4 for each parent annotation.

The following screenshot illustrates steps 1 to 4:



After you have done the time-alignment, you can export the file back to Shoebox – in this case, the time code information will be kept (see section 3.2.16). If you then re-import the file back into ELAN, ELAN automatically assigns the Shoebox records to their correct time intervals.

An imported Shoebox file can be saved as an ELAN file (see section 3.2.4), exported back into Shoebox (see section 3.2.16), or exported as a tab-delimited text (see section 3.2.17).

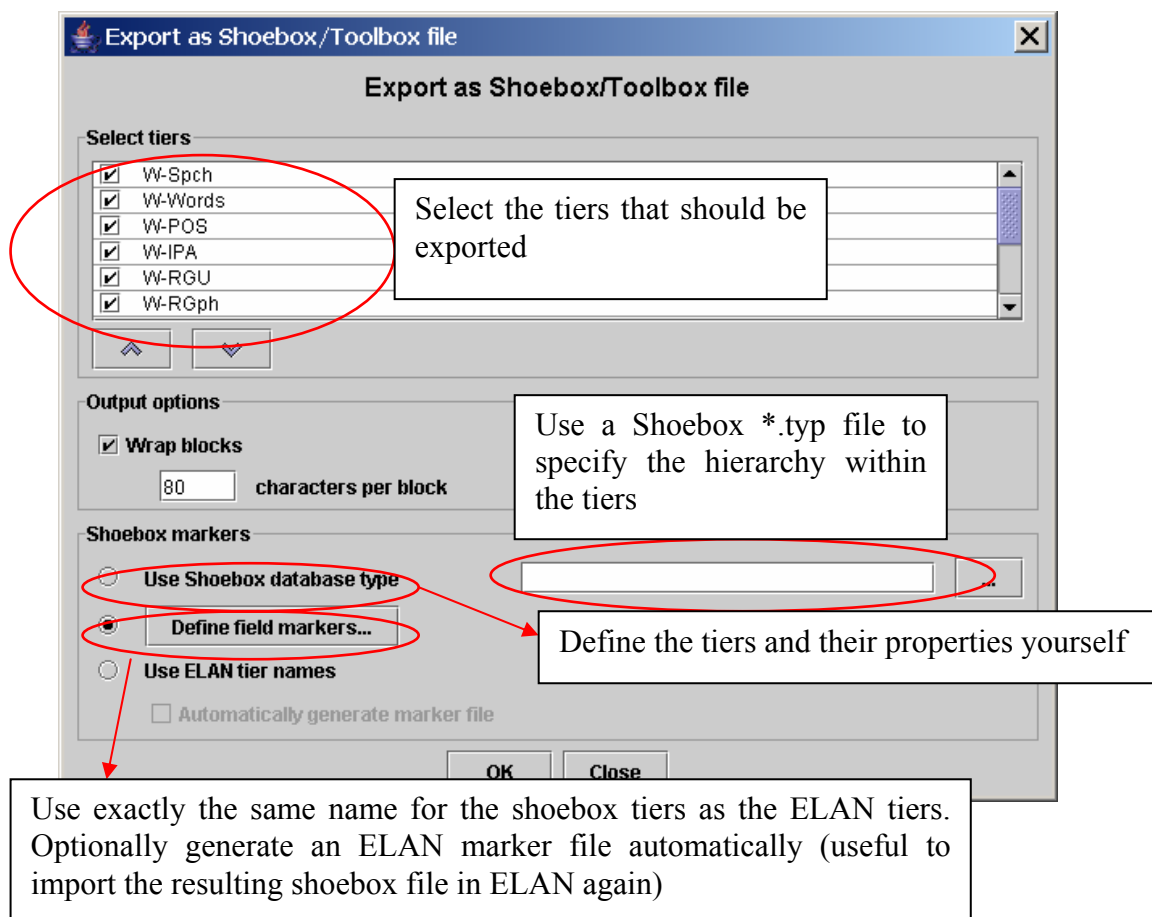
### 3.2.16 Exporting a document to Shoebox

All Shoebox files that were imported into ELAN (see sections 3.2.14) can be exported back into Shoebox. In this case, the time code information is kept.

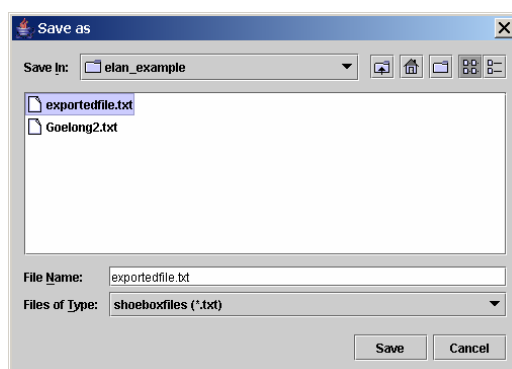
To export a file into Shoebox, do the following:

1. Click on **File** menu.
2. Click on **Export as > Shoebox file ....**

The Shoebox Export dialog box appears. Make a choice and click on **OK** to continue.



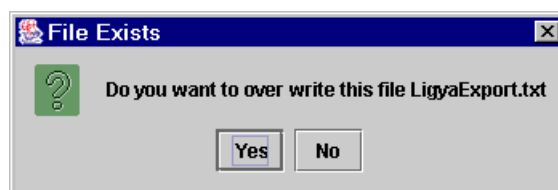
3. Specify the name and directory of the exported file, e.g.:



4. Click **Save** to export the file; otherwise click **Cancel** to exit the dialog box without exporting the file.

The file is exported as a **\*.txt** file.

If there already exists a file of the same name, ELAN will ask you whether or not it should overwrite the existing file, e.g.:



5. Open the exported file in Shoebox.

It contains the following information:

(1) All tiers and annotations.

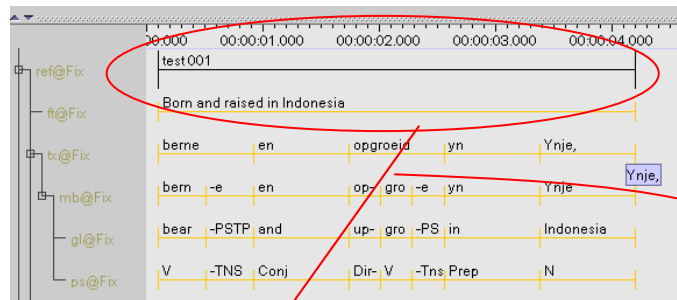
Each ELAN parent annotation (including all its referring annotations) corresponds to one Shoebox record. E.g., in the illustration below, the ELAN parent annotation “Ligya-001” corresponds to the Shoebox record “Ligya-001”.

(2) The time code information for each parent annotation.

Each ELAN parent annotation (i.e., each Shoebox record) contains the additional field markers \ELANBegin and \ELANEnd (i.e., the begin and end time of the parent annotation).

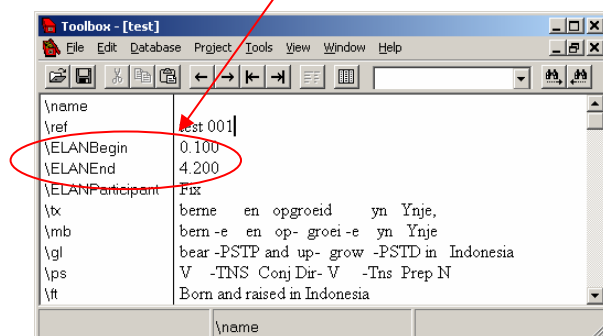
This timecode information allows you to import the Shoebox file back into ELAN, without having to manually re-align the file (see section 3.2.15).

ELAN  
file:



exported  
file:

timecode  
information



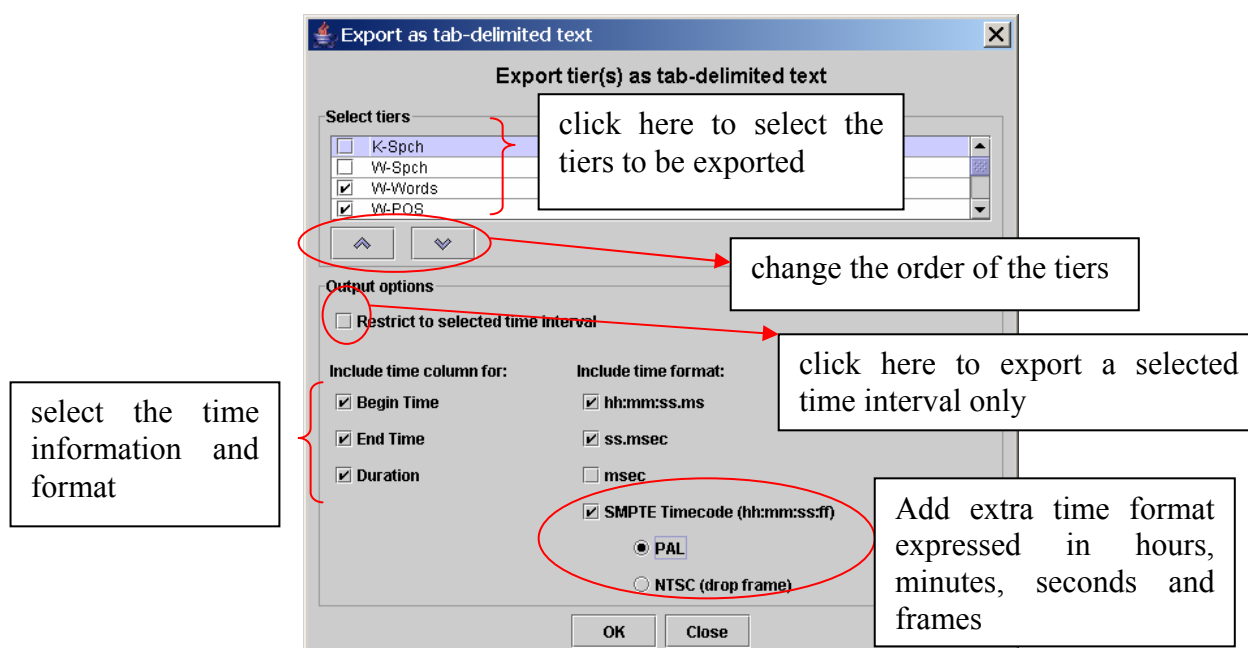
annotation  
test 001

### 3.2.17 Exporting a document as a tab-delimited text file

All documents can be exported into a tabular format for purposes of further analysis and/or printing. This includes documents that were created by ELAN itself (see sections 3.2.1 and 3.2.3) as well as documents that were imported into ELAN from Shoebox (see section 3.2.14) Do the following:

1. Click on File menu.
2. Click on Export as tab-delimited text ....

The Export as tab-delimited text dialog window is displayed, e.g.:



3. By default, ELAN exports all annotations, but it is possible to restrict the export process to selected annotations. The following three options are available:

- (a) Export only those annotations that correspond to a selected time interval. Do the following:
  1. In the ELAN window, select the desired time interval (see section 4.5.1).
  2. In the Export as tab-delimited text dialog window, click in the box to the left of Restrict to selected time interval. A checkmark appears indicating that this option has been selected.

- (b) Export only those annotations that are contained on particular tiers. Do the following:

In the Export as tab-delimited text dialog window, select those tiers that you want to export. A checkmark appears next to any selected tier.

- (c) Export only those annotations that (a) correspond to a particular time interval and (b) are contained on particular tiers. To do this, combine the two steps under (a) and (b) above.



4. Select the time markers you want to export (begin time, end time and/or duration of every annotation unit).
5. Choose the time format (hh:mm:ss.ms, ss.msec, milliseconds and/or SMPTE timecode)

If you choose the SMPTE (hh:mm:ss.ff) format, the selected video standard (PAL or NTSC) just indicates the way seconds and milliseconds are converted to frame numbers. This is independent of the actual video standard of the associated video(s).

6. Click **OK** to start the export process; otherwise click **Cancel** to exit the dialog box without exporting the annotations.
7. Finally you will see a save dialog window. In the Encoding drop down box a text encoding can be selected (either iso-latin, UTF-8 or UTF-16). Make an appropriate choice and click on **Save**.

Some Mac applications, like TextEdit, have difficulties to load UTF-8 encoded files. This is most noticeable for “special” characters, e.g. IPA. Using UTF-16 is recommended in that case.

A message appears to inform you that the file has been exported. The exported file has the extension **\*.txt**.

The exported file contains the following information: begin time of each annotation, end time, total length, content, and tier. It can be opened with any program that can handle tab-delimited texts, e.g., Microsoft Excel.

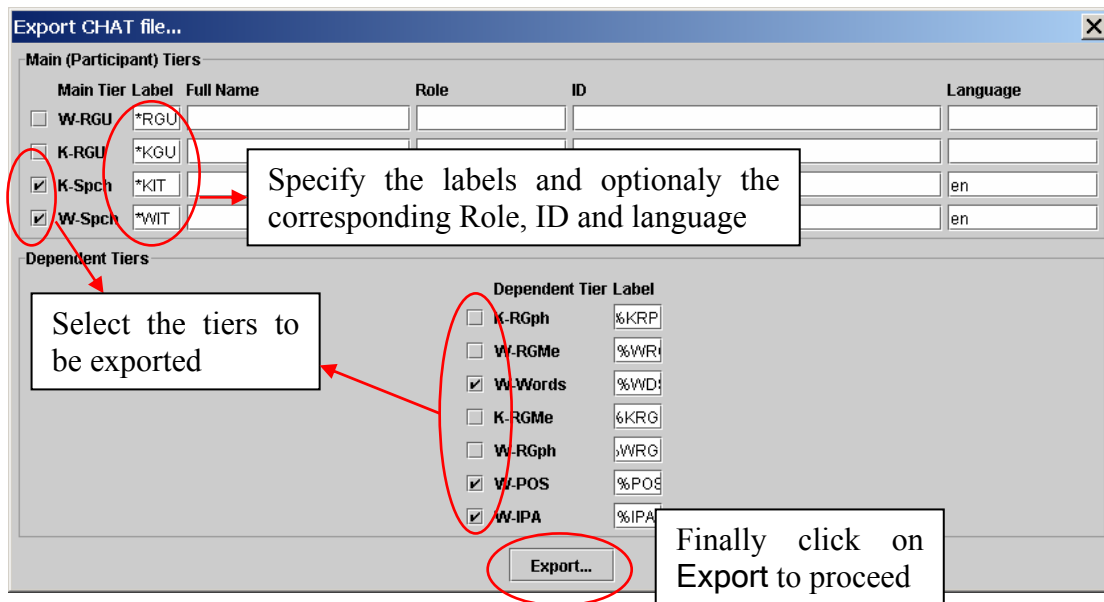
The image shows a screenshot of an Excel spreadsheet with a table of data. The table has columns labeled A through M. The data is organized into rows, with the first row (row 1) containing the following values: A: W-Words, B: 00:00:00.78, C: 0.78, D: 780, E: 00:00:01.34, F: 1.34, G: 1340, H: 00:00:00.56, I: 0.56, J: 560, K: so, L: , M: . The subsequent rows follow a similar pattern, with the first column (A) containing the text 'W-Words' and the other columns containing numerical values. The table is divided into sections by red circles and labels. The labels are: 'begin time' (pointing to column B), 'end time' (pointing to column C), 'total length' (pointing to column D), 'annotations' (pointing to column K), and 'tier labels' (pointing to column A). The spreadsheet interface includes a menu bar at the top with 'File', 'Edit', 'View', and 'Format'. The status bar at the bottom shows 'Sheet1', 'Default', '100%', 'STD', and 'Sum=0'.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	W-Words	00:00:00.78	0.78	780	00:00:01.34	1.34	1340	00:00:00.56	0.56	560	so		
2	W-Words	00:00:01.34	1.34	1340	00:00:01.57	1.57	1570	00:00:00.23	0.23	230	you		
3	W-Words	00:00:01.57	1.57	1570	00:00:01.81	1.81	1810	00:00:00.24	0.24	240	go		
4	W-Words	00:00:01.81	1.81	1810	00:00:02.00	2	2000	00:00:00.19	0.19	190	out		
5	W-Words	00:00:02.00	2	2000	00:00:02.19	2.19	2190	00:00:00.19	0.19	190	of		
6	W-Words	00:00:02.19	2.19	2190	00:00:02.57	2.57	2570	00:00:00.38	0.38	380	the		
7	W-Words	00:00:02.57	2.57	2570	00:00:03.06	3.06	3060	00:00:00.49	0.49	490	Institute		
8	W-Words	00:00:03.06	3.06	3060	00:00:03.20	3.2	3200	00:00:00.14	0.14	140	to		
9	W-Words	00:00:03.20	3.2	3200	00:00:03.34	3.34	3340	00:00:00.14	0.14	140			
10	W-Words	00:00:03.34	3.34	3340	00:00:03.48	3.48	3480	00:00:00.14	0.14	140			
11	W-Words	00:00:03.48	3.48	3480	00:00:03.74	3.74	3740	00:00:00.26	0.26	260			
12	W-Words	00:00:03.74	3.74	3740	00:00:04.09	4.09	4090	00:00:00.35	0.35	350			
13	W-Words	00:00:04.09	4.09	4090	00:00:04.82	4.82	4820	00:00:00.22	0.22	220			
14	W-Words	00:00:04.82	4.82	4820	00:00:05.02	5.02	5020	00:00:00.20	0.2	200	then		
15	W-Words	00:00:05.02	5.02	5020	00:00:05.37	5.37	5370	00:00:00.35	0.35	350	you		
16	W-Words	00:00:05.37	5.37	5370	00:00:05.50	5.5	5500	00:00:00.13	0.13	130	go		
17	W-Words	00:00:05.50	5.5	5500	00:00:05.70	5.7	5700	00:00:00.20	0.2	200	the		
18	W-Words	00:00:05.70	5.7	5700	00:00:05.90	5.9	5900	00:00:00.20	0.2	200	other		
19	W-Words	00:00:05.90	5.9	5900	00:00:06.12	6.12	6119	00:00:00.22	0.22	219	Saint		
20	W-Words	00:00:06.12	6.12	6119	00:00:06.29	6.29	6290	00:00:00.17	0.17	171	Anna		
21	W-Words	00:00:06.29	6.29	6290	00:00:06.59	6.59	6591	00:00:00.30	0.3	301	Straat		
22	W-Words	00:00:06.59	6.59	6591	00:00:06.80	6.8	6802	00:00:00.21	0.2	211	to		
23	W-Words	00:00:06.80	6.8	6802	00:00:07.15	7.15	7150	00:00:00.35	0.35	348	this		
24	W-Words	00:00:07.15	7.15	7150	00:00:07.51	7.51	7510	00:00:00.36	0.36	360	to		
25	W-Words	00:00:07.51	7.51	7510	00:00:07.69	7.69	7690	00:00:00.18	0.18	180	the		
26	W-Words	00:00:07.69	7.69	7690	00:00:08.02	8.02	8020	00:00:00.33	0.33	330	center		
27	W-Words	00:00:08.02	8.02	8020	00:00:08.13	8.13	8130	00:00:00.11	0.11	110	of		

Some versions of Excel seem to have problems importing tab-separated files (white rectangles are shown instead of the column borders). As a workaround you can open the text file first in a text editor (e.g. notepad) and copy and paste the content into Excel.

### 3.2.18 Exporting CHAT files

1. Choosing File > Export as > CHAT file ... will give you the following screen:



2. Fill in the necessary fields.

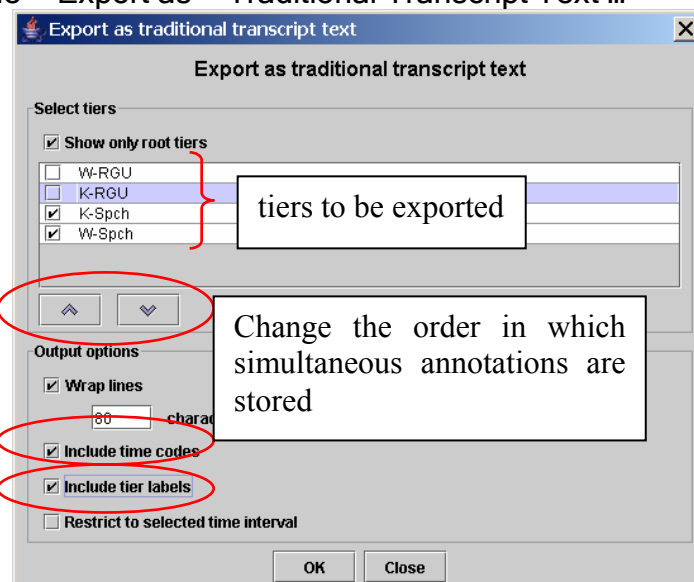
! Note: Chat labels have to contain exactly 3 characters, and must be preceded by \* (for root tiers) or % (for dependent tiers).

3. Click on **Export...**

4. Fill in a chat file name and choose **Save**

### 3.2.19 Exporting traditional transcript files

In some situations a straight-forward list of the annotation units, one after another, can be a handy. For that cause an export option to a “traditional transcript text” has been added to ELAN. In its simplest form it just will create a text file containing the successive annotations of several tiers, in chronological order. This feature can be found under **File > Export as > Traditional Transcript Text ...**



Clicking on OK will produce a file containing the following text:

W-Spch so you go out of the Institute to the Saint Anna Straat.  
TC 00:00:00.780 - 00:00:04.090

K-Spch    yeah  
TC        00:00:04.040 - 00:00:04.400

...

### 3.2.20 Exporting a part of a clip

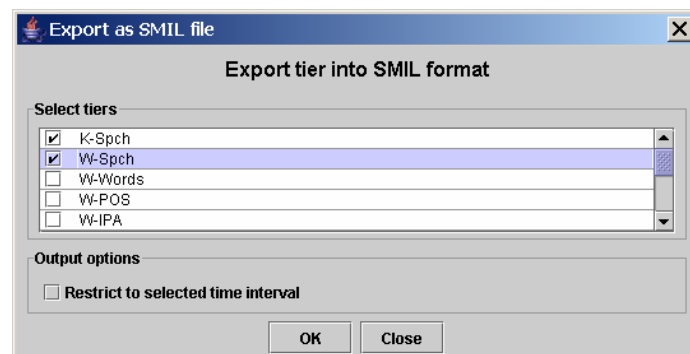
When a command line tool for extracting clips from video files is installed Elan is able to use that tool. At this moment only M2-edit-cl<sup>2</sup> from Mediaware Solutions is supported. If the edit tool is in the user path and a selection is made there is a menu item to export the video clip from the current selection. In that case, follow these steps:

1. Select the part of the video you want to export as a clip
2. Choose **File > Export As > Media Clip**
3. Enter a filename and press **Save**

### 3.2.21 Exporting a SMIL clip

ELAN supports export to SMIL<sup>3</sup>-compliant clips. With a suitable player this enables you to view media files and the associated annotations as a subtitled movie.

1. Select the **File > Export As > SMIL...** menu. This will bring up this dialog box:



2. Select the tiers you want to export. Keep Ctrl pressed and click to select multiple tiers, press Shift and click to select multiple successive tiers.
3. Check **Restrict to selected time interval** if you only want to export the current selection. Otherwise the whole media file and associated annotations will be exported.
4. Click on the suggested filename to change the location where the SMIL clip will be saved.
5. Choose OK to export the clip.

### 3.2.22 Exporting ELAN's document view

To export ELAN's document view (i.e. to make a *screenshot*):

1. choose **File > Export As > Image from Elan Window**

<sup>2</sup> See <http://www.mediaware.com.au/>

<sup>3</sup> For a description of this standard and players see <http://www.w3.org/AudioVideo/>

2. Enter a filename and an extension (\*.jpg, \*.jpeg or \*.png)
3. click on **Save**.

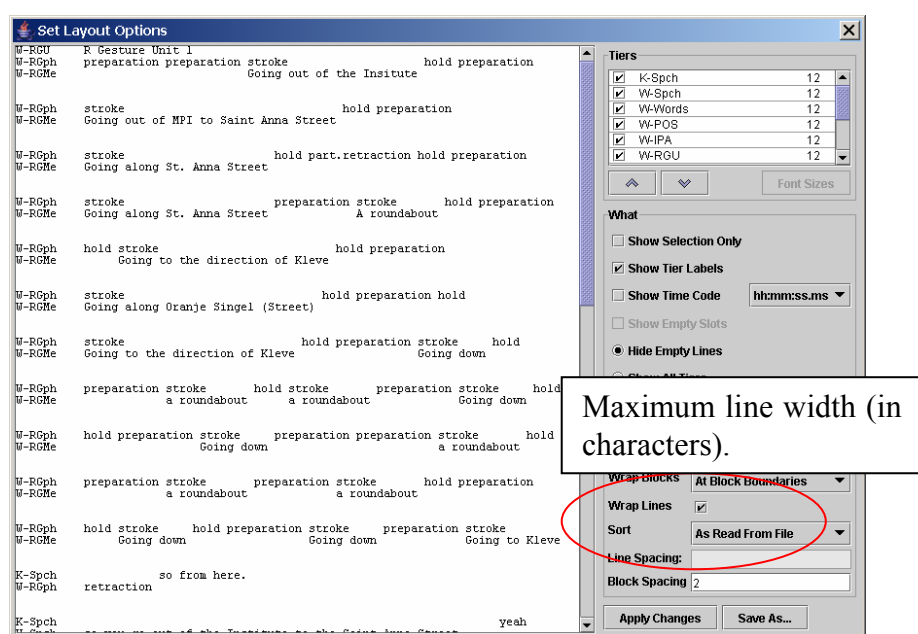
If you are using Windows, it sometimes happens that ELAN's video window is black on the picture created using this function. This can be solved by temporary disabling the hardware video acceleration:

1. Right-click on the desktop
2. choose properties
3. select the Settings tab
4. Click on the advanced... button
5. Select the Troubleshooting tab
6. move the Hardware Acceleration slider tot None

Don't forget to re-enable the hardware acceleration afterwards, because this has a strong effect on the system's graphical performance.

### 3.2.23 Exporting to interlinear text

This function (File > Export as > Interlinearized Text...) is very similar to ELAN's printing system. Therefore more information can be found in section 3.2.12. The main difference is that the width of the exported text depends in this case on the number of characters that fits on one line.

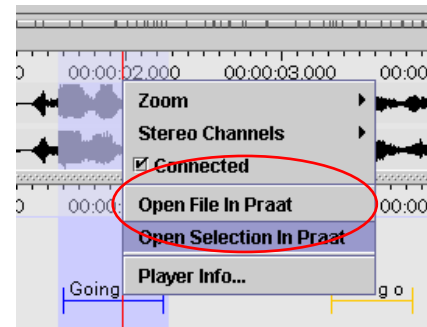


After selecting an appropriate layout click on **Save as** and choose a location and file name. These files can afterwards easily be edited with any text editor (preferably using a fixed-width font).

### 3.2.24 Opening a wave file in Praat

ELAN offers the possibility to open wave files (or a part of them) in Praat<sup>4</sup>. To achieve this, follow the steps below:

1. If desired, make a selection first (See section 0)
2. Right click in the waveform viewer
3. Choose **Open file in Praat** or **Open selection in Praat**
4. If you haven't specified the location of the praat and sendpraat<sup>5</sup> program yet, you will have to locate them now in the file dialog
5. The praat program containing the wavefile appears



Make sure you are using a recent version of Praat (higher than 4.0.5), otherwise this feature will not work.

*Note for advanced users: compiling SendPraat on linux*

- Make sure a C compiler (like gcc) is installed.
- You will also need the X11 or Xorg development packages (they can be found in a package called xorg-x11-devel or something likewise)
- Replace `if0` with `if1` in `sendpraat.c`, as described on <http://www.fon.hum.uva.nl/praat/sendpraat.html>
- Link the executable with the X11 libraries when you compile sendpraat:  

```
cc -o sendpraat sendpraat.c -L /usr/X11R6/lib -lX11
```

<sup>4</sup> See <http://www.fon.hum.uva.nl/praat/>

<sup>5</sup> For downloads and installation instructions see <http://www.fon.hum.uva.nl/praat/sendpraat.html>.

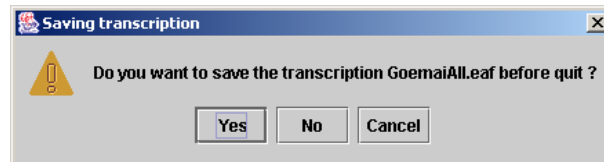
### 3.2.25 Exiting ELAN

To close the active ELAN window, choose one of the following options:

(a) Either use the menu options:

1. Click on **File** menu.
2. Click on **Exit**.

If you exit ELAN without having saved the changes (see section 3.2.4), the **Saving transcription** dialog window appears, e.g.:



Click **Yes** to save the changes; click **No** to close the window without saving; click **Cancel** to return to ELAN.

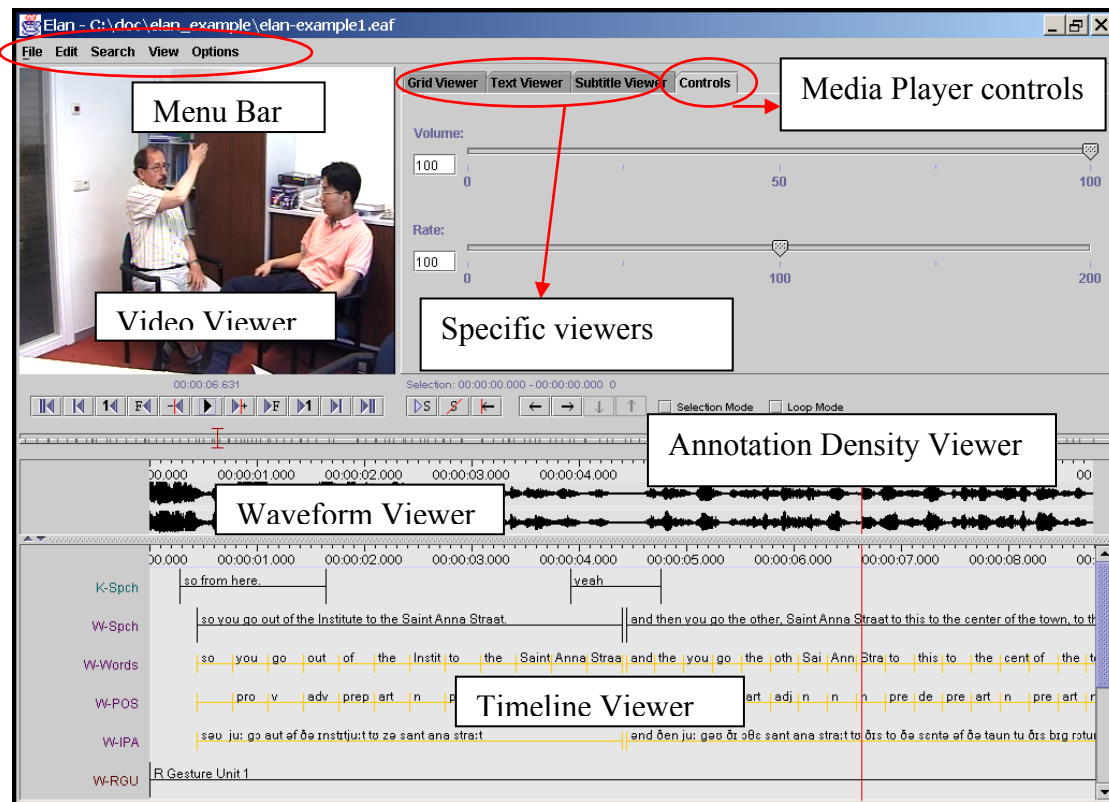
Note that the **Exit** option does not necessarily close all open windows, only the active window! This behavior is subject to change in future versions.

## 3.3 Screen display and navigation through a document

Whenever you open, create or import a document, the **ELAN** window appears. This section introduces you to the setup of the **ELAN** window (section 3.3.1), and explains the navigation through it (sections 3.4 and 3.5).

### 3.3.1 Basic Information: The ELAN window

The ELAN window displays the Menu bar, the Media Player options, and up to six Viewers, e.g.:



All Viewers are synchronized and thus display the same point(s) in time. I.e., whenever you access a point in time in one of the Viewers, all the other Viewers will immediately jump to the corresponding point in time.

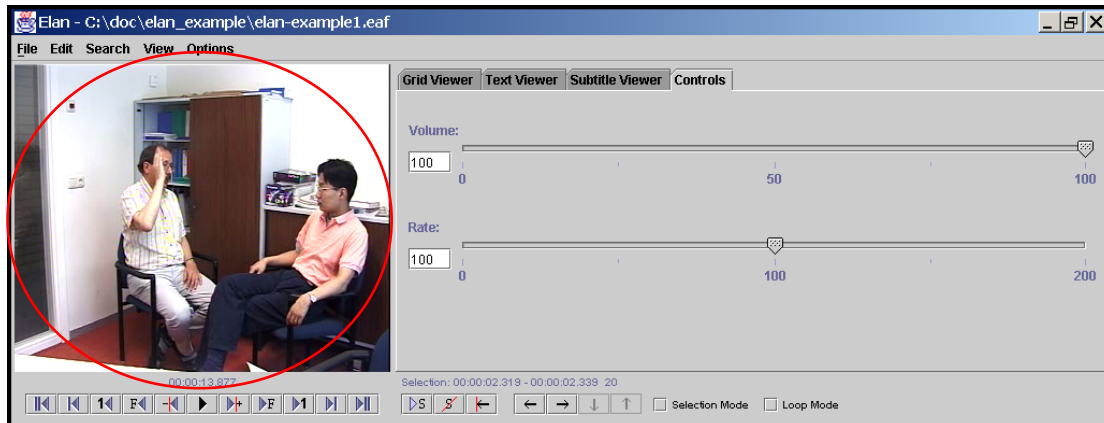
In all Viewers, color coding is used to facilitate the orientation in the document.

This section introduces the setup of the Viewers, the Menu bar, the Media Player options and the color coding. Detailed information about how to navigate through the ELAN window follows in the subsequent sections.



### 3.3.2 The Video Viewer

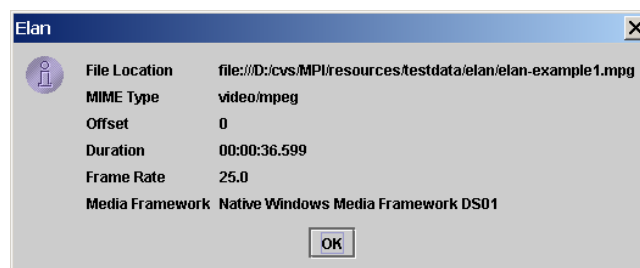
The Video Viewer displays the video image (of the \*.mpg/\*.mov file).



Note that you can right click on the video viewer to *detach* it, i.e. create a separate window for the video. The size of that window can be altered, in contrast to the normal video viewer within the ELAN window. To re-attach the video window, right click on it and select *attach*.

! Note: In the very unlikely case that you encounter problems while playing video files, uncheck **Options > Use native media platform** in ELAN's initial screen.

Right clicking in the video window and selecting **Player Info...** will display a dialog with information about the video file, e.g.:



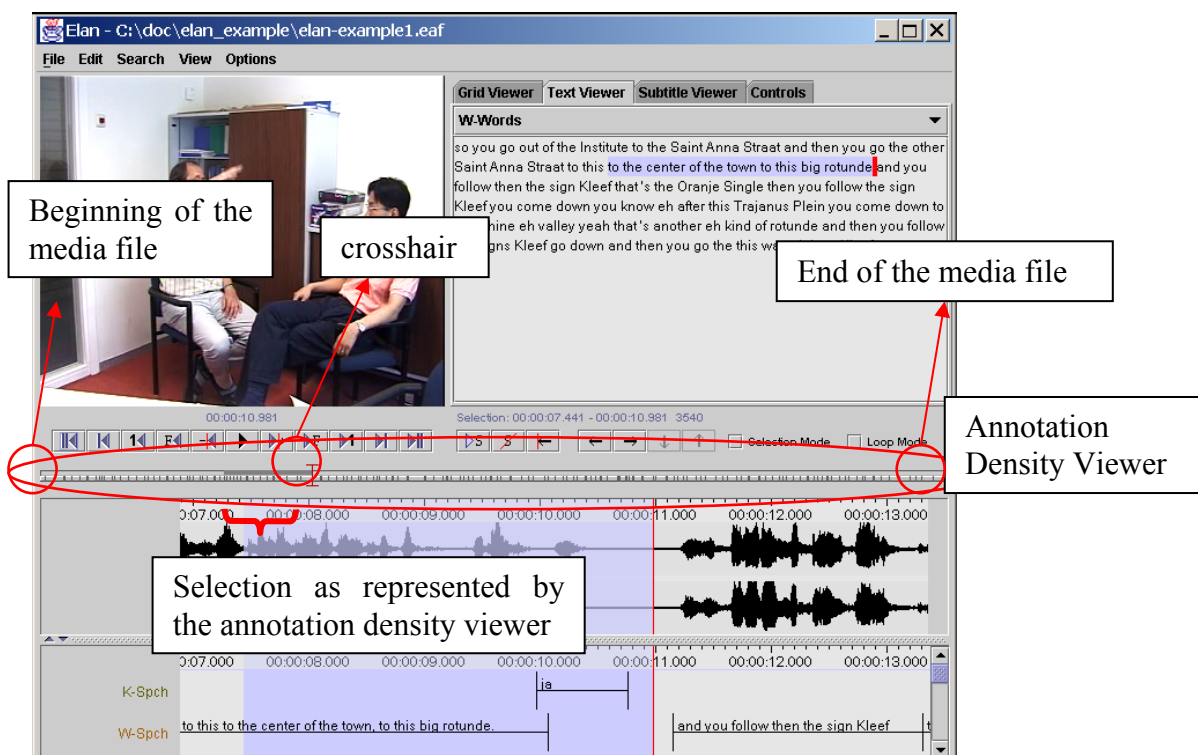
A static picture containing the currently displayed frame can be stored using the context menu of the video window (right click > **Save Current Frame as Image...**)

Saving a static picture might sometimes cause a freeze of the program on MacOS X. This issue is being investigated.

### 3.3.3 The Annotation Density Viewer

The Annotation Density Viewer is a kind of a timeline which allows you to:

- Navigate through the whole media file. The length of this viewer always corresponds to the whole media file, so e.g. by clicking in the middle you will always go to the middle of the media file. The selection is represented as a small grey bar.
- See how many annotations are concentrated at a particular moment of the time (the *Annotation Density*). The more annotations available for a particular moment, the more the Annotation Density bar is filled. This can be useful to track places in the media file that still have to be annotated.



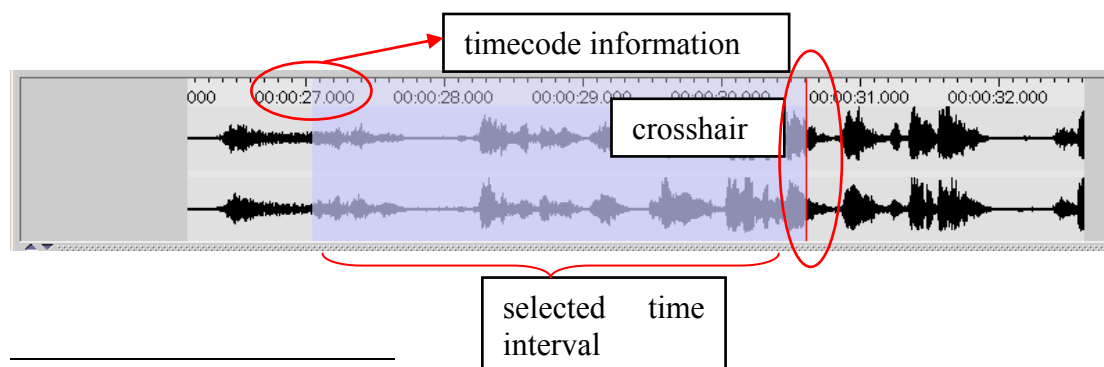
### 3.3.4 The Waveform Viewer

The Waveform Viewer displays the waveform of the audio file<sup>6</sup> (\*.wav). Above the waveform, timecode information is displayed. While listening to the sound, a red vertical bar, the crosshair, moves through the waveform and indicates which part of the waveform corresponds to the current point in time. Furthermore, whenever you have selected a time interval, the corresponding part of the waveform will be highlighted in light blue color.

At any time, you can press ALT and drag the time axis for a panning effect (i.e. go to the left to go back in the time or to the right to go further).

In the case of video files, the waveform is only displayed if there exists an additional \*.wav file (see section 3.1). If this is not the case, the Waveform Viewer will not be available.

On slower machines, the Waveform Viewer may not always update properly when moving to the next page.

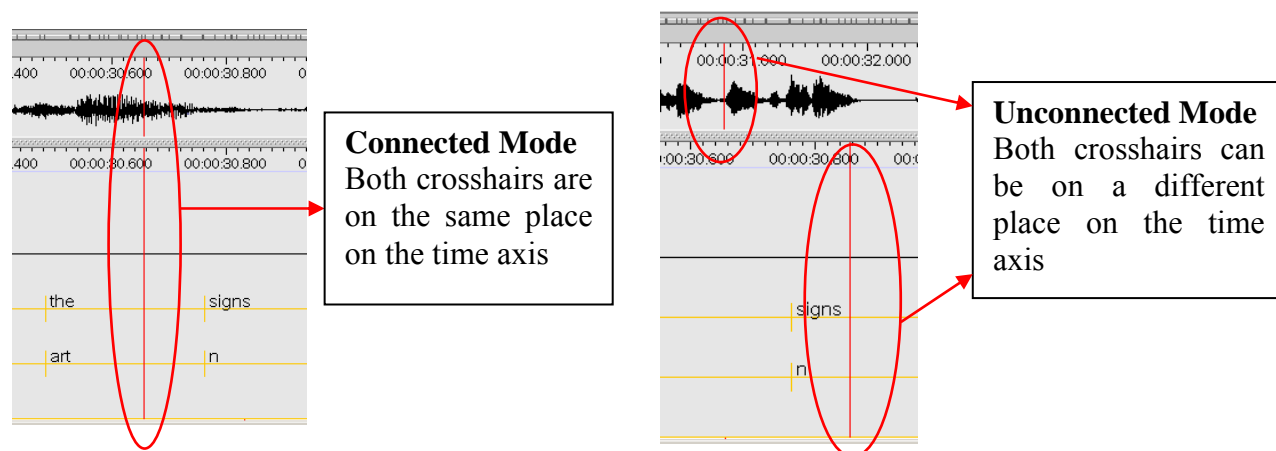


<sup>6</sup>Different resolutions are supported: 8 bits (mono and stereo), 16 bits (mono and stereo) and 24 bits (stereo). Both PCM and A-law encoded wave files can be loaded.

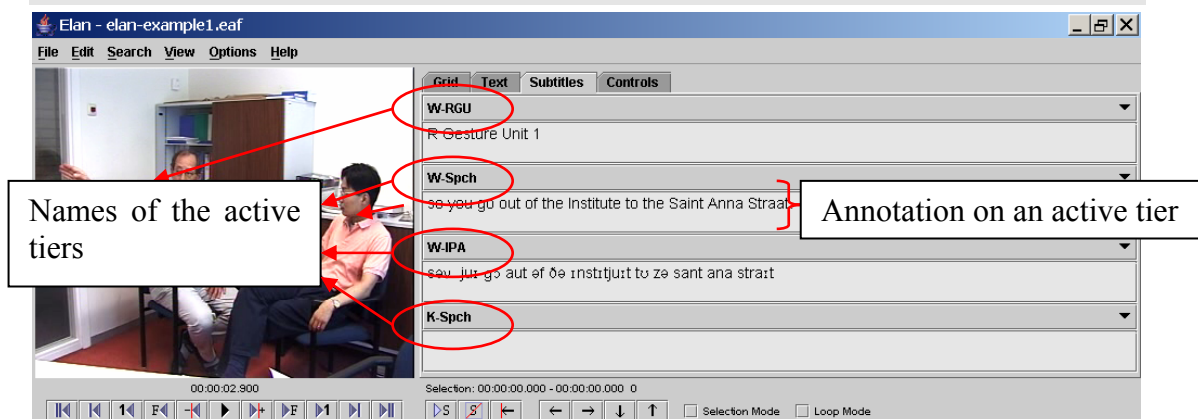
The Waveform Viewer supports 3 modes. You can select the active mode by a right click on the Waveform Viewer. In the menu *Stereo Channels*, the following options are available:

1. **Separate.** 2 Waveforms are showed, one for each channel.
2. **Merged.** The 2 channels are merged and the result, one waveform is displayed.
3. **Blended.** Both channels are displayed on 1 waveform, differences are designated with colors.

Another option in the context-menu (right click) of the Waveform Viewer is *connected*. If this option is checked, the timescale of the Waveform Viewer and the Timeline Viewer are connected:



### 3.3.5 The Subtitle Viewer

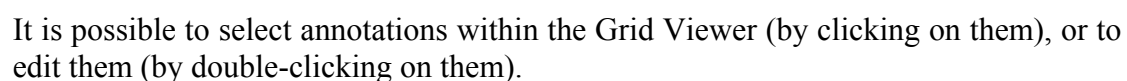


You can turn on the subtitle viewer for a tier by selecting that tier from the pulldown menu in the tab **Subtitle Viewer**. During playback, the Subtitle Viewer displays the annotations of the selected tiers at the current media time, both during playback and in static situations.

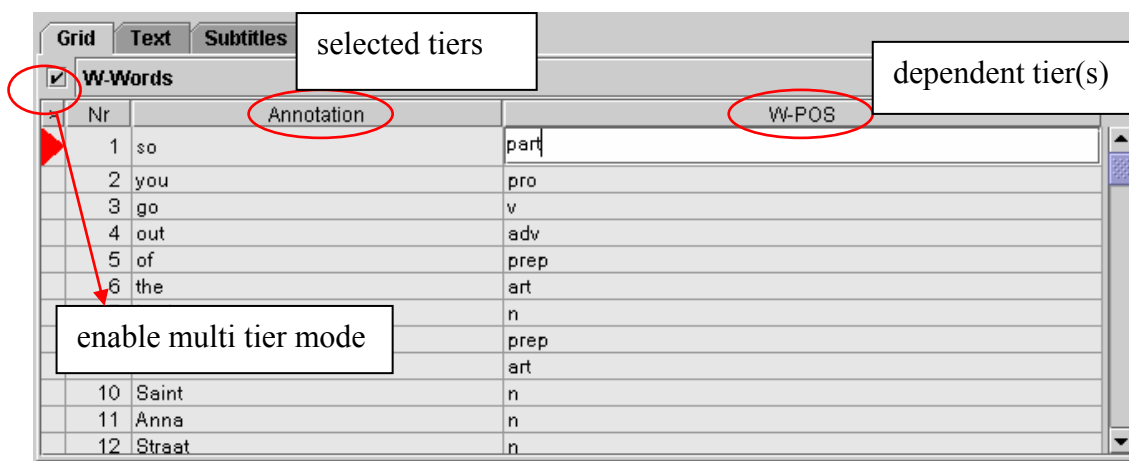
It is possible to display up to four of the tiers as a subtitle in the Subtitle Viewer:

- 
- The screenshot shows the ELAN software interface. At the top is a menu bar with 'File', 'Edit', 'Search', 'View', and 'Options'. Below the menu bar is a toolbar with various playback controls. The main window is divided into several sections. On the left is a video player showing two men, one of whom is playing a flute. To the right of the video player is a list of annotation tracks. The tracks are labeled as follows: 'W:Words', 'W:POS', 'W:IPA', 'W:Empty', 'W:K-Spch', 'W:Spch', 'W:Words', 'W:POS', 'W:IPA', 'W:RSG'. A red circle highlights the 'W:Words' track, which contains the text 'you' and 'pro'. Below the tracks is a timeline with multiple tracks, including 'W:IPA', 'W:Words', 'W:Spch', 'W:POS', and 'W:RSG'. The 'W:Words' track shows the text 'you go out of the institute to the Saint Anna Street' and 'and then you go the other side'.

The Grid Viewer displays the content as well as the begin and end time of all annotations from a single tier. You have to activate this Viewer through selecting the Grid tab next to the video window.



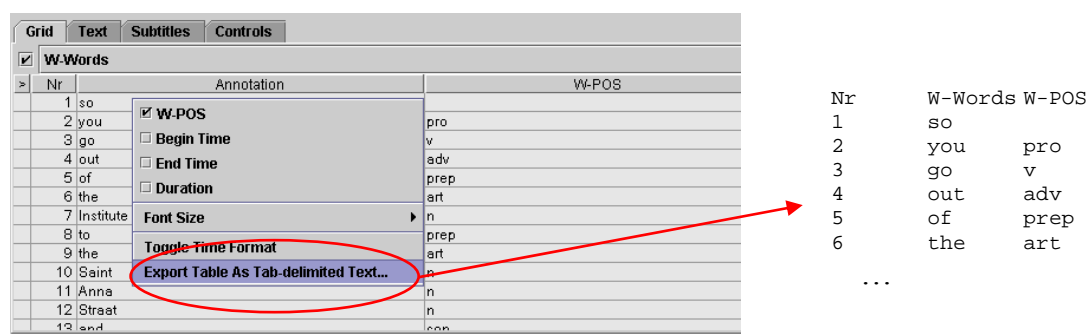
44



### Exporting from the Grid Viewer

The results as shown in the Grid Viewer can be exported to a (tab-separated) text file, for later manipulation with e.g. a spreadsheet. Follow these steps:

1. First make a selection of the tiers you want to be displayed (and thus exported)
2. Right click on the Grid Viewer and select **Export Table as tab-delimited text...**
3. Enter the name of the file to be created and choose **Save**

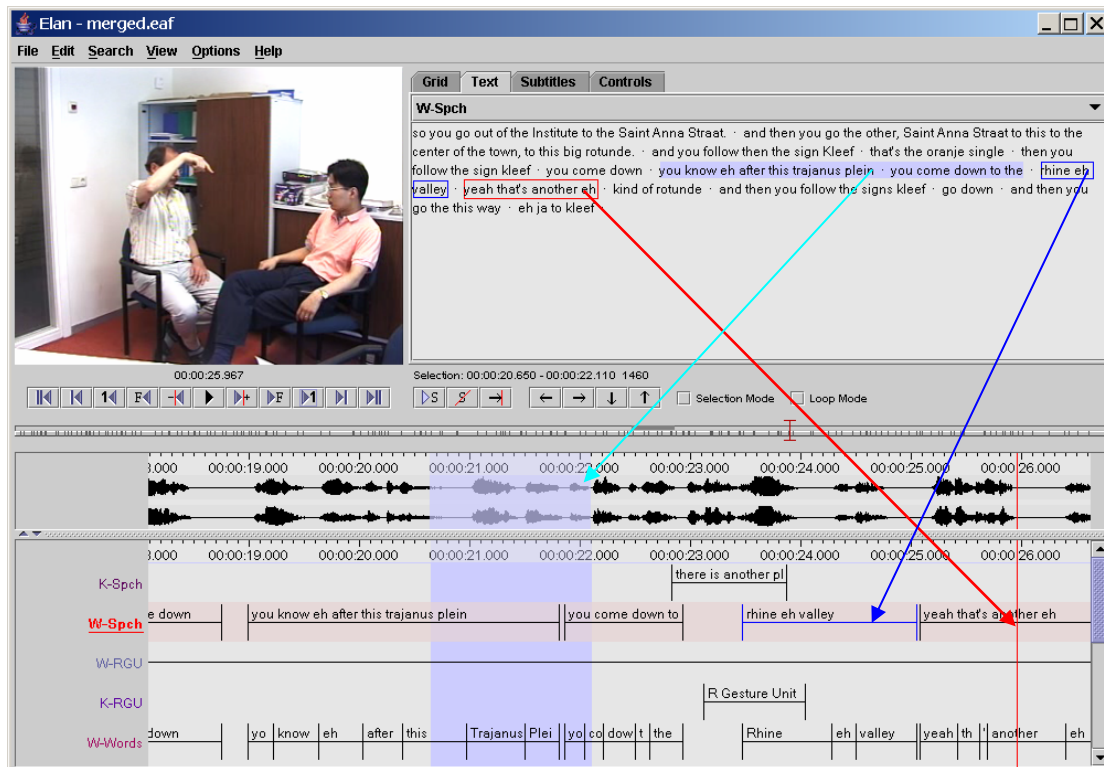


### 3.3.7 The Text Viewer

The text viewer shows all values of all annotations on a selected tier as ongoing text. Within this viewer, there are 3 kinds of highlighting shown:

- Text inside a red box: the value of the annotation on the selected annotation tier that matches the current point of time.
- Text inside a dark blue box: the active annotation (see also next section)
- Text with a light blue background: the selected time interval.

Optionally, you can make the annotation boundaries visible in the text viewer. Right click in the text viewer and select **Toggle visualization** to enable this. The boundaries are marked by a dot.

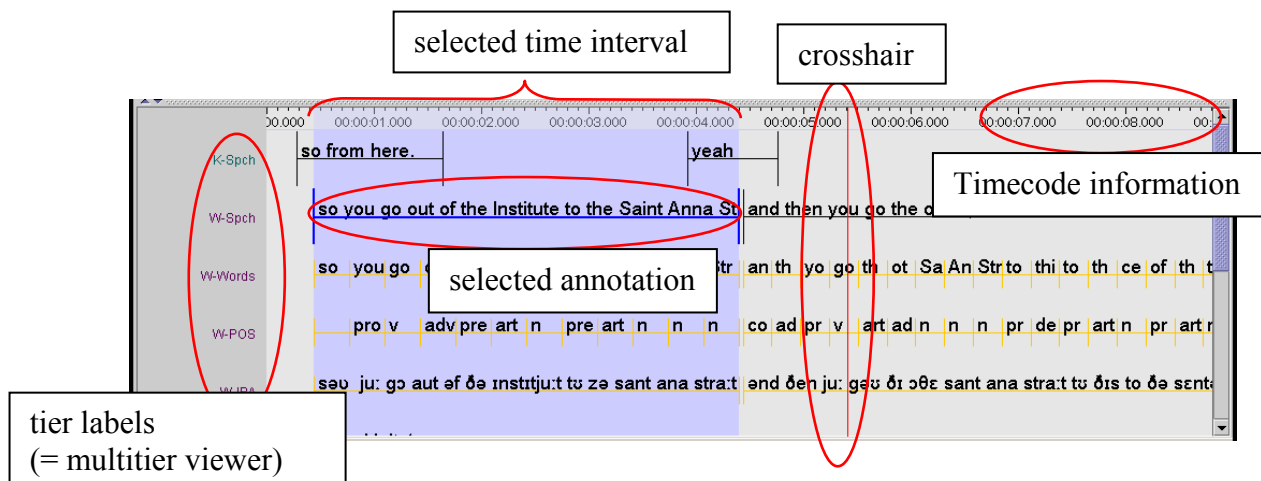


### 3.3.8 The Timeline Viewer and the Interlinear Viewer

All annotations can be displayed in the Timeline Viewer or the Interlinear Viewer. Only one of the two Viewers can be switched on at a time.

The Timeline Viewer is always shown when a document is opened in ELAN. It displays the tiers and their annotations, whereby each annotation corresponds to a specific time interval. Above the tiers, a timescale is displayed. During playback, a red vertical bar, the crosshair, moves through the annotations and indicates the current point in time. Furthermore, whenever you have selected a time interval, it will be highlighted in light blue color; and whenever you have selected an annotation, this becomes the active annotation and will be highlighted in a dark blue frame.

If desired the latter can also be indicated with a bold line. To activate this, right click on an annotation somewhere in the timeline viewer and check the **Active Annotation Bold** box in the context menu.



In the Timeline Viewer you can (a) select and modify time intervals (see section 0) and (b) enter annotations (see section 4.6).

The Interlinear Viewer offers an alternative perspective on the tiers and their annotations. It shows parent-child relations between annotations using vertical text alignment (interlinearization). You can enable it selecting the **Show Interlinear Viewer** radio button when you right-click on the tier name's panel and select **Viewer**. Switching it on, will automatically switch off the Timeline Viewer.

The following screenshots compare how information is displayed in the two Viewers.



## Timeline Viewer

Elan - C:\doc\elan\_example\elan-example1.eaf

File Edit Search View Options

Grid Viewer Text Viewer Subtitle Viewer Controls

W-Spch

so you go out of the Institute to the Saint Anna Straat and then you go the other, Saint Anna Straat to this to the center of the town, to this big rotunde. and you follow then the sign Kleef that's the oranje single then you follow the sign kleef you come down you know eh after this trajanus plein you come down to the rhine eh valley yeah that's another eh kind of rotunde and then you follow the signs kleef go down and then you go the this way eh ja to kleef

00:00:02.420

Selection: 00:00:00.440 - 00:00:04.400 3960

W-Words

so you go out of the Institute to the Saint Anna Straat and then you go the oth Sai Ann Stra to this to the cent of the t

W-POS

pro v adv prep art n prep art n n n con adv pro v art adi n n n pre de pre art n pre art n

W-IPA

səu ju: go aut əf ðə Instɪtju:t to zə sɛnt ənə stræt and ðen ju: gəʊ ðə sɛnt ənə stræt to ðɪs to ðə sɛnt əf ðə taun tu ðɪs bɪg rɒtʊnd

W-RGU

R Gesture Unit 1

K-Spch

so from here. yeah

active  
annotation (on  
tier W-words)

Selected time interval

## Interlinear Viewer

Elan - C:\doc\elan\_example\elan-example1.eaf

File Edit Search View Options

Grid Viewer Text Viewer Subtitle Viewer Controls

W-Spch

so you go out of the Institute to the Saint Anna Straat and then you go the other, Saint Anna Straat to this to the center of the town, to this big rotunde. and you follow then the sign Kleef that's the oranje single then you follow the sign kleef you come down you know eh after this trajanus plein you come down to the rhine eh valley yeah that's another eh kind of rotunde and then you follow the signs kleef go down and then you go the this way eh ja to kleef

00:00:02.420

Selection: 00:00:00.440 - 00:00:04.400 3960

W-Words

so you go out of the Institute to the Saint Anna Straat and then you go the oth Sai Ann Stra to this to the cent of the t

W-POS

pro v adv prep art n prep art n n n con adv pro v art adi n n n pre de pre art n pre art n

W-IPA

səu ju: go aut əf ðə Instɪtju:t to zə sɛnt ənə stræt and ðen ju: gəʊ ðə sɛnt ənə stræt to ðɪs to ðə sɛnt əf ðə taun tu ðɪs bɪg rɒtʊnd

W-RGU

R Gesture Unit 1

K-Spch

so from here. yeah

click here to move  
backward/forward to  
the next annotation



Whenever the Interlinear Viewer is switched on, it displays an annotation block (i.e., an independent, time-alignable parent annotation together with its referring annotations, see section 4.1). To move forward/backward to the next block, click on the **arrow** icons at the top of the Viewer. During playback, the Viewer automatically moves forward to the next annotation block.

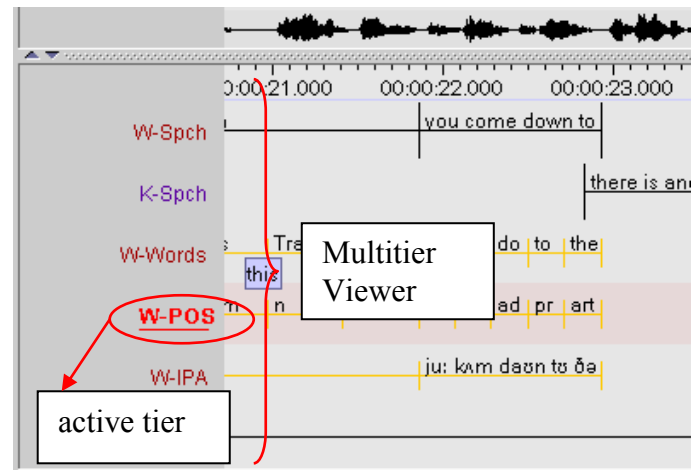
The Interlinear Viewer differs from the Timeline Viewer in that it does not allow to modify the time interval or to enter new annotations. It is similar to the Timeline Viewer in that it allows to edit existing annotations.

### The Tier Name Panel

The Tier Name Panel gives an overview of the different existent tiers. One of these tiers (underlined and in red) is the *active tier*, which means that new annotations will be added to this tier (when pressing ALT+ N).

To make a tier the active tier, choose one of the following actions:

- Double click on a tier label.
- Right click in the Timeline or Interlinear Viewer and choose Active Tier.
- Select the active tier with the keyboard shortcut CTRL + ARROW UP/DOWN



To select the tiers to display (and their order) see sections 3.3.16 and 3.3.18.

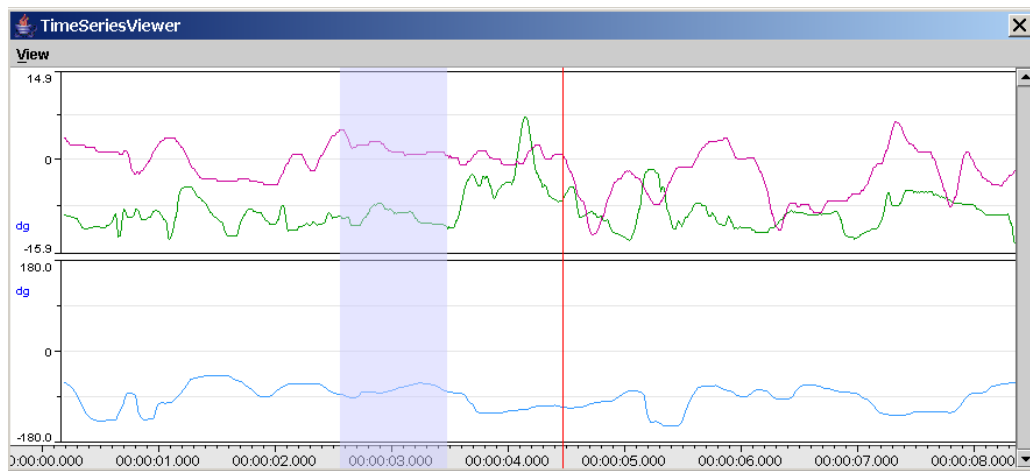
### 3.3.9 The Timeseries Viewer

The Timeseries Viewer can display time series data as line graphs. Like the Timeline and Waveform viewer, it has a horizontal timescale bar, a red vertical crosshair indicating the media time and a light blue rectangle to highlight the selected time interval. It has also the same zoom and pan options.

It can display multiple “trackpanels” and each trackpanel can display multiple “tracks”. Trackpanels and tracks can be added and removed via a popup menu. Each trackpanel derives its value range (vertical axis) from one of the tracks. The viewer has a facility to transfer data from a track to annotation values. Based on the time intervals of the annotations on a chosen (time-alignable) tier, the minimum, maximum or average of the data within these intervals of the selected track will be copied to annotations on a dependent, symbolically associated tier.

The Timeseries Viewer will be created after at least one supported timeseries data file<sup>7</sup> has been associated to the transcription via menu **Edit > Linked Files** and then the tab “Linked Secondary Files”. These data files can be synchronized to the media files in the “Media Synchronization Mode”.

<sup>7</sup> Currently supported file formats are a proprietary .log file produced by MPI CyberGlove software and a special kind of plain text (.txt) file, containing a time-value pair on each line. Software developers can add support for other formats by implementing a Service Provider Interface (more information can be found in the source code release notes).



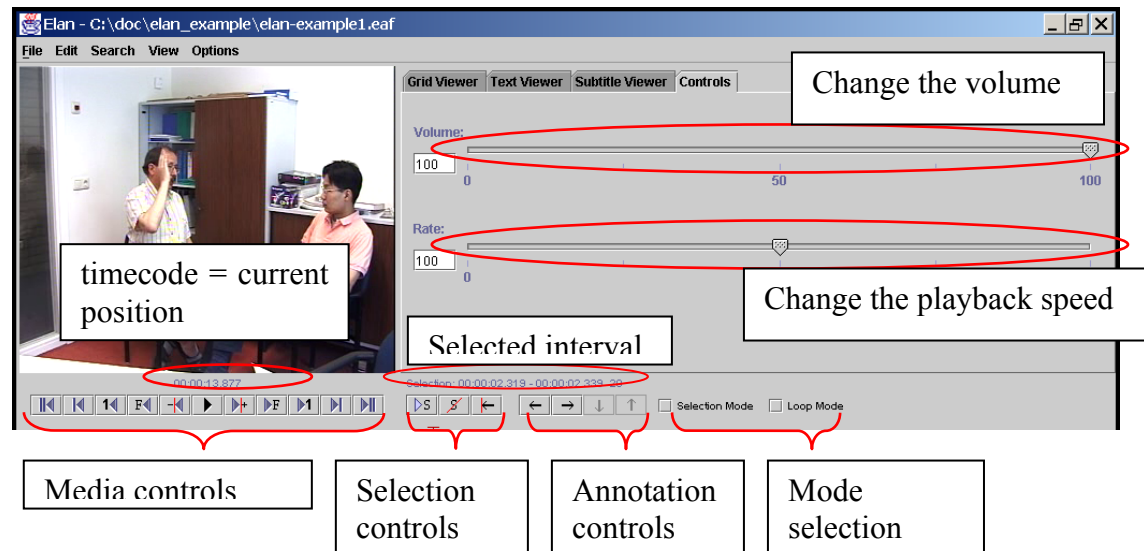
### 3.3.10 The Menu options

The following Menu options are available at the top of the ELAN window:

- **File:** use this menu to open, create, save, im-/export or exit a document (see section 3) and to configure automatic backups.
- **Edit:** use this menu to define, modify and delete annotations, tiers and linguistic types (see section 4).
- **Search:** use this menu to search for text (see section 4.7).
- **View:** use this menu to get an overview of the tier dependencies (see section 4.1) and the shortcut keys.
- **Options:** use this menu to (de)activate the Bulldozer mode (see section 4.5.7), to choose between annotation mode and synchronization mode and to select a language and video standard.
- **Help:** use this menu to read information about ELAN.

### 3.3.11 The Media Player options



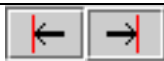
With the Media Player options, you can control the playback of the file. The following options are available at the bottom and at the left side of the ELAN window:







#### Media Controls

Icon	Meaning	Shortcut
	Go to the beginning of the video/audio fragment	CTRL+B
	Go to the previous scroll view (make the beginning point of the current timeline view the end point)	CTRL+PAGE UP
	Go back one second	SHIFT+LEFT
	Go back one frame ( = 40 ms for PAL, 33.4 ms for NTSC)	CTRL+LEFT
	Go back one “pixel” on the timeline viewer (smallest unit, depends on the zoom factor of the timeline viewer, default value 10 ms)	CTRL+SHIFT+LEFT
	Start / Pause the playback	CTRL+SPACE
	Go to the next “pixel” on the timeline viewer (smallest unit)	CTRL+SHIFT+RIGHT
	Go to the next frame	CTRL+RIGHT
	Go to the next second	SHIFT+RIGHT
	Go to the next scroll view	CTRL+PAGE DOWN
	Go to the end of the media fragment	CTRL+E


## Selection Controls

Icon	Meaning	Shortcut
	Play the selected interval.	SHIFT+SPACE
	Clear the selection.	CTRL+C
	Move the crosshair to the begin / end of selection	CTRL+/,

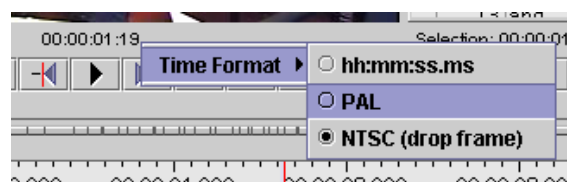
## Annotation Controls

Icon	Meaning	Shortcut
	Go to the previous annotation on the active annotation tier	ALT+LEFT
	Go to the next annotation on the active annotation tier	ALT+RIGHT
	Go to the annotation above.	ALT+UP
	Go to the annotation below.	ALT+DOWN

## Selection Mode

Icon	Meaning	Shortcut
<input type="checkbox"/> Selection Mode	While playing, select an interval automatically	CTRL+K
<input checked="" type="checkbox"/> Loop Mode	Keep playing the selected interval (if used together with <i>play selection</i>  )	CTRL+L

- The timecode can be displayed in the following formats: "hours:minutes:seconds.milliseconds" and "hours:minutes:seconds.frames". It can be used to access points in time (by clicking on it and entering a value in "hh:mm:ss.ms" format). To change the display format, right click on it and select a format from the menu. The SMPTE timecode formats PAL and NTSC only indicate the way time values are converted to frame numbers; this is independent of the actual video standard of the associated video(s).



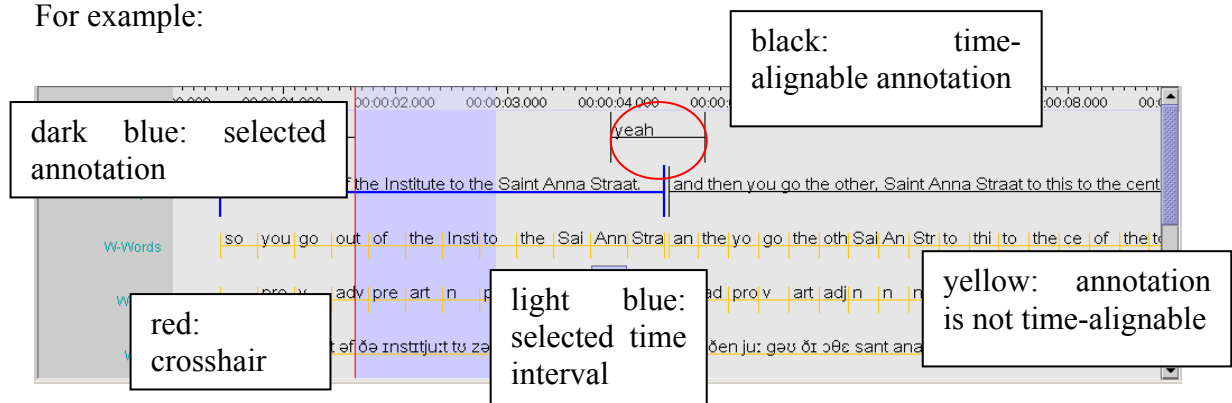
- The sliders available when the Controls tab is selected allow you to control the playback rate and the volume.

### 3.3.12 The color coding

In all its displays, ELAN makes use of recurring colors in order to facilitate the orientation in the document. The following colors are used:

- Red: Position of the crosshair (i.e., current point in time);
- Light Blue: Selected time interval;
- Dark Blue: Active annotation.
- Black with long segment boundaries: Annotations that can be aligned to the time axis.
- Yellow with short segment boundaries: Annotations that cannot be aligned to the time axis.

For example:



## How to change the display of the ELAN window

The ELAN window setup as described and illustrated in section 3.3.1 above is the default display. But you can easily change the display according to your needs. The following options are available:

- increasing/decreasing the size of the ELAN window (section 3.3.13);
- switching Viewers on/off (section 3.3.14);
- increasing/decreasing the size of Viewers (section 3.3.15);
- switching tiers on/off (section 3.3.16);
- rearranging the order of tiers (section 3.3.17);
- displaying a tier in the any of the tab pane's viewers (section **Error! Reference source not found.**);
- changing the time resolution (section 3.3.19);
- changing the font size (section 3.3.20);
- de-attach/re-attach the video window (section 3.3.2).

### 3.3.13 Increasing/decreasing the size of the ELAN window

The size of the ELAN window can be increased or decreased. Do one of the following:

- (a) Go with the mouse to the borders of the ELAN window. The mouse will turn into a double-headed arrow. Click and move it to increase/decrease the size of the window.
- (b) In the top right corner of the ELAN window, click on the **Maximize** icon to activate the full-screen modus; click on the **Restore Down** icon to return to the previous size.

### 3.3.14 Switching Viewers on/off

Depending on the type of media file, ELAN automatically displays three Viewers (Video, Waveform and Timeline Viewer). Furthermore one can choose an additional viewer in the tabs on the right of the video Viewer: a text viewer, a grid viewer or subtitle viewer. Note that it is not possible to activate both the Timeline and the Interlinear Viewer at the same time.

If a media file is not available (e.g., the \*.mpg/\*.mov file in case of audio data, or the \*.wav file in case of some video data), the corresponding Viewer is not available either.

### 3.3.15 Increasing/decreasing the size of Viewers

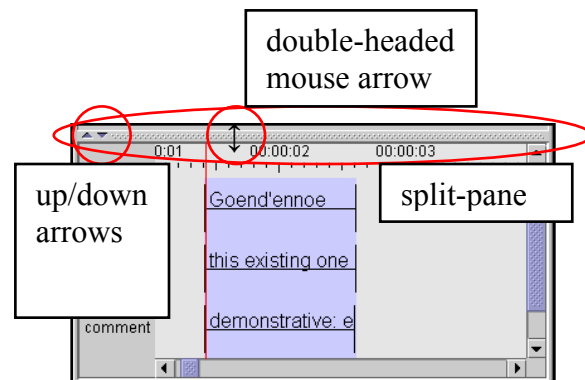
The size of all Viewers (except for the Video Viewer) can be increased and decreased relative to the size of other Viewers. Do one of the following:

**(a) Use the up/down-arrows of the split-pane.**

Click on the up/down-arrow to increase/decrease the size of the corresponding Viewer.

**(b) Use the mouse.**

Go with the mouse to the split-pane. The mouse will turn into a double-headed arrow. Click and move it up/down to increase/decrease the size of the corresponding Viewer.



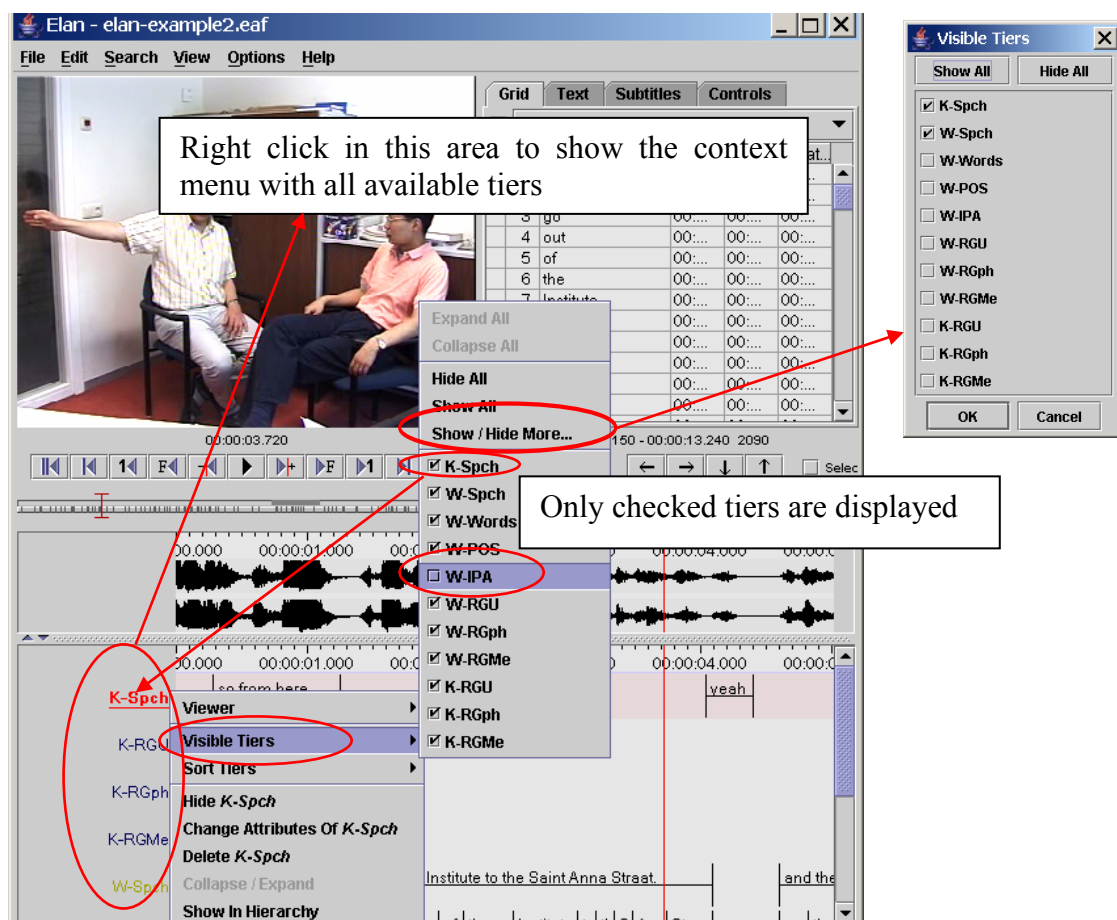
### 3.3.16 Switching tiers on/off

By default, ELAN automatically displays all available tiers, but each tier can be switched on or off manually, allowing you to focus only on the tiers of interest for the task at hand.

To switch tiers on/off, do the following:

1. In the Timeline Viewer, right-click in the tier name panel.
2. In the pull-down menu select the submenu **Visible Tiers** and (un)check the tiername

Switching off a tier can be done directly by right clicking on its name and selecting **hide <tiername>** from the pull down menu. Alternatively you can open a window containing all tier names by selecting **Show/Hide More...** in the popup menu.

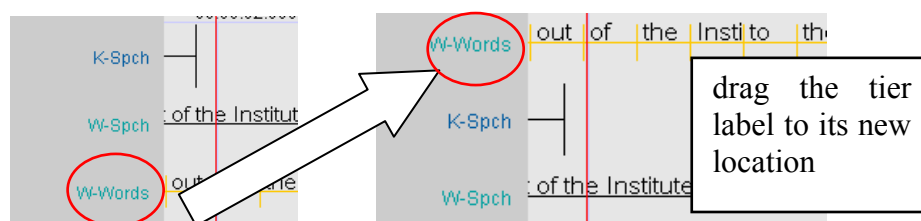


If you switch a tier on, it will be put on the place where you clicked.

### 3.3.17 Rearranging the order of tiers

Within the Timeline or Interlinear Viewer, you can rearrange the order in which the tiers are displayed. Just drag the tier label to its new location. The tiers will be displayed in the new order.





If you exit the document, ELAN will save the order of tiers in the following way: first, all activated tiers (in the order as they appear in the Timeline or Interlinear Viewer), followed by all non-activated tiers in alphabetical order.

### 3.3.18 Sorting tiers

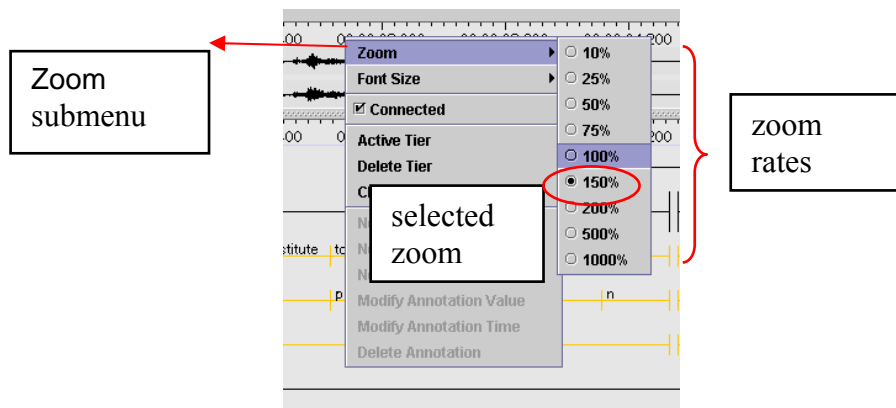
The order of the visible tiers in the timeline and interlinear viewer can be altered. To achieve this, right click in the tier name panel and select the submenu **Sort Tiers**. Then choose one of the following options:

<p>Unsorted: no specific order.</p>	
<p>Sort by Hierarchy: display a tree with the hierarchical structure of the tiers</p>	
<p>Sort by Linguistic Type: group tiers by their linguistic type</p>	
<p>Sort by Participant: group tiers with annotations of a single participant</p>	

### 3.3.19 Changing the time zoom

The default zoom for the Waveform and the Timeline Viewer is 100%, corresponding to 10 milliseconds per pixel. The zoom can be changed simultaneously for both Viewers. Do the following:

1. Click with the right mouse button on either the Waveform Viewer or the Timeline Viewer.
2. Go to **Zoom**. The following menu appears:

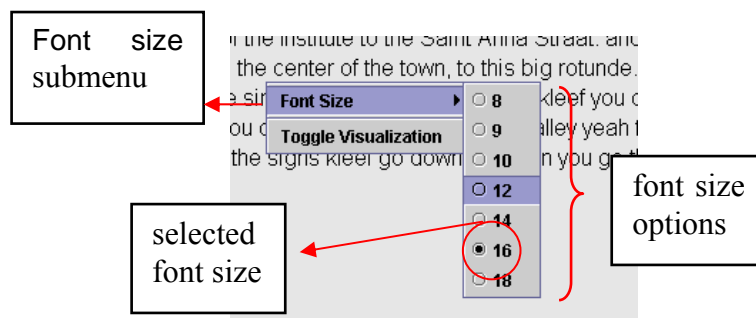


3. Click on a zoom rate to select it. A checkmark appears next to the selected zoom rate.
  - Click on a lower percentage to get a finer zoom.
  - Click on a higher percentage to get a wider zoom.

### 3.3.20 Changing the font size

The default font size is 12 pt., but it can be changed separately for the different annotation viewers in ELAN. Do the following:

1. Right click on one of the viewers (Grid, Subtitle, Text, Timeline, Interlinear Viewer).
2. Go to **Font size**. The following menu appears:



3. Click on a font size to select it. A checkmark appears next to the selected font size.

### 3.4 How to navigate through a document

ELAN supports the following options for navigating through a document:

- accessing points in time (section 3.4.1);
- going to a selection (section 3.4.2);
- jumping step by step through a document (section 3.4.3);
- navigating using the Timeline Viewer (section 3.3.11)
- navigating using the Grid Viewer (section 3.4.4);
- navigating using the Interlinear Viewer (section 3.4.5).

All Viewers are synchronized in time, i.e., when you navigate to a specific point or selection in one Viewer, all other Viewers will immediately jump to the corresponding point or selection:

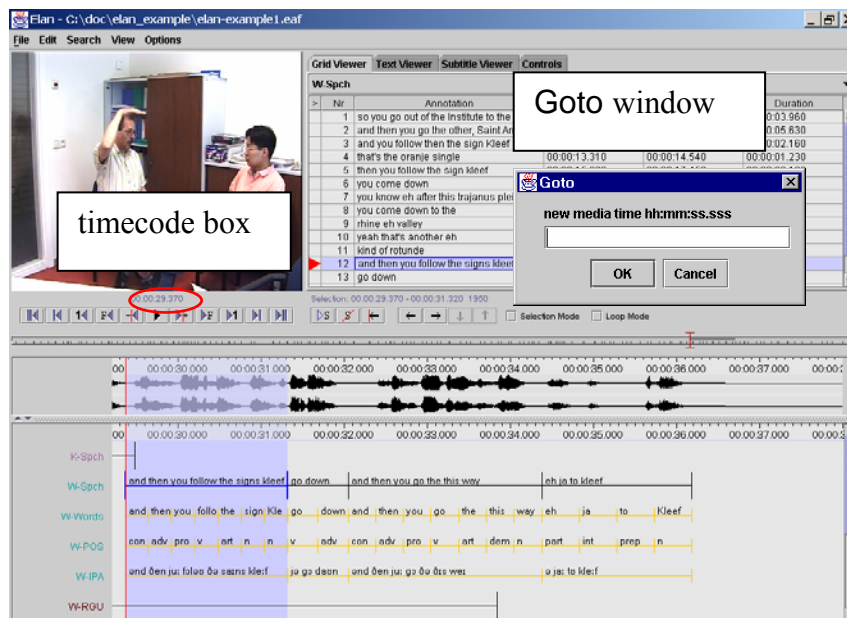
- The Video Viewer will display the corresponding video frame
- The Waveform Viewer will display a crosshair at the corresponding location in the waveform.
- The Subtitle, Timeline and Interlinear Viewers will display the corresponding annotation(s).

### 3.4.1 Accessing points in time

To access a point in time, do one of the following:

**(a) Use the timecode box.**

1. Click on the timecode above the media playback controls (left side of the ELAN window). The Goto dialog window appears.



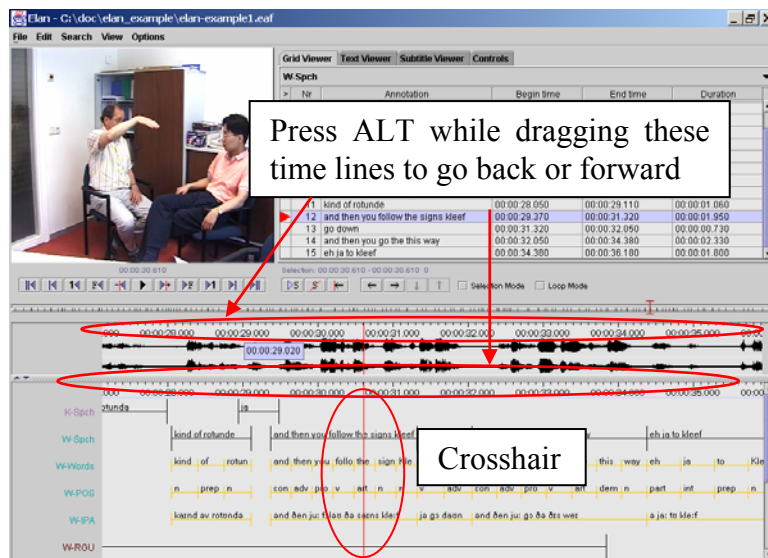
2. Enter the timecode in the following format: “hh:mm:ss.sss” (hours:minutes:seconds.milliseconds). The crosshair jumps immediately to the selected timecode.

You can omit part of the timecode information:

- If the digits are “00”, you can omit “hours:”, “hours:minutes:”, or “hours:minutes:seconds.”
- It is optional to specify “.milliseconds”.

The character preceding the milliseconds is a dot, not a colon.

## (b) Use the crosshair





Click somewhere in the Text, Subtitle, Timeline, Waveform or Grid Viewer. The crosshair will jump to that point. By holding the ALT button and dragging the time axis to the left or to the right you can scroll through the annotations.

## (c) Use the media control buttons

See section 3.3.11

### 3.4.2 Going to the borders of a selection

If you have made a selection (see section 0), you can move the crosshair to the begin or end of that selection from anywhere within the file.

To jump to the begin, click on the  button, which is part of the Selection Controls. Then the button will show an arrow in the other direction , which brings the crosshair to the end of the selection.

### 3.4.3 Jumping step by step through a document

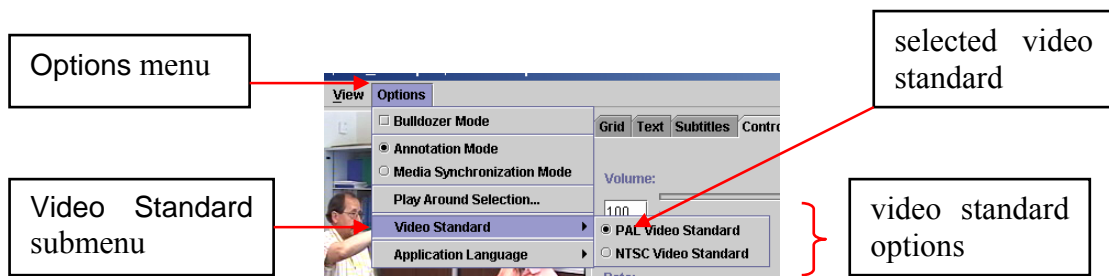
It is possible to move the crosshair back and forth step by step. To achieve this, use one of the media controls as described in 3.3.11.

When using the next/previous frame control, you can alter the step size. This is useful in order to work with a “natural” frame duration, depending on the video format that is used (i.e. 25 frames/second for PAL or +/- 30 frames/second for NTSC).

Do the following:

Click on Options menu.

1. Go to Video Standard. The following menu appears:



Select an option from the menu:

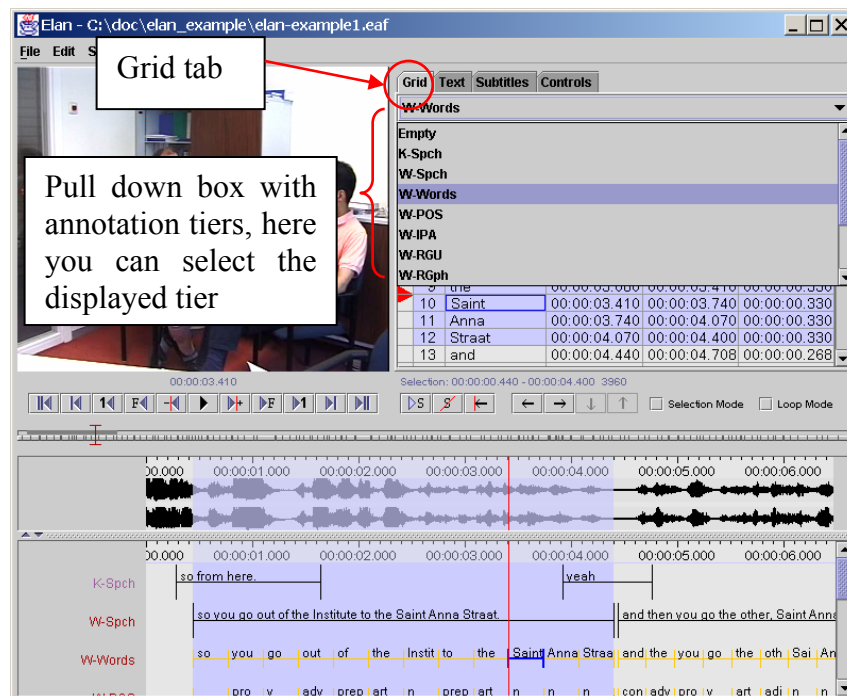
- PAL: The 1 frame stepsize for video data corresponds to one PAL frame (40 ms)
  - NTSC: The 1 frame stepsize for video data corresponds to one NTSC frame (33 ms)
2. Click on a Video Standard mode to select it. A radio bullet appears next to the selected step mode.

If you are using the DirectX framework (the default for the Windows version of ELAN), the video mode is detected automatically and therefore the manual Video Standard setting will be ignored in this case.

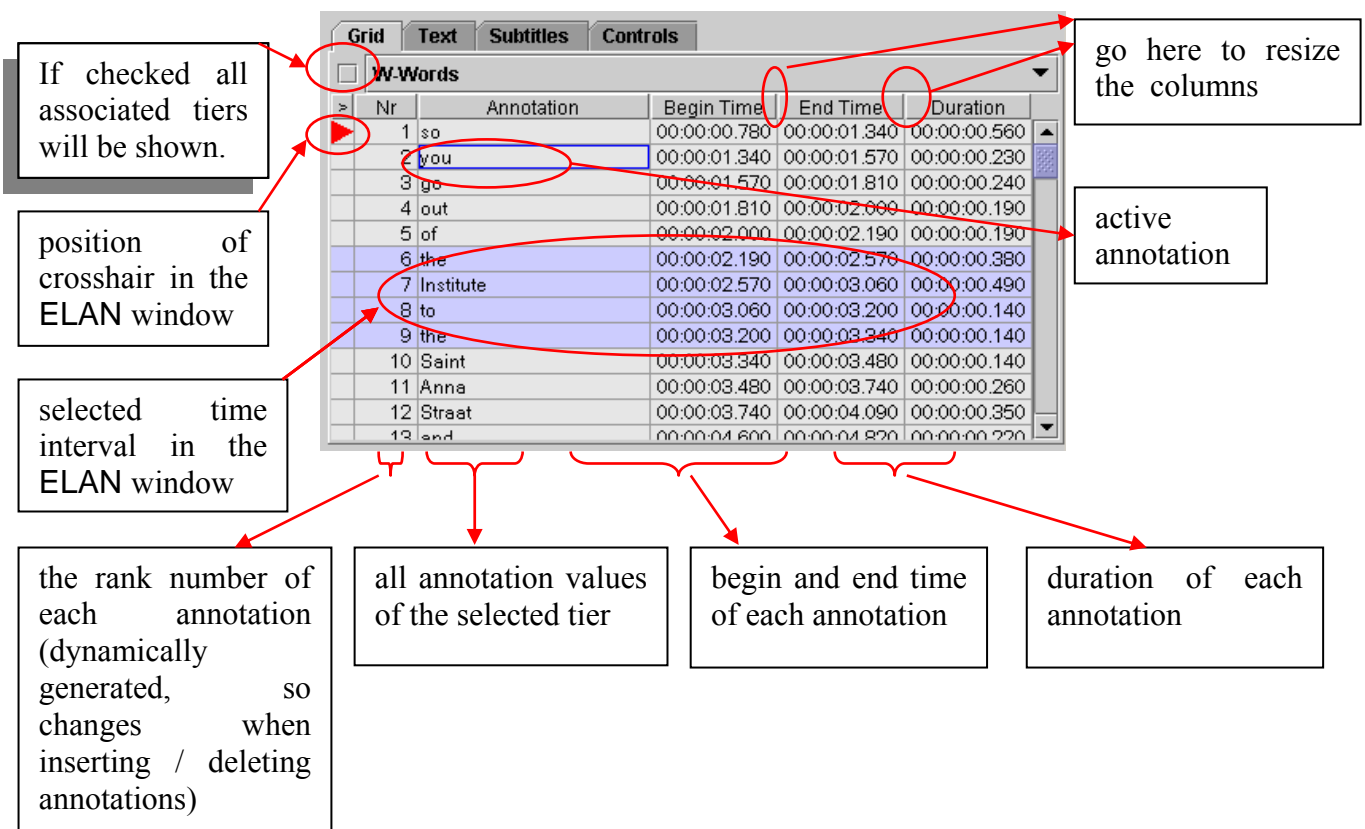
### 3.4.4 Viewing a list of annotations within one tier (Grid Viewer)

ELAN allows you to list all annotations from a single tier, and then select one annotation and jump to it.

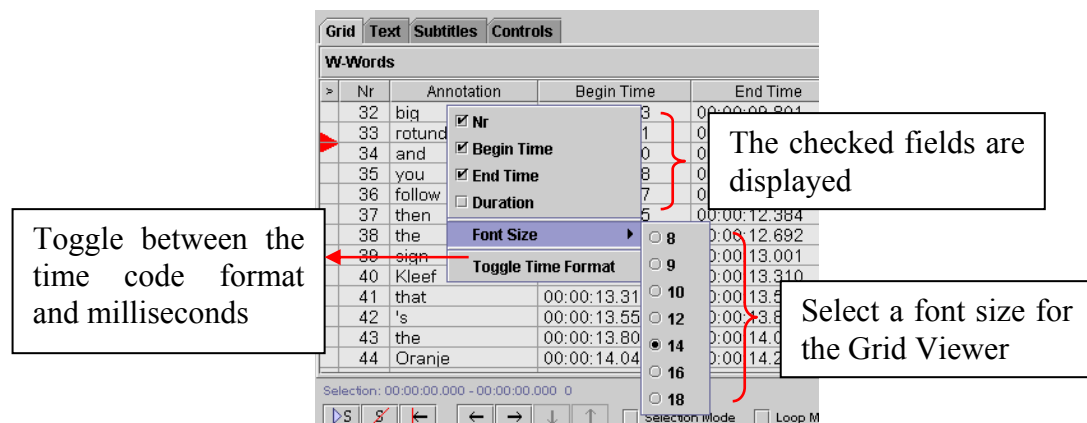
To activate the Grid Viewer, click on the the Grid tab:



The structure of the Grid View window is as follows:



You can change the Grid Viewer's interface at any moment by right clicking in the Grid Viewer. A context menu will appear:



In this context menu, you can choose between the following options:

- choose the columns displayed in the Grid Viewer
- changing the Grid Viewer's font size
- toggle between the time code format (hh:mm:ss.mmm) and milliseconds

You can use the Grid Viewer window to navigate to an annotation in the ELAN window. You have the following two options:

**(a) Move the crosshair to the begin of an annotation.**

Do the following:

1. In the Grid Viewer window, click with the mouse button on an annotation.  
A red triangle appears next to the annotation in the Grid Viewer window, and the crosshair moves to the beginning of that annotation in the ELAN window.

**(b) Select a time interval.**

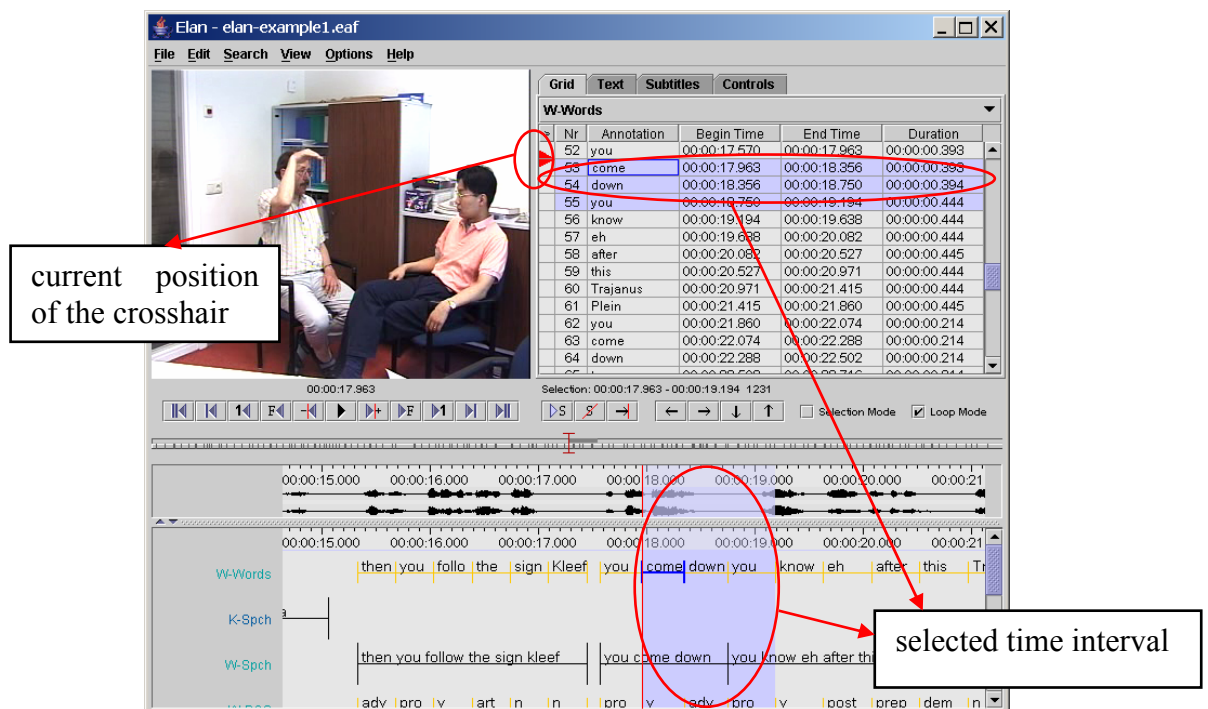
Do the following:

1. In the Grid Viewer window, click with the mouse button on the first annotation that you want to select.
2. Keep the mouse button down and drag the mouse to another annotation.

In the Grid Viewer window, all selected annotations are highlighted in light blue color. In all other windows, the corresponding time interval is selected and highlighted in light blue color (starting with the beginning of the first annotation and ending with the endpoint of the last).

Selecting a time interval also changes the current time. This happens implicitly by moving the crosshair to the begin of the annotation.

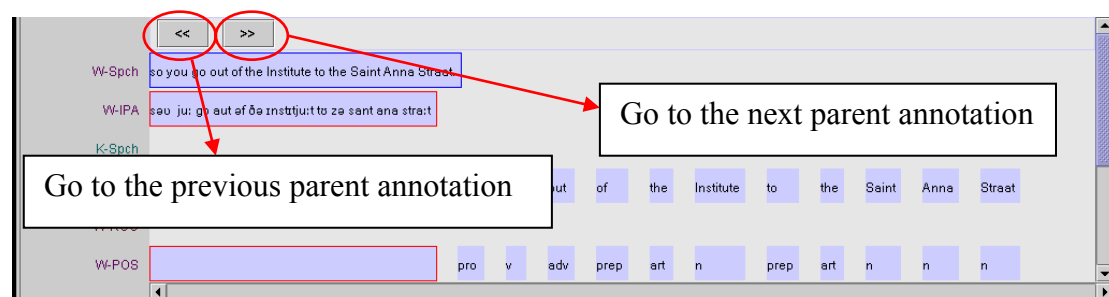




### 3.4.5 Jumping from annotation to annotation (Interlinear Viewer)

In the Interlinear Viewer, you can jump from one annotation block forward/backward to the next block. Do one of the following:

- Click on the left arrow button at the top of the Interlinear Viewer to move to the previous annotation block.
- Click on the right arrow button at the top of the Interlinear Viewer to move to the next annotation block.



### 3.5 How to play a document

ELAN supports the following options for playing a document:

- playing a document (section 3.5.1);
- playing a selection (section 3.5.2);
- playing around the crosshair/selection (section 3.5.3);
- playing in slow motion (section 3.5.4).

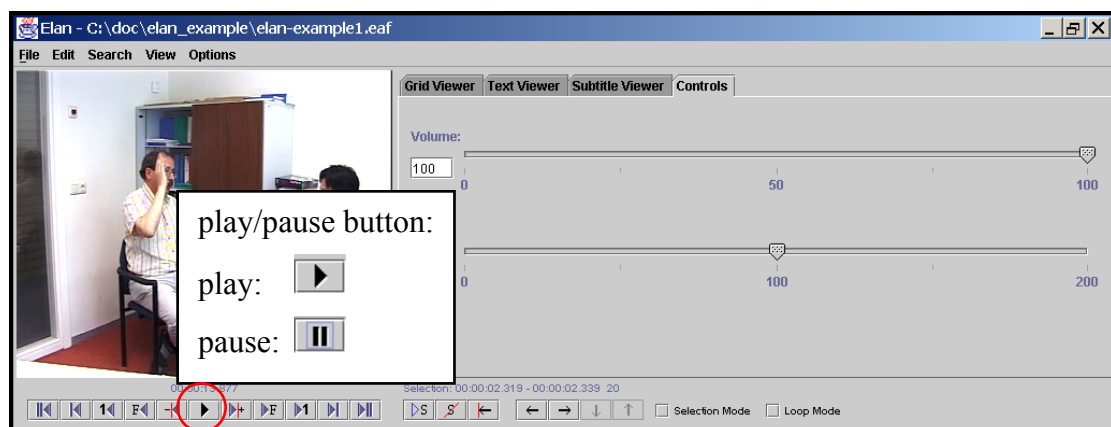
#### 3.5.1 Playing a document

During playback, the red crosshair moves through the Waveform, Timeline, Interlinear, Annotation Density, Text and Grid Viewers, indicating the point in time that is being replayed. If you stop the playback, the crosshair will stop at this point in time. The playback will resume from this point and will continue until the end of the document, or until the playback is paused again.

Make use of either one of the following three options to start/pause the playback:

(a) Use the **Play/Pause button** located under the Video window:

1. Click on the **Play** icon to start playback. After the playback starts, the **Play** icon turns into a **Pause** icon.




2. Click the **Pause** icon to pause the playback again.

(b) Use the **Shortcut key** CTRL+SPACE to start the playback. Use it again to pause the playback.

If two or more ELAN documents are open at the same time, the sound may not work properly. Should this happen, close all documents except for one.

### 3.5.2 Playing a selection

It is possible to play only a selected part of the document. Do the following:

1. Make a selection (see section 0), or click on an annotation.
2. To play the selection, do one of the following:
  - (a) Click on the Play Selection  icon.
  - (b) Or use the shortcut key SHIFT + SPACE.

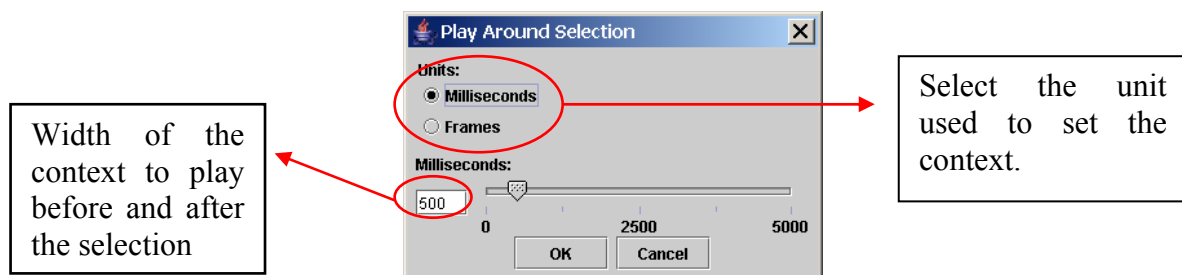
The selected part is played immediately. To listen to the selection again, repeat step 2.

If the crosshair is positioned somewhere within the selection (i.e., if it had been manually moved forward or backward, see section 3.4.3), playback will start from that position and stop at the end of the selection. Otherwise, the whole selection will be played.

### 3.5.3 Playing around a selection

When playing a selection it is possible to extend the playback interval with a context up to a few seconds before and after the selection. This can e.g. be useful to get an idea about the preparation and the finishing of a gesture.

1. Click on Options menu.
2. Go to Play around selection. The following dialog appears:

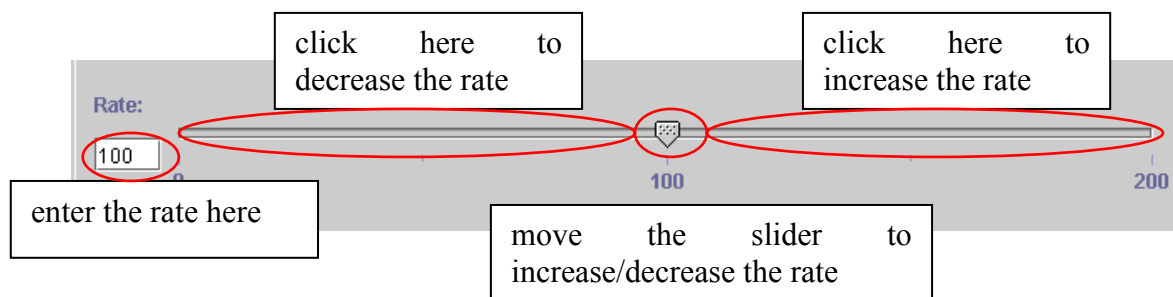


3. Choose how many (milli)seconds or frames there should be played before and after the selection. Click on the OK button.
4. Now make a selection and press CTRL-SHIFT-SPACE. This keyboard shortcut is the only way to play around a selection.

### 3.5.4 Playing in slow motion

To change the playback rate, use the *Rate* slider to the right of the video window. The following options are available:

- Click somewhere above or below the slider to increase or decrease the playback rate by 1%.
- Move the slider up or down to increase or decrease the playback rate.
- Enter the desired playback rate in the box at the left of the slider and press the key ENTER. ELAN accepts rates between 1% and 200%.



Once you have selected a playback rate, the document will be played at the selected rate. If you want it to be played at a different rate, you have to manually change the playback rate, repeating the steps above.

**! Note: On slower machines, the slow motion playback may not work properly.**

### 3.5.5 Changing playback rate and volume via the keyboard

In the **Options > Rate and volume toggle...** menu, a preset value can be specified for the playback rate and volume which can be activated through a keyboard shortcut:

- CTRL+ALT+V : alternate the current volume with the preset value
- CTRL+ALT+R : alternate the current playback rate with the preset value

## 4 Annotations

You can use the ELAN program for annotating your data. This annotation process involves three steps: defining linguistic types and tiers (see sections 4.2.1 and 4.3), selecting time intervals (see section 0), and entering annotations (see section 4.6).

### 4.1 Basic Information: Annotations, tiers and linguistic types

The following illustration shows an example of an annotation document:

The screenshot displays the ELAN software interface with the following components and annotations:

- Video View:** Shows two men in a room. A red box labeled "selected time interval" highlights a portion of the video.
- Subtitle Viewer:** Displays three tiers: "K-Spch" (so from here.), "W-Spch" (so you go out of the Institute to the Saint Anna Straat.), and "W-RGph" (preparation). Red arrows point from the "Real-time displaying of annotations" label to these tiers.
- Timeline:** A horizontal axis with time markers from 00:00 to 00:16. A red box labeled "selected annotation" highlights a segment of the timeline.
- Tiers List:** On the left, a list of tiers includes "W-Spch", "W-IPA", "K-Spch", "W-Words", and "W-RGU". A red box labeled "parent tier:" points to "W-Spch", and another labeled "referring tier:" points to "W-IPA".
- Annotations:** The bottom section shows various linguistic annotations aligned with the timeline, including "so from", "yeah", "ja", "sāu ju:t go aut əf ðə Instɪtju", "Saint Anna", "ana stræt:t", "and you follow that's th", "then you follo", "ay f t t t si kt t t", "thy fo th si kt", and "R Gesture Unit 1". A red box labeled "time-alignable tiers" points to the "W-Words" tier.
- Controls:** A toolbar at the bottom includes buttons for selection and loop mode, with a red box labeled "Loop Mode" pointing to the "Loop Mode" button.

Each annotation is entered on a tier and assigned to a time interval (either directly or to the time interval of another annotation).

All tiers can be displayed simultaneously in the Timeline and Interlinear Viewer, but four of them can be displayed additionally in the Subtitle Viewer. It is useful to select the tier you are currently working on in a Subtitle Viewer because this viewer is bigger and supports line wrapping (which makes it easier to read along during playback).

It is also possible to select one tier as the *active tier*. This can be done by double clicking on the tier name in the Timeline or Interlinear Viewer. When a tier is active, it's name is underlined and displayed in red. Adding a new annotation to a tier by the keyboard shortcut ALT+N is only possible when that tier is active (see section 4.6).

A tier is a set of annotations that share the same characteristics, e.g., one tier containing the orthographic transcription of the speaker's utterances, and another tier containing the free translation of these utterances.

The following two types of tiers exist:

- **Independent tiers**, which contain annotations that are linked directly to a time interval, i.e., they are “time-alignable”.
- **Referring tiers**, which contain annotations that are linked to annotations on another tier (i.e., to annotations on their so-called “parent tier”). They are usually not linked directly to the time axis. (Some of them may be linked – but only within the time interval determined by their independent parent tier, see below.)

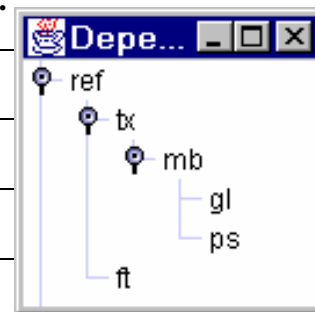
One example: a transcription tier could be independent and time-alignable, as it is linked directly to the time intervals of the speaker's utterances. A translation tier, by contrast, would be referring and not time-alignable: it refers to the transcription tier – not directly to the time axis. By definition, it inherits its time alignment from the transcription tier, i.e., from its parent tier.

In the Timeline and Interlinear Viewers, the label of a referring tier is assigned the same color as the label of its independent parent tier.

It is possible to build up nested hierarchies, i.e., tier **A** can be the parent tier of tier **B**, and tier **B** can be the parent tier of tier **C**, etc.

For example:

tier:	type:	hierarchical relation:
<b>ref</b>	independent	parent of <b>tx</b> and <b>ft</b>
<b>tx</b>	referring	parent of <b>mb</b>
<b>mb</b>	referring	parent of <b>gl</b> and <b>ps</b>
<b>gl, ps</b>	referring	–



Parent and child tiers are linked in such a way that some changes made on a parent tier will also affect its child tiers (but not vice versa):

- If you delete a parent tier, all its child tiers are automatically deleted as well. Similarly, when you delete an annotation on a parent tier, all corresponding annotations on its child tiers are deleted as well.
- If you change the time interval of an annotation on a parent tier, the time interval of the corresponding annotation on all its child tiers are changed accordingly. The time interval of a child tier cannot be changed independently.

You can view the existing dependency relations by clicking on View menu, and then on Tier Dependencies.

Each tier is assigned to a linguistic type. Each linguistic type specifies a number of constraints that hold for all tiers assigned to that type. Such constraints are bundled into so-called ‘stereotypes’. The following four stereotypes are currently available:

- None
 

The annotation on the tier is linked directly to the time axis, i.e., the annotation is entered on an independent tier. Two annotations cannot overlap.
- Time Subdivision:
 

The annotation on the parent tier can be sub-divided into smaller units, which, in turn, can be linked to time intervals. Note that there are no time gaps allowed, i.e., the smaller units have to immediately follow each other.

E.g., an utterance transcribed on a parent tier can be sub-divided into words – each of which is then linked to its corresponding time interval.

Annotations on such tiers are time-alignable. They differ from annotations on independent tiers in that they are assigned to an interval that is contained within the interval of their parent annotation.
- Symbolic Subdivision:
 

Similar to Time Subdivision, except that the smaller units cannot be linked to a time interval.

E.g., a word on a parent tier can be sub-divided into individual morphemes (which are not linked to a time interval).
- Included In<sup>8</sup>:
 

All annotations fall within the borders of the parent tier. However, there can be gaps between the child annotations.

E.g., a sentence with a silence can be split up into words while the silence corresponds to a gap in the child annotations (i.e. the separate words).
- Symbolic Association:
 

The annotation on the parent tier cannot be sub-divided further, i.e., there is a one-to-one correspondence between the parent annotation and its referring annotation.

E.g., one sentence on a parent tier has exactly one free translation. Or one word has exactly one gloss.

<sup>8</sup> A similar stereotype exists in Media Tagger, so it is especially useful for the import of such files.

The following example illustrates the four different stereotypes:



**Time Subdivision:** each word is linked individually to the time axis. If gaps between the annotation items are allowed choose the **Included In** stereotype.

**None:** independent parent tier: transcript is linked directly to the time axis

<b>tx</b> (text)	<i>T'eng nd'yemmnnoe.</i>			
<b>wd</b> (word)	<i>t'eng</i>	<i>nd'yemmnnoe</i>		
<b>mb</b> (morpheme break)	<i>t'eng</i>	<i>n-</i>	<i>d'yem-</i>	<i>nnoe</i>
<b>gl</b> (gloss)	<i>tree</i>	<i>ADVZ</i>	<i>Cl:stand(sg)</i>	<i>DEM.PROX</i>
<b>ft</b> (free translation)	<i>This standing tree.</i>			

**Symbolic Subdivision:** word is divided into morphemes, which are not linked to the time axis

**Symbolic Association:** gloss and free translation have the same number of subdivisions as their corresponding parent tiers

You can define an unlimited number of tiers. It is useful to make decisions about the type of information that you want to enter (and consequently about the type of tiers that you need) at a relatively early stage in the annotation process. However, it is always possible at a later stage to change the parent of a dependent tier (see section 4.3.5) or to copy a tier (section 4.8.2) and to alter the copy.



## 4.2 How to define a linguistic type

ELAN supports the following options:

- adding a new linguistic type (sections 4.2.1 and 4.2.2);
- changing the attributes of a linguistic type (section 4.2.2);
- deleting a linguistic type (section 4.2.3).

### 4.2.1 Adding new linguistic types

Do the following to add a new linguistic type:

1. Click on **Edit** menu.
2. Click on **Add New linguistic type...** The **Add type** dialog window appears.
3. Define the linguistic type (see section 4.2.2).
4. Click **Add** to save the linguistic type. Otherwise click **Close** to exit the window without saving.

### 4.2.2 Creating linguistic types

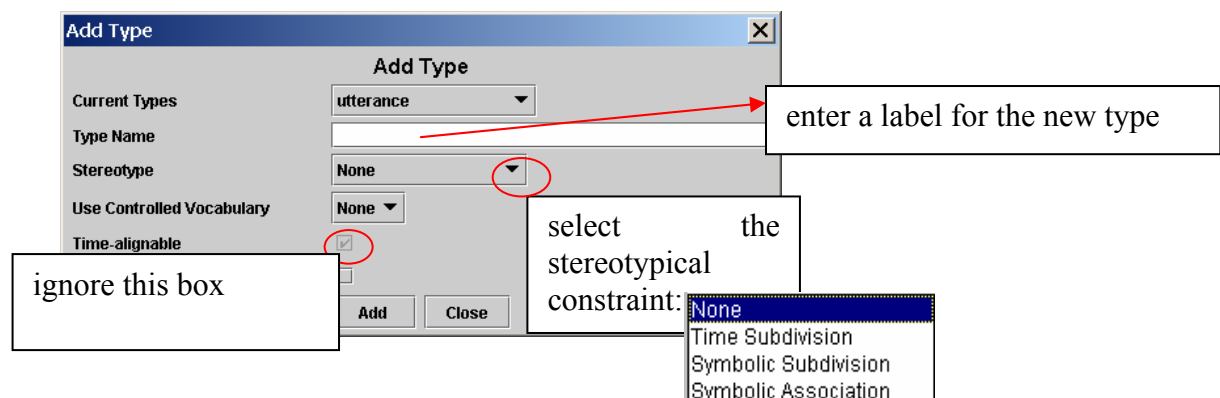
Every tier is assigned to a linguistic type. The linguistic type specifies the stereotypical constraints (and, as a consequence, whether or not the tiers are time-alignable).

Information about the relationship between tiers is given in two different places: for each individual tier it is given in the **Add tier attributes** dialog window (see section **Error! Reference source not found.**), and for all tiers belonging to one linguistic type it is given in the **Add type** window (this section), i.e.:

- **Add tier attributes** window: specify the parent tier of the individual tier.
- **Add type** window: specify the stereotypical constraints of tiers belonging to one type.

To create/change a linguistic type, do the following:

1. Click on the **Edit** menu.
2. Either go to **Add New linguistic type...** or to **Change linguistic type...** In the latter case, click on the linguistic type that you want to change when the **Add/Change type** dialog window appears:



Enter the following information:

- (1) Go to **Type name**. Enter/change the name for the type.
- (2) Go to **Stereotype**. Select the stereotypical constraint relevant to its tiers (see section 4.1).
- (3) Optionally select a controlled vocabulary (see section 4.4).

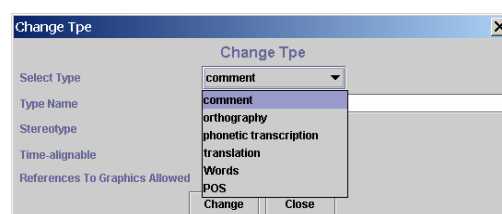
After you have selected a stereotypical constraint, ELAN automatically enters the correct value next to the box **Time-alignable**.

Click **OK** to save the settings; otherwise click **Cancel** to exit the window without saving them.

#### 4.2.3 Changing linguistic types

Do the following to change the attributes of an already existing linguistic type:

1. Click on **Edit > Change linguistic type...** The **Change type** dialog window appears.
2. The labels of all available linguistic types are displayed in the pull down menu, e.g.:
3. Click on the linguistic type whose attributes you want to change.
4. Change the linguistic type (see section 4.2.2).
5. Click **Change** to save the changes. Otherwise click **Close** to exit the window without saving the changes.

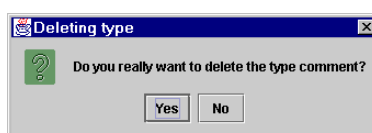


#### 4.2.4 Deleting linguistic types

To delete a linguistic type, do the following:

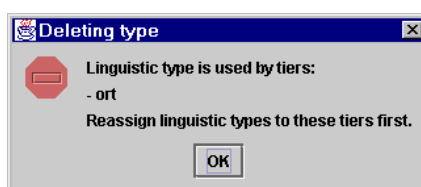
Click on **Edit** menu.

1. Go to **Delete linguistic type...**
2. A dialog window appears. The names of all available linguistic types are displayed in the pull down menu, e.g.:
3. Click on the type you want to delete. A warning appears, e.g.:



4. Click **Yes** to delete the type; otherwise click **No**.

You can only delete a linguistic type if it is not used by any of the tiers. If it is used, the following error message appears:



### 4.3 How to define a tier and its attributes

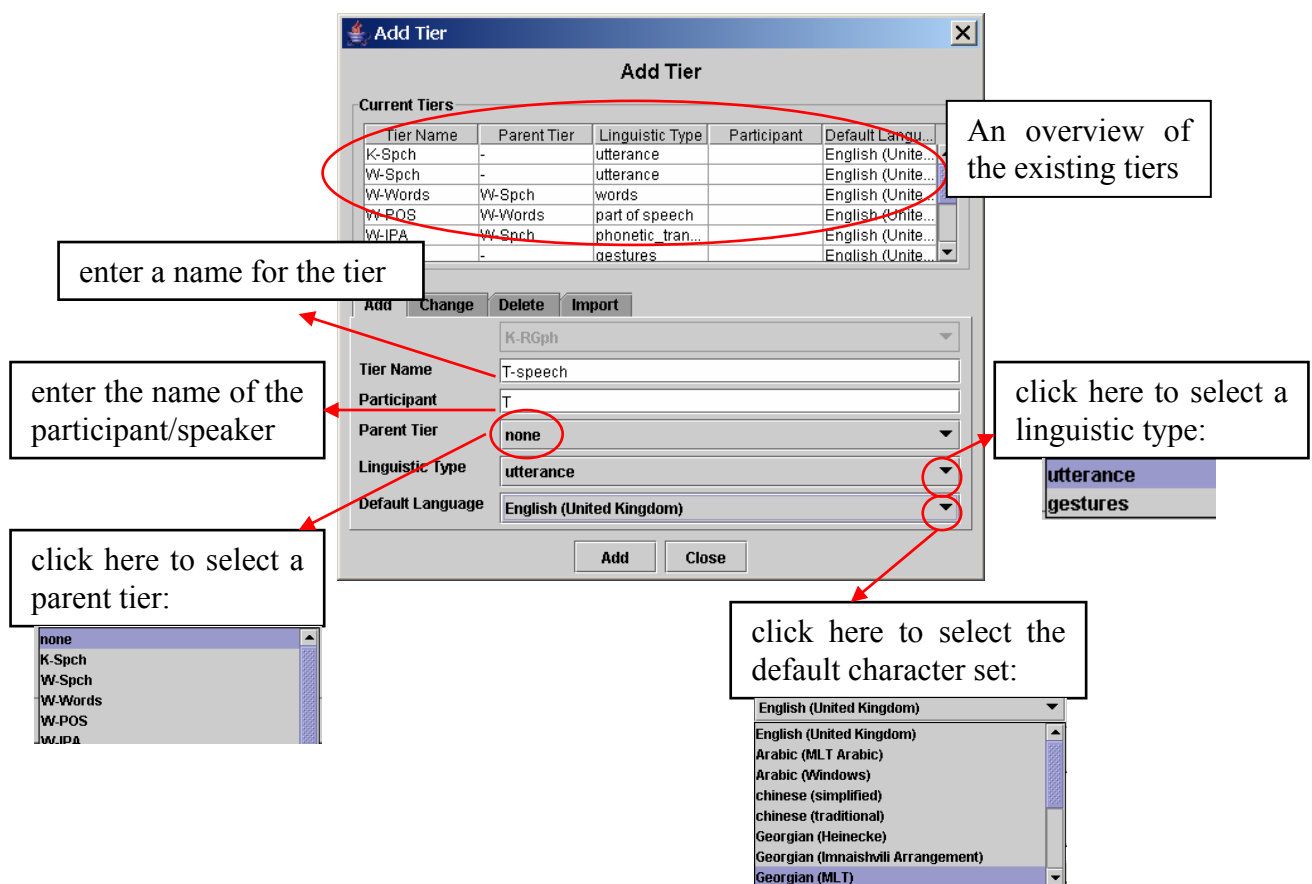
ELAN offers the following options:

- adding a new tier (sections 4.3.1 and **Error! Reference source not found.**);
- changing the tier attributes (sections 4.3.3 and **Error! Reference source not found.**);
- deleting a tier (section 4.3.4).

#### 4.3.1 Adding new tiers

Do the following to add a new tier:

1. Click on **Edit > Add new tier**. The Add tier dialog window appears.
2. Define the tier attributes:



Enter the following information:

- (1) Go to **Tier name**. Enter the tier name.  
The tier name is the name that is displayed in the Timeline, Interlinear and Subtitle Viewer.
- (2) Go to **Participant**. Enter the name of the participant whose utterance is being transcribed.
- (3) Go to **Linguistic type**. Select a linguistic type from the predefined list in the pull-down menu (see section 4.2).

(4) Go to **Parent tier**. Select a parent tier from the pull-down menu.

- If the tier has no parent tier, select **none**.
- Otherwise select the appropriate parent tier.

If the tier is associated with the stereotype “none” (see section 4.2), only the option **none** will be available.

(5) Go to **Default Language**. Select the default character set from the pull-down menu.

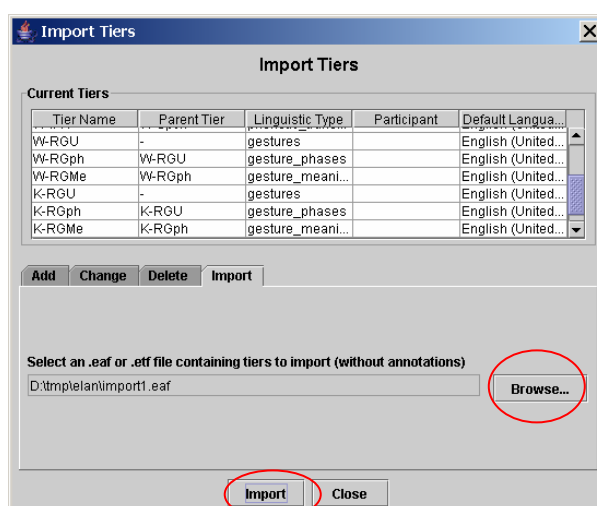
Whenever you enter or change annotations on that tier, the text entry box is automatically preconfigured for the default character set.

Finally click **Add** to save the tier and its attributes. Otherwise click **Cancel** to exit the window without saving.

#### 4.3.2 Importing tiers

You can reuse the tiers (together with their linguistic types) you created before in other eaf or etf (templates, see section 3.2.7) files:

- (1) Select **Edit > Add new tier...**
- (2) Click on the **Import** tab
- (3) Click on **Browse...**
- (4) Select the eaf or etf file from which you want to import a tier and confirm your choice by clicking on **Select**
- (5) Finally click on **Import**.



If you import a linguistic type that already exists, a postfix like **-cp1** or **-cp2** will be added to the imported version.

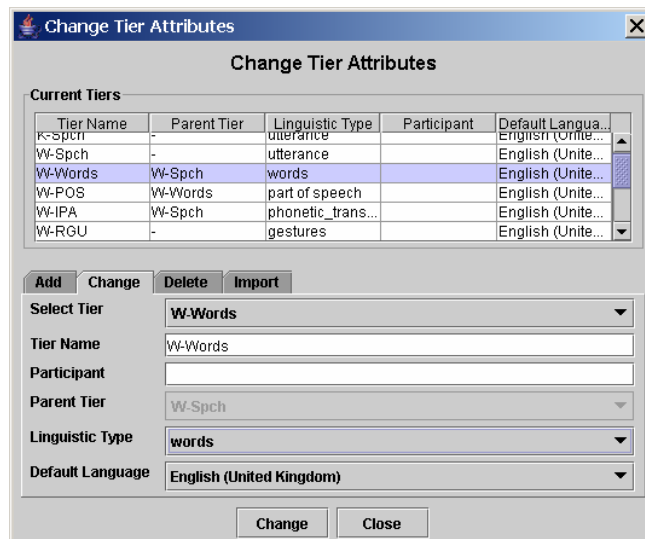
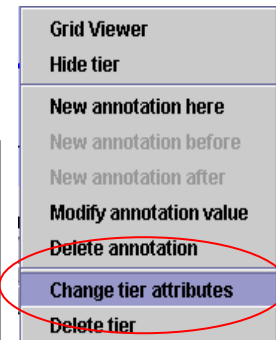
### 4.3.3 Changing tier attributes

Do one of the following to change the attributes of an already existing tier:

#### (a) Use the mouse.

1. In the Timeline Viewer, right-click at about the height of the tier that you want to change. A pull-down menu appears.
2. In the pull-down menu, click on **Change tier attributes**. The **Change tier attributes** dialog window for that tier appears.
3. Change the tier attributes in the dialog that now appears and confirm by clicking on **Change**.

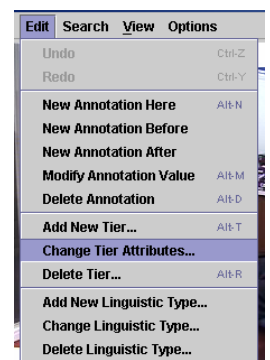
click here  
to change  
the tier  
attributes



You can't change the Parent Tier nor the Linguistic type in this dialog because of possible data loss this alterations would cause. To change these settings in a safe way (i.e. creating a copy of the tier and changing this copy) see section 4.3.3 and 4.3.5).

#### (b) Use the Edit menu.

1. Click on Edit menu.
2. Go to **Change tier attributes...**
3. Click on the drop down box and select the tier which attributes you want to change. The **Change tier attributes** dialog window for that tier appears.
4. After making the changes, click on **Change** to save them. Otherwise click **Cancel** to exit the window without saving.

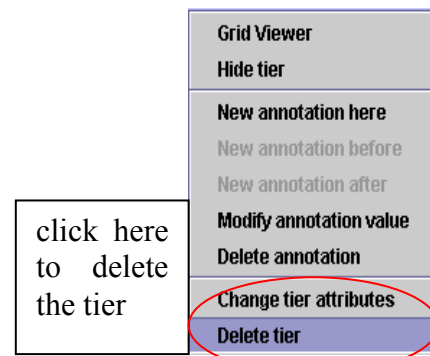


#### 4.3.4 Deleting tiers

Do one of the following to delete a tier:

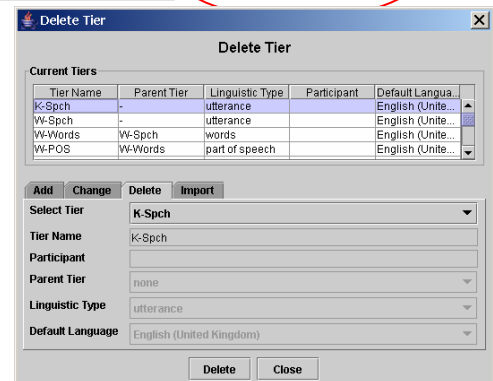
**(a) Use the mouse.**

1. In the Timeline Viewer, right-click at about the height of the tier that you want to delete. A pull-down menu appears.
2. In the pull-down menu, click on Delete tier.
3. A warning dialog appears asking you to confirm the deletion (see below).

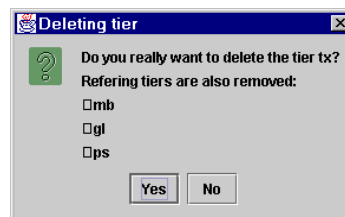


**(b) Use the Edit menu.**

1. Click on Edit menu.
2. Go to Delete tier... A dialog window appears.
3. The labels of all available tiers are displayed a pull down box, e.g.:



4. Click on the tier that you want to delete. A warning dialog appears asking you to confirm the deletion., e.g.:



5. Click Yes to delete the tier and all its child tiers; click NO to not delete them.

If you delete a tier, you will lose all its annotations.

If you delete a parent tier, all its child tiers will be automatically deleted as well. Please make sure that you do not accidentally delete a child tier.

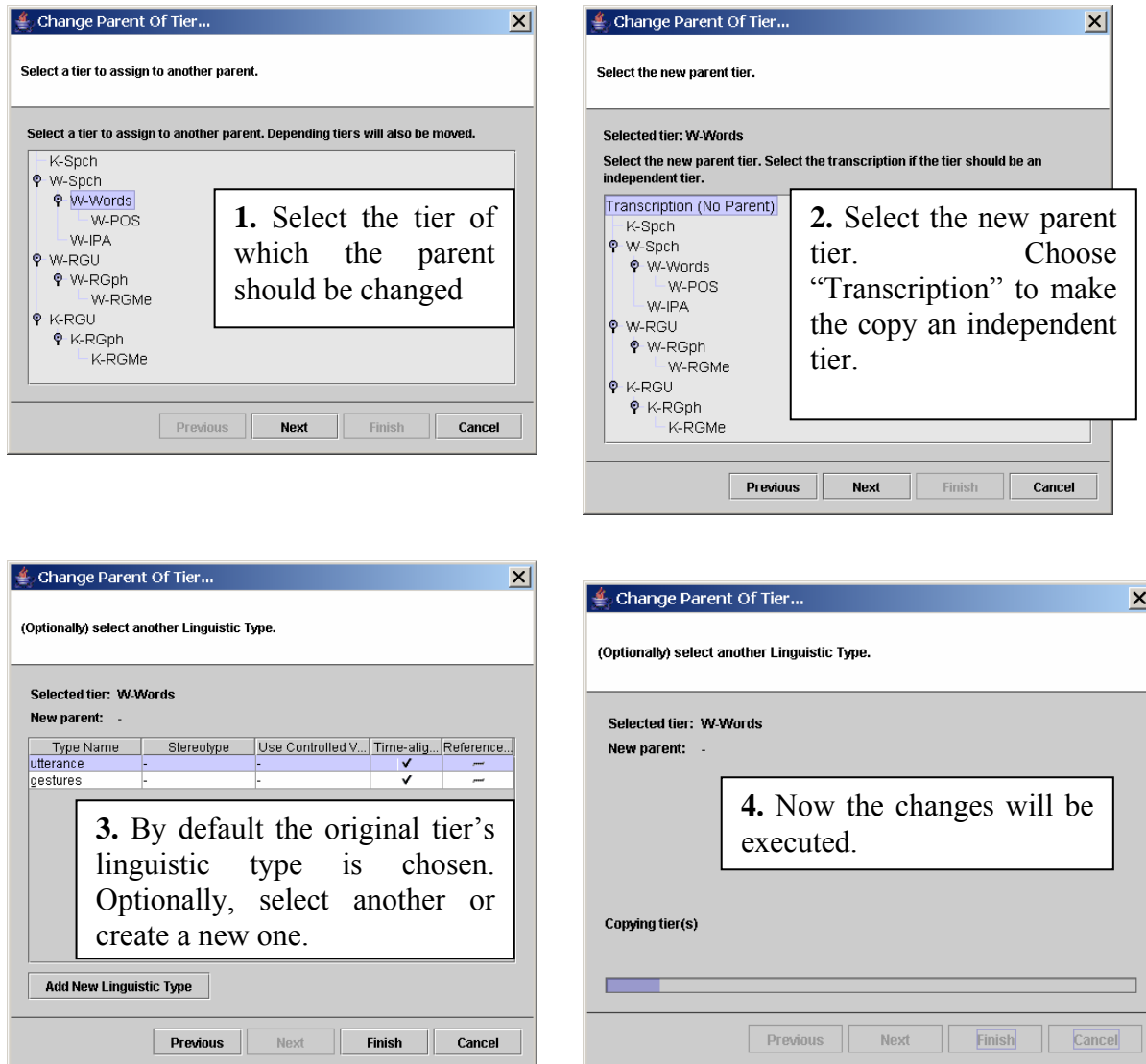
To delete a parent tier without deleting its child tiers, you have to assign the child tier to another parent or make it an independent tier. Afterwards you can safely remove the parent tier. For instructions on how to change a tier's parent, see section 4.3.5.

#### 4.3.5 Changing the parent of a tier

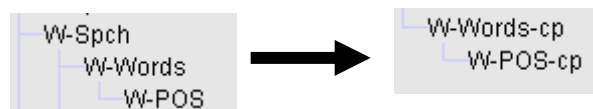
Previous versions of ELAN offered an option to change the parent of a tier. As this operation could cause data loss when not performed carefully, this function has been disabled. However, what can be done now as an alternative is making a copy of a tier and altering that copy. This prevents the possible loss of data, as the original tier stays

the same. At the same time the linguistic type of the copy can be changed as well. In order to maintain the overview during this potentially complicated operation a wizard guides you through this process, which can be started via the Edit > Change parent of tier... menu.

Though an undo option is available it still is a good idea to make a backup of your files before proceeding.



As the result of this process the selected tier (and its children) will be copied and they will become dependent upon the newly chosen parent tier. In our example the W-Words tier, previously a child of W-Spch, became an independent tier:



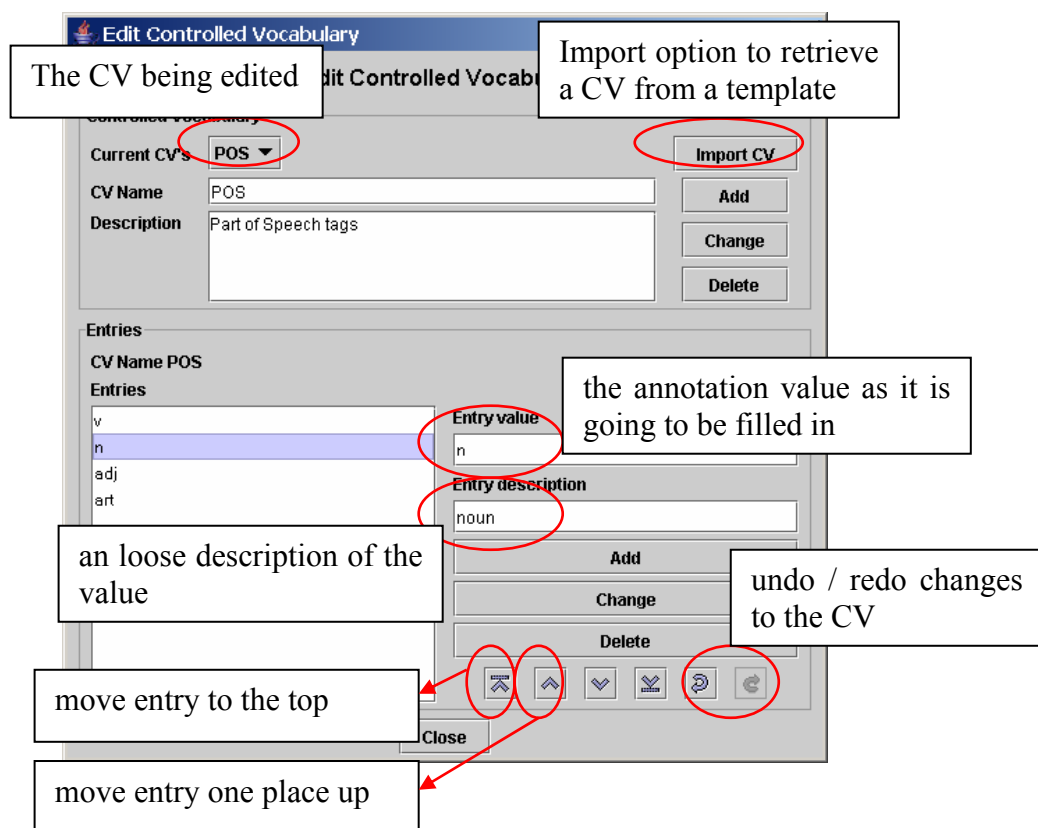
Note that as the tier is not moved but copied the names have been changed: a postfix “-cp” has been added to the copies. The original can be deleted afterwards if you are

satisfied with the result of the operation, while the copies can be renamed to reflect the original tier names.

If you decide to assign a tier to a different parent tier, ELAN will automatically align its annotations with that of the new parent tier (based on overlapping time intervals). In this case, if there is an annotation on the referring tier, but no overlapping annotation on the parent tier, ELAN will delete this annotation. Be very careful that you do not lose such annotations accidentally. A referring tier can be turned into an independent time-alignable tier without any problem.

### The creation of a CV

When you frequently use a certain linguistic type with a limited number of annotation values, it might be a good idea to associate a Controlled Vocabulary (CV) with it. Such a CV consists of a number of predefined values that a user can choose from when editing an annotation, thus making the task of the annotator less error-prone. E.g. one can choose to create a controlled vocabulary for part of speech tagging, as the tags that are used often reoccur. In order to do this select **Edit > Edit Controlled Vocabularies**. The following dialog will appear:



To create a new CV, do the following:

1. Enter a CV Name and a description
2. Click on the **Add** button
3. Now enter each CV entry and its description.
4. Confirm every entry addition by clicking on the **Add** button



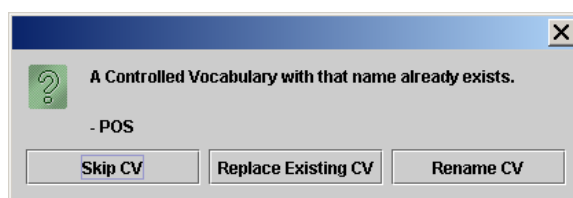
The undo function in the CV dialog window only works as long as that window is active. Once it is closed changes cannot be undone anymore.

### Importing an existing CV

Instead of creating a Controlled Vocabulary again every time ELAN also supports the possibility to reuse existing CVs. A first way to achieve this is to use a template file that contains a CV when creating a new \*.eaf file. (See section 3.2.1) Alternatively you can import a CV for already existing files:

1. select **Edit > Edit Controlled Vocabularies**
2. Click on **Import CV**
3. Select the template from which you want to import a CV
4. Choose **Open**
5. Now all CVs that are stored in the selected template file will be imported

If you try to import a CV with the same name as an already existing CV a dialog will pop up asking what to do:



- **Skip CV:** ignore the CV from the template
- **Replace Existing CV:** overwrite the existing CV with that from the template
- **Rename CV:** opens a dialog asking you to give a new name for the imported CV

### Exporting a CV

A CV can be exported by saving the current file as a template. (See section 3.2.5)

### Using CVs

When editing an annotation that belongs to a linguistic type associated with a CV, a drop down box appears. Just select the value to be filled in:



It is possible to by-pass the controlled vocabulary constraints by holding shift and double clicking on the active annotation (right clicking and selecting “Modify annotation value” while holding shift does the same).

## 4.4 Associating Controlled Vocabularies with linguistic types

### 4.5 How to make a selection

In order to make an annotation, you first have to select a time interval. ELAN supports the following options:

- making and saving a selection on an independent tier (section 4.5.1);
- making and saving a selection on an independent tier while playing (section **Error! Reference source not found.**);
- making and saving a selection on a referring tier (section 4.5.4);
- deselecting a selection (section 4.5.5);
- changing the time alignment of an existing selection (section 4.5.6);
- activating and deactivating the Bulldozer mode (section 4.5.7).

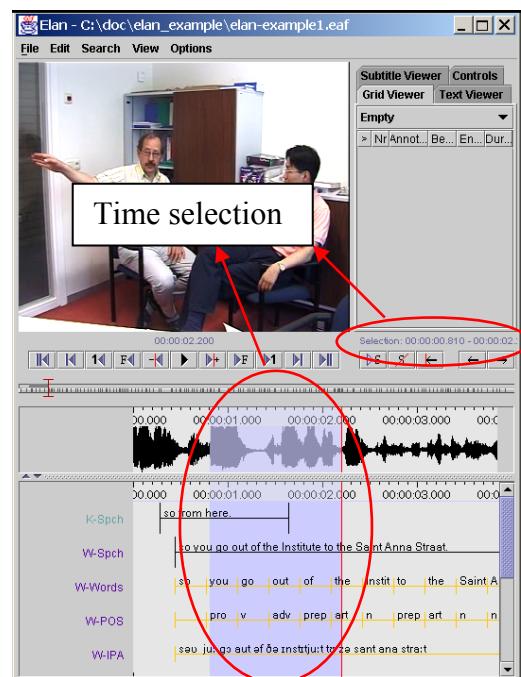
#### 4.5.1 Making a selection on an independent tier

To make a selection, do the following:

1. Go either to the Waveform or the Timeline Viewer.
2. Go with the mouse to the beginning of the time interval you want to select.
3. Click the mouse button, keep it clicked and drag it to the endpoint of the time interval you want to select.

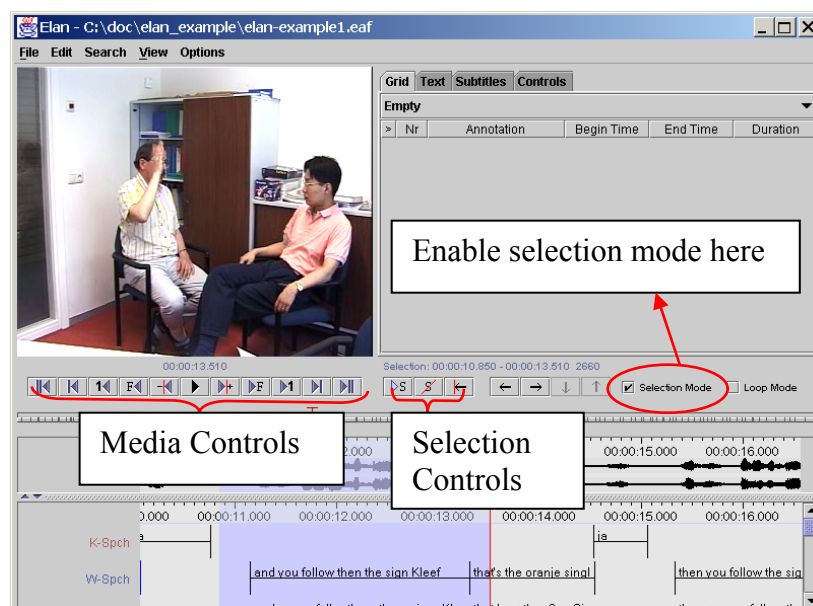
The video image will be continuously updated. The selected part is highlighted in light blue color.

The selection can be extended beyond the size of the current window. The display in all Viewers will automatically move along.



You can change the beginning and endpoints of the selection. Choose one of the following options:

- (a) Either use the mouse: press the key **Shift**, keep it pressed and click with the mouse to the left/right of the selected part. The selection will be extended to include this point.
- (b) Or enable the *Selection Mode* by selecting the **Selection Mode** checkbox. When selection mode is enabled, you can use the media controls to edit the selected part. When moving the crosshair in Selection Mode, the current selection is narrowed or broadened, depending on the direction the crosshair is moved to. For a complete overview of the use of the media controls, see section 3.3.11.



#### 4.5.2 Using the selection controls

The selection controls allow you to navigate through or to change the active selection. For their use, see again section 3.3.11.

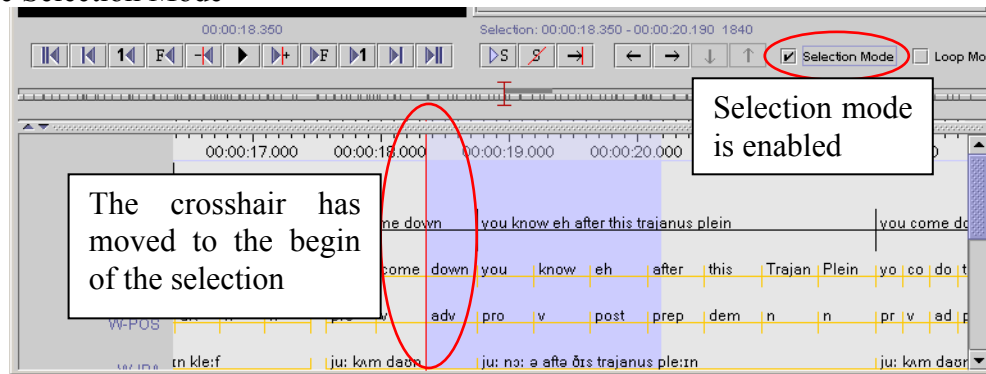
##### **Example: extending the begin of a selection with one second**

Assume that you have made a selection and that you want to add a time interval of 1 second to the begin of it. In that case you should perform the following actions:

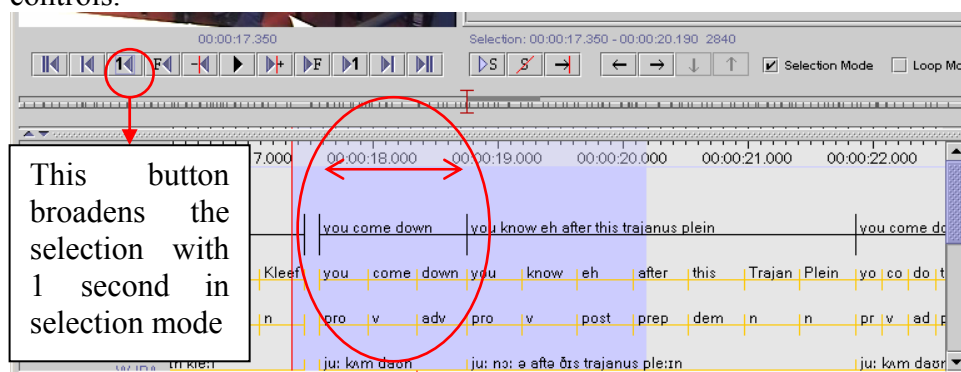
1. Move the crosshair to the begin of the active selection



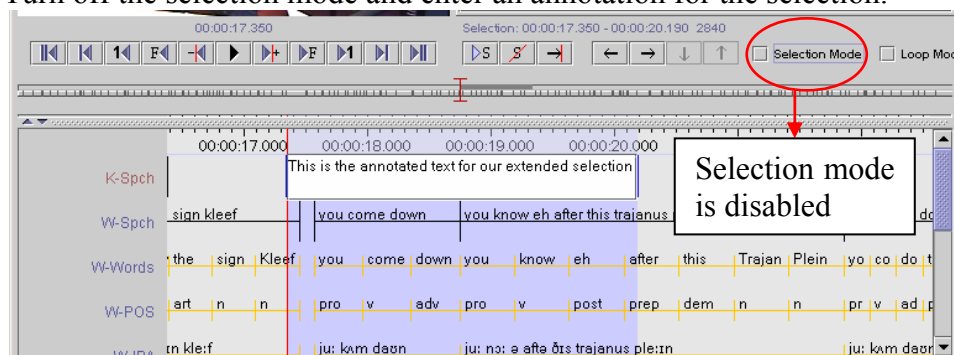
## Enable Selection Mode



2. Go back one second by clicking the corresponding button from the media controls.



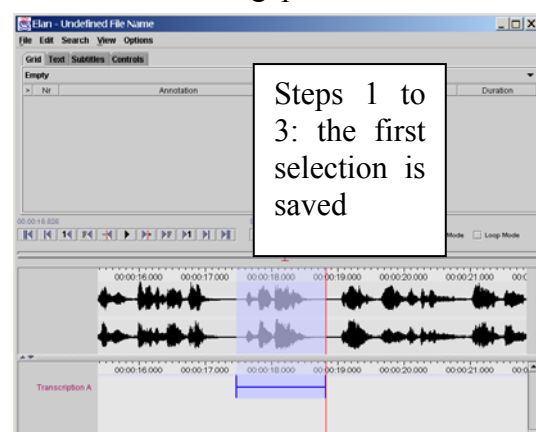
3. Turn off the selection mode and enter an annotation for the selection.



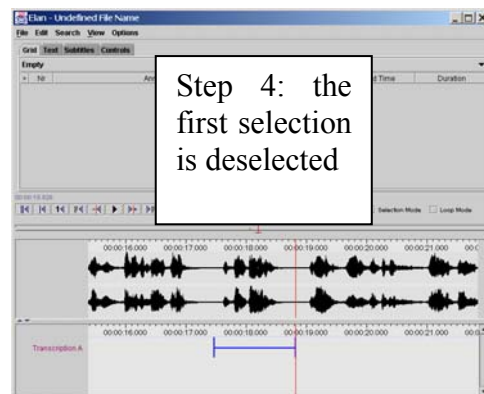
### 4.5.3 Creating annotation units without gaps

It is possible to start a new selection immediately after a previous selection has ended, i.e., the endpoint of one selection will be taken as the starting point for the next selection. Do the following:

1. Make a selection on an independent tier (see section 4.5.1).
2. Double-click on the selection.
3. Optionally enter the content of the annotation unit. Press the keys CTRL+ENTER. The selection is saved.



4. Press the key CTRL-C or click on the clear selection icon to deselect the selection (see section 4.5.5 for deselecting a selection).

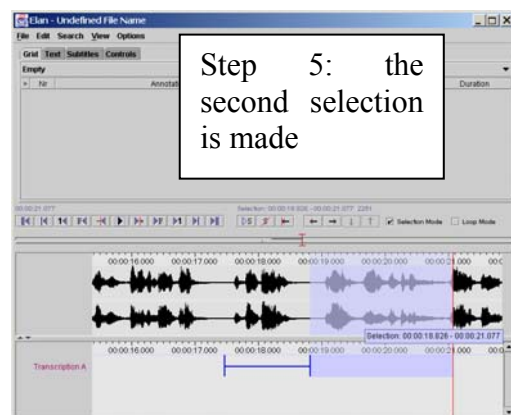


5. Enable the selection mode. Then, play the video or sound file until the

The playback stops. The new selection extends from the endpoint of the previous selection until the point when the playback was stopped.

Repeat steps 2 and 3 to save the new selection.

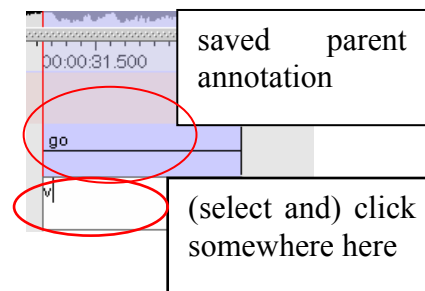
Repeat steps 3 to 5 to add another selection.



#### 4.5.4 Making and saving a selection on a referring tier

A referring tier inherits all its time alignments from its parent tier. To make and save a selection on a referring tier, do the following:

1. Select and save a time interval on the corresponding parent tier (see sections 4.5.1 and **Error! Reference source not found.**).
2. Double-click somewhere within the time interval of the parent annotation at about the height of the referring tier. The Inline Edit box appears.



3. Do one of the following:
  - (a) Enter an annotation (see section 4.6), and then press the keys CTRL+ENTER to save the selection.
  - (b) Press the keys CTRL+ENTER (without entering an annotation) to save the selection.

#### 4.5.5 Deselecting a selection

To deselect a selection, do one of the following:

- (a) Either use the Deselection icon from the selection controls:



- (b) Or use the shortcut key CTRL-C.

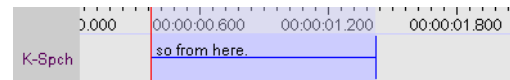
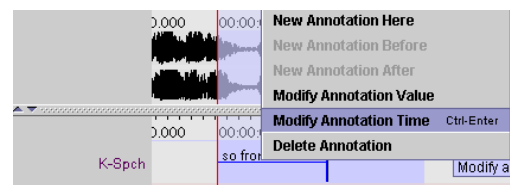
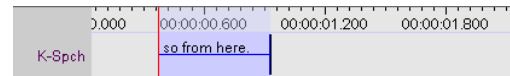
**! Note:** Whenever you select another time interval, the old selection is automatically deselected, unless you enabled Selection Mode.

#### 4.5.6 Changing the boundaries of an existing selection

##### By selection

In the timeline viewer, go through these steps:

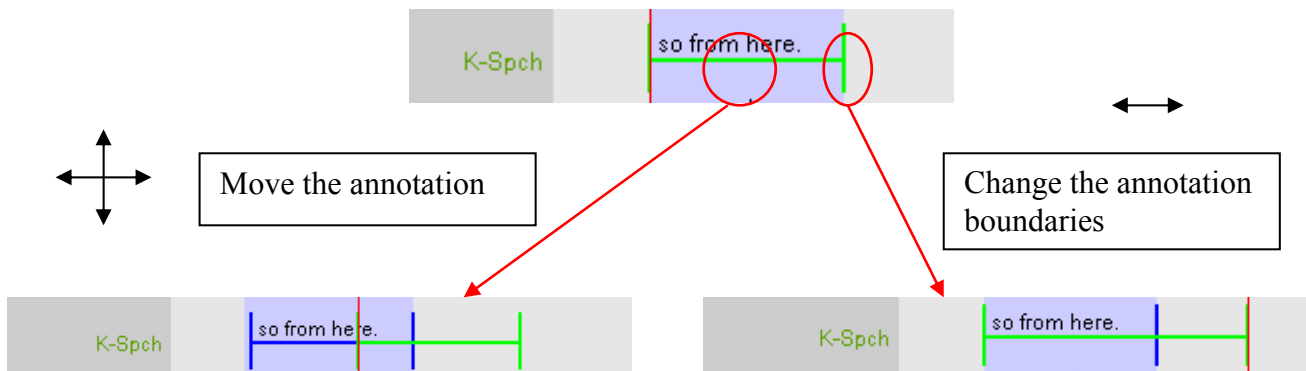
1. Click on an annotation unit to select it.
2. Select the region where you want the modified annotation to be placed.
3. Right click on the original annotation and select **Modify annotation time** or press CTRL+ENTER
4. Now the length of the annotation becomes that of the selection from the second step.



##### By dragging the mouse

If you press ALT, the active annotation is indicated in green and becomes adjustable with the mouse:

- drag in the middle of the annotation and drop it somewhere else to move it
- drag and drop the borders to change the boundaries of the annotation unit



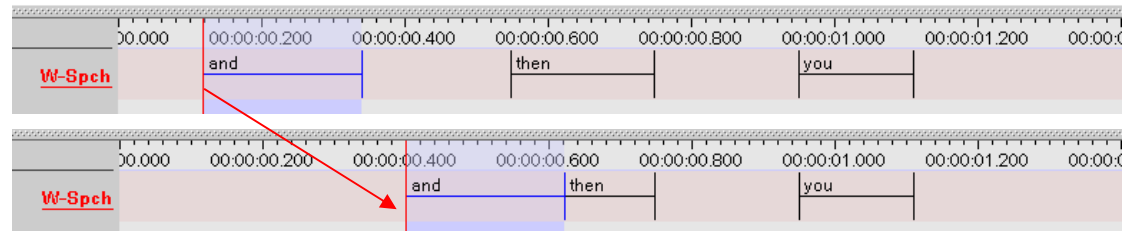
Only the time-alignment of annotations on the following types of tiers can be modified: annotations on independent tiers, and annotations on referring tiers that fall under the Time Subdivision stereotype (but note that in the latter case, the alignment cannot be extended beyond the boundaries of its parent annotation, see section 4.1). To modify the time alignment of annotations on all other tiers, change the time alignment on the corresponding parent tier (following the steps above). The time alignment on all referring tiers is automatically updated.



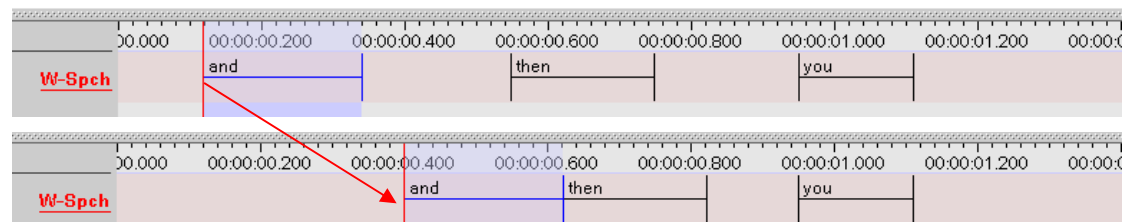
#### 4.5.7 Activating and deactivating the Bulldozer mode or Shift mode

ELAN supports three editing modes: Overwrite mode, Bulldozer mode and Shift mode. These two modes are somehow comparable to the Overwrite and Insert modes that are supported by many text editor programs such as, e.g., Microsoft Word.

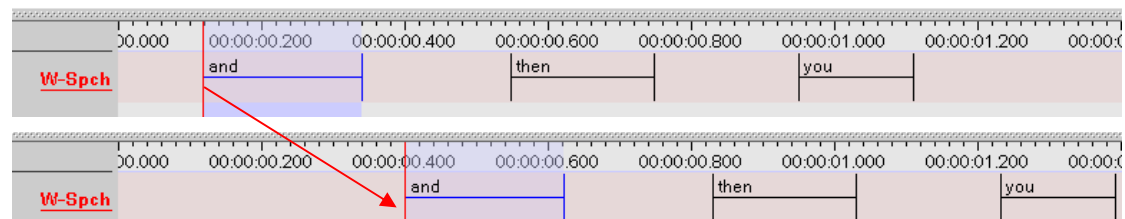
- **Normal (i.e. overwrite) mode:** if you extend a selection into a time interval that is already occupied by an annotation, that annotation is (partly or wholly) overwritten.



- **Bulldozer mode:** if you extend a selection into a time interval that is already occupied by an annotation, that annotation is moved to the right/left. Think about it as a bulldozer which pulls all annotations together, discarding the spaces in between.



- **Shift Mode:** like Bulldozer Mode, but the spaces between annotations are preserved too. This resembles most to the insert mode of text editors.



The following conventions apply for the Bulldozer mode:

- Annotations are moved to the right if you extend your selection from left to right. They are moved to the left if you extend your selection from right to left.
- If a moved annotation extends into the time-interval of yet another annotation, that other annotation is moved accordingly. If it extends into empty space, no other annotations are affected.

Moving annotations may thus affect the whole document, and may thereby destroy previous time alignments. Please make sure that the Bulldozer Mode is not accidentally switched on.

The Overwrite mode is the default mode. To switch to another mode, do the following:

1. Click on Options menu.
2. Click on Normal Mode, Bulldozer Mode or Shift Mode. A checkmark appears next to it. To switch back to the Overwrite mode, repeat steps 1 and 2 above. The checkmark disappears.

## **4.6 How to enter annotations**

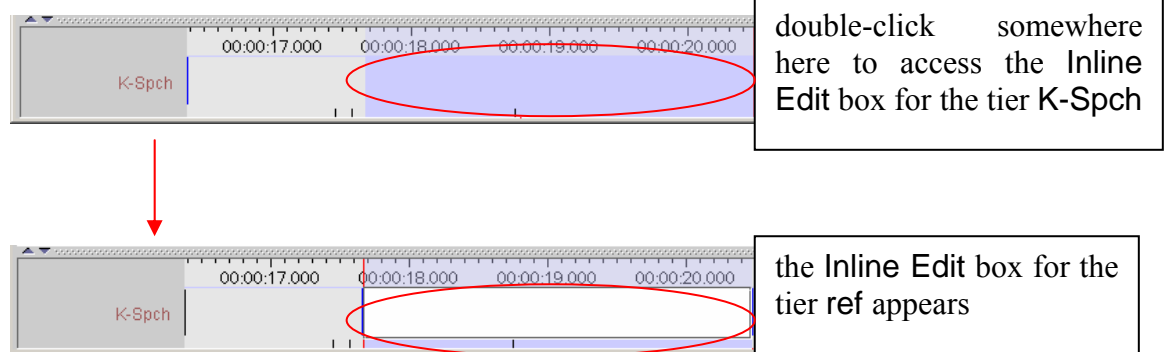
All annotations are entered either into the **Inline Edit** box or into the **Edit Annotation** box, which are accessed via the **Timeline**, **Interlinear**, **Grid**, **Text** and **Subtitle Viewers**. ELAN supports the following options:

- entering annotations (sections 0 and 4.6.2);
- entering annotations before/after other annotations (section 4.6.3);
- modifying the content of annotations (section 0);
- deleting annotations (section 4.6.6);
- entering annotations in different character sets (section 4.6.7).

#### 4.6.1 Entering annotations (into the Inline Edit box)

1. Select the time span in which you want to place the annotation.
2. Access the Inline Edit box by doing one of the following:
  - (a) Either double-click in the Timeline Viewer on the selection at about the height of the tier where you want to enter the annotation.
  - (b) Or click on Edit menu, then click on New annotation here (active tier only).
  - (c) Or use the key equivalent ALT+N (active tier only).

In all cases, the Inline Edit box appears, e.g.:

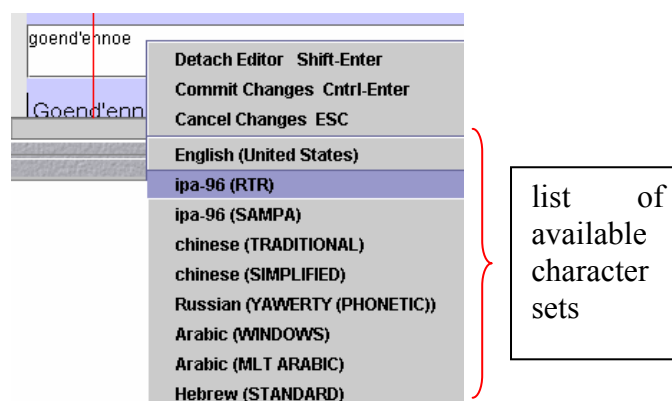


3. Do one of the following:
  - (a) Press the keys CTRL+ENTER (without entering an annotation) to create an empty annotation.
  - (b) Enter an annotation and then press the keys CTRL+ENTER to save the selection.

It is possible to enter text that contains line breaks. The text entry box automatically displays a scrollbar if necessary.

The Inline Edit box is automatically preconfigured for the default character set of the tier (see section **Error! Reference source not found.**). If you want to use a different character set, do the following:

1. Right-click in the Inline Edit box. A pull-down menu appears that displays the available character sets.



2. Click on the appropriate character set. From now on, the characters are entered in the selected set.
3. To switch back to the default character set, repeat the steps above and select the default set from the pull-down menu.

Only selections on time-alignable tiers can be saved in this way. To save a selection on a referring tier, see section 4.5.4.

3. Save the annotation by doing one of the following:
  - (a) Use the shortcut keys CTRL+ENTER.
  - (b) Or right-click in the **Inline Edit** box and click on **Commit Changes** in the pull-down menu.

To exit the **Inline Edit** box without saving, do one of the following:

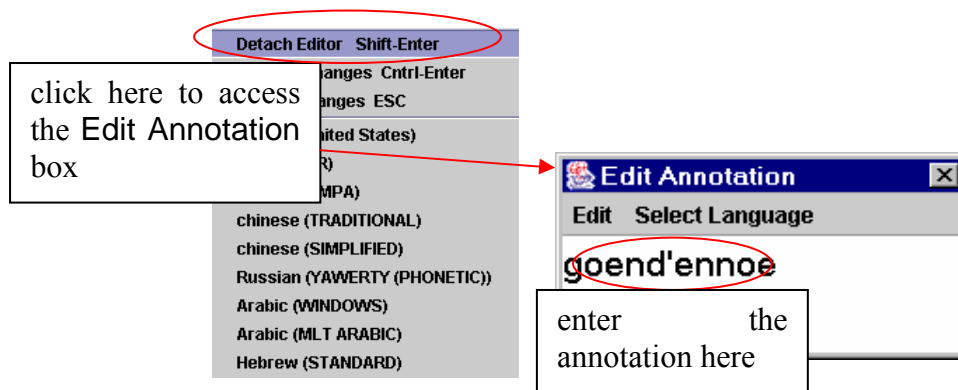
- (a) Use the shortcut key ESC.
- (b) Or right-click in the **Inline Edit** box and click on **Cancel Changes** in the pull-down menu.

#### 4.6.2 Entering annotations (into the Edit Annotation box)

The Edit Annotation box differs from the Inline Edit box in that it has a bigger, resizable, display and supports line wrapping. It is therefore better for entering longer texts.

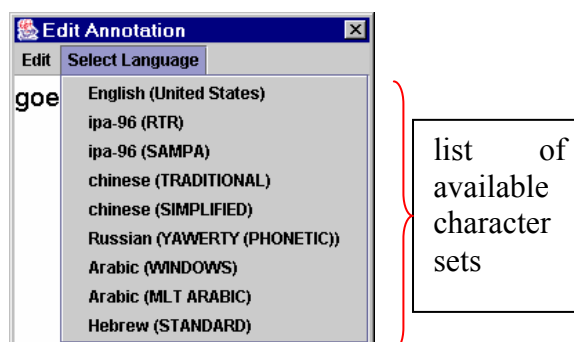
To enter an annotation into the Edit Annotation box, do the following:

1. Either make a selection in the Timeline Viewer (see section 4.5.1), or click on an existing annotation in the Timeline or Interlinear Viewer.
2. Access the Inline Edit box (see section 0).
3. Do one of the following:
  - (a) Either use the shortcut keys SHIFT+ENTER.
  - (b) Or right-click in the Inline Edit box. A pull-down menu appears. Click on Detach Editor.
4. The Edit Annotation box appears:



The Edit Annotation box is automatically preconfigured for the default character set of the tier (see section **Error! Reference source not found.**). If you want to use a different character set, do the following:

1. Click on Select Language. A pull-down menu appears that displays the available character sets, e.g.:



2. Click on the appropriate character set. From now on, the characters are entered in the selected set. (For an overview of the input methods for the character sets see section 4.6.7 ).
3. To switch back to the default character set, repeat the steps above and select the default set from the pull-down menu.
5. Save the annotation by doing one of the following:
  - (a) Use the shortcut keys CTRL+ENTER.
  - (b) In the **Edit Annotation** box, click on **Edit** and then click on **Commit Changes** in the pull-down menu.

To exit the **Edit Annotation** box without saving, do one of the following:

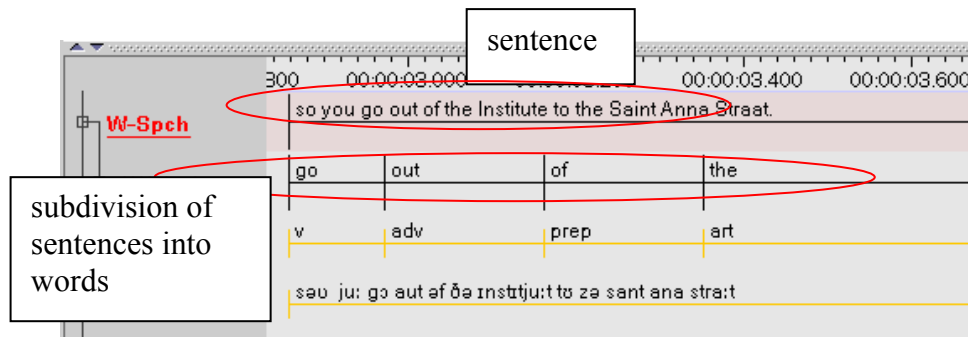
- (a) Use the shortcut key ESC.
- (b) In the **Edit Annotation** box, click on **Edit** and then click on **Cancel Changes** in the pull-down menu.

To return to the **Inline Edit** box, do one of the following:

- (a) Use the shortcut keys SHIFT+ENTER.
- (b) In the **Edit Annotation** box, click on **Attach Editor** in the pull-down menu.

#### 4.6.3 Entering annotations before/after other annotations

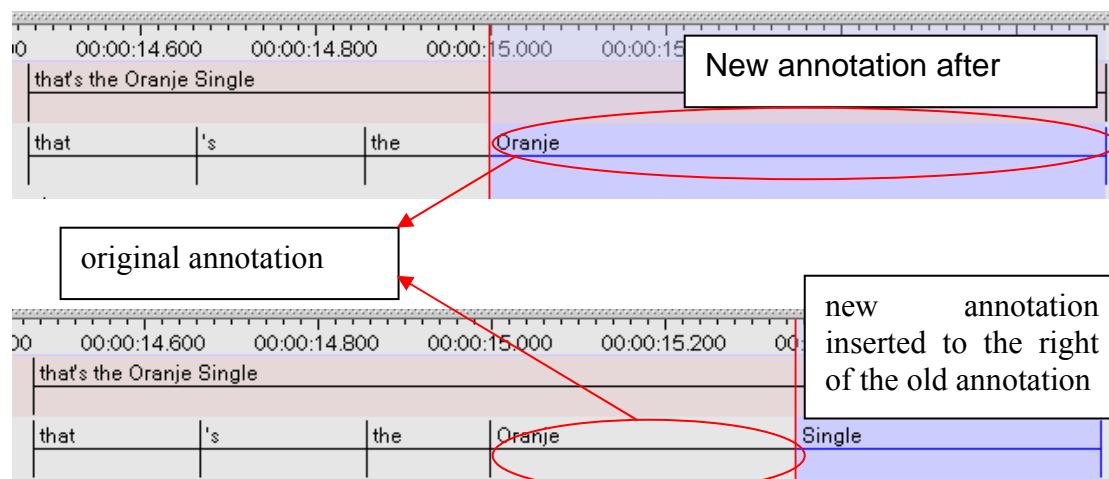
Annotations on some tiers can be subdivided into smaller units. In this way, you can for example break up a sentence into different words (or words into morphemes, etc.), as in the following illustration:



To divide an annotation into smaller units, do the following:

1. In the Timeline or Interlinear Viewer, click on the annotation that you want to subdivide. It appears in a dark blue frame.
2. Do one of the following:
  - (a) Right-click on the selected annotation. A pull-down menu appears. Click on either **New Annotation before** or on **New Annotation after** to subdivide the annotation.
  - (b) Or click on **Edit** menu. Then click on either **New Annotation before** or on **New Annotation after** to subdivide the annotation.

If you click on **New annotation before**, the original annotation is divided and the new annotation is inserted to its left (as in the illustration below). If you click on **New annotation after**, it is inserted to its right.





This option is only available for those tiers that are assigned to the stereotypes Time Subdivision and Symbolic Subdivision (see section 4.1).

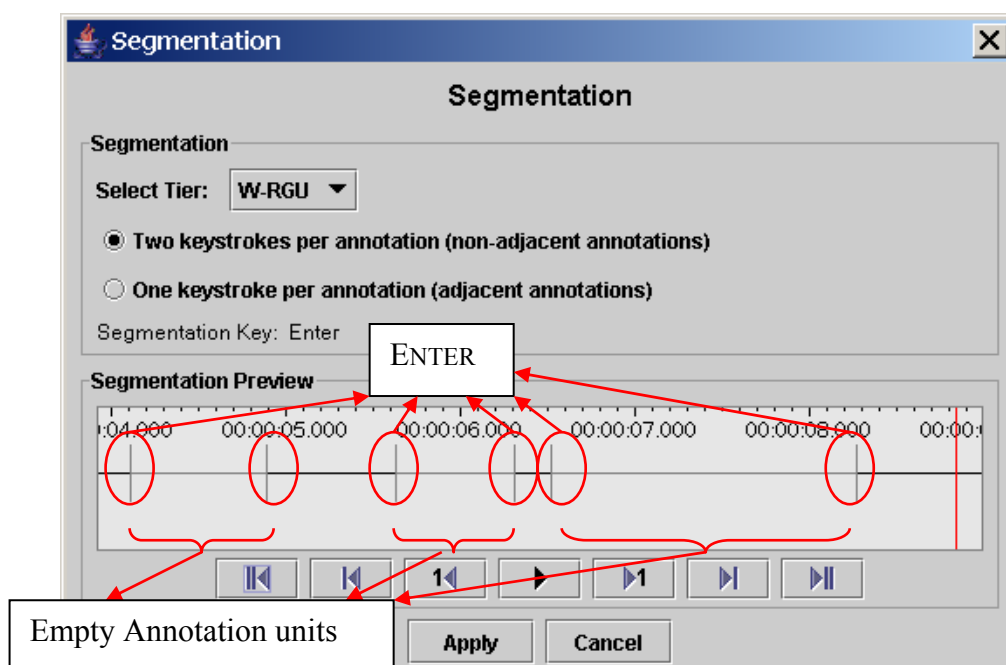
An annotation is always subdivided into two units. If you need further subdivisions, repeat the steps above.

#### 4.6.4 On-the-fly segmentation

Creating annotation units can be a time consuming job. Therefore, ELAN offers a faster way to initially mark the boundaries of annotation units on a root tier (i.e. where the linguistic type stereotype is “None”): on-the-fly segmentation. This way of working is similar to Transcriber<sup>9</sup>’s segmentation mode. This mode can be accessed through **Edit > Segmentation ...** and works as follows:

1. Select the tier to be segmented
2. Choose between 2 keystrokes/annotation (non-adjacent annotations) and 1 keystroke/annotation (this makes all annotations adjacent)
3. Use the media controls to play the media file.
4. Press enter to start/end an annotation unit, while the movie or sound file is playing
5. Once finished, choose **Apply** to save the newly created annotations or **Cancel** to discard the changes.

The segmentation is considered as a single operation and thus can be reversed by a single undo command.



<sup>9</sup> See <http://www.etca.fr/CTA/gip/Projets/Transcriber/>

#### 4.6.5 Modifying the content of annotations

To modify an annotation, do one of the following:

**(a) In the Timeline or Interlinear Viewer:**

1. Click on the annotation that you want to modify. It appears in a dark blue frame.
2. Do one of the following:
  - (a) Right-click on the selected annotation. A pull-down menu appears. Click on **Modify annotation value**.
  - (b) Or click on **Edit** menu. Click on **Modify annotation value**.
  - (c) Or use the key equivalent ALT+M.

The Inline Edit box appears (see section 0).

**(b) In the Grid Viewer:**

1. Double-click on the annotation that you want to modify.

The Inline Edit box appears (see section 0).

#### 4.6.6 Deleting annotations

To delete an annotation, do the following:

1. In the Timeline or Interlinear Viewer, click on the annotation that you want to delete. It appears in a dark blue frame.
2. To delete it, do one of the following:
  - (a) Right-click on the selected annotation. A pull-down menu appears. Click on **Delete annotation**.
  - (b) Or click on **Edit** menu. Then click on **Delete annotation**.
  - (c) Or use the key equivalent ALT+D.

The annotation is deleted without further warning.

If you delete an annotation on a parent tier, the corresponding annotations on all its child tiers will be automatically deleted as well. Please make sure that you do not accidentally delete a child annotation. An annotation on a child tier can be deleted without consequences for the annotation on its parent tier.

#### 4.6.7 Entering annotations in different character sets

ELAN supports different character sets, making use of the following input methods:

### (a) GATE Unicode Kit (GUK)

Characters can be entered by using a different keyboard mapping. This method is implemented using the GATE Unicode Kit developed at Sheffield University, Department of Computer Science.

If you select a character set that is based on GUK, a visual representation of a keyboard appears on the screen, which informs you about the implemented keyboard mapping. The following illustrations show the mappings of “ipa-96 (SAMPA)” and “Arabic (WINDOWS)”:



To enter a character, do one the following:

- Either press the corresponding key on your keyboard.
- Or click on the corresponding key of the keyboard displayed on your screen.

The visual representation has the layout of the standard UK keyboard. If you do not have a UK keyboard, there may be mismatches between the characters and their visual representation.

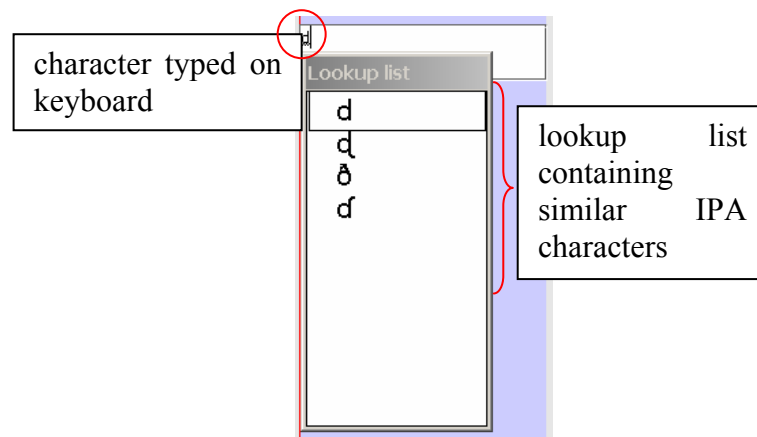
For example, the IPA character “ə” is matched to the key “@”, i.e., in order to get “ə”, you have to type “@”. On a standard UK keyboard, the key “@” is located to the left of the key “enter” (see the illustration above). On other keyboards, however, “@” may be located on a different key. In such cases, if you press the key to the left of “enter”, you will not get the character “ə”. To get “ə”, you have to search for the

location of “@” on your keyboard, and then press that key. (Note that these mismatches only arise if you use the physical keyboard, but not if you use the visual representation on the screen.)

The character set “ipa-96 (SAMPA)” can be used to enter IPA characters. However, the current version of ELAN only supports SAMPA, but not X-SAMPA. As a consequence, some of the characters that you require may not be available yet (see <http://www.phon.ucl.ac.uk/home/sampa> for further information).

## (b) Roman Typographic Root (RTR)

IPA characters can be entered using the RTR input method. This method is based on the following principle: whenever you type a character, all typographically similar characters are displayed in a lookup list, as shown in the following illustration:



To select a character from the lookup list, do the following:

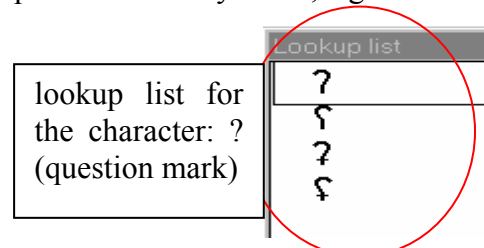
1. Use the UP and DOWN arrow keys to navigate to the desired character.
2. Press ENTER or SPACE to confirm the selection.

Do not use the mouse within the lookup window. If you do, the window will disappear.

The input of IPA characters is restricted to the official IPA-96 character set. Withdrawn or superseded characters are not supported.

The RTR-mapping is generally obvious, but please pay attention to the following features:

- Use lower case letters only.
- IPA characters can map to non-letter symbols, e.g.:

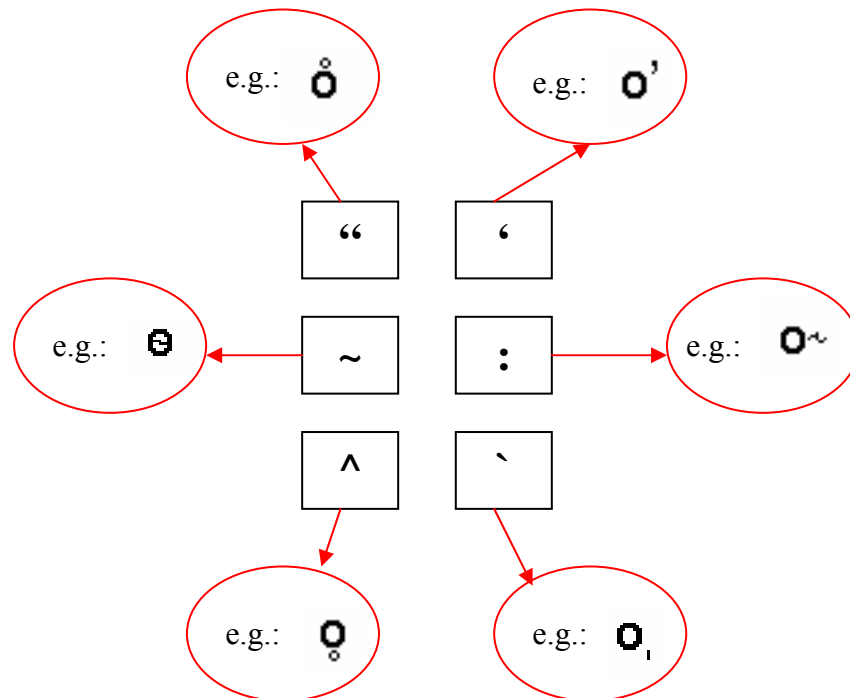


- Diacritics are entered with the help of the following keys:

key      position of diacritic in relation to character

“      (quotation mark)      above  
 ‘      (apostrophe)      in the upper right  
 :      (colon) to the right  
 `      (grave accent) in the lower right  
 ^      (circumflex accent)      under  
 ~      (tilde) on

For example:

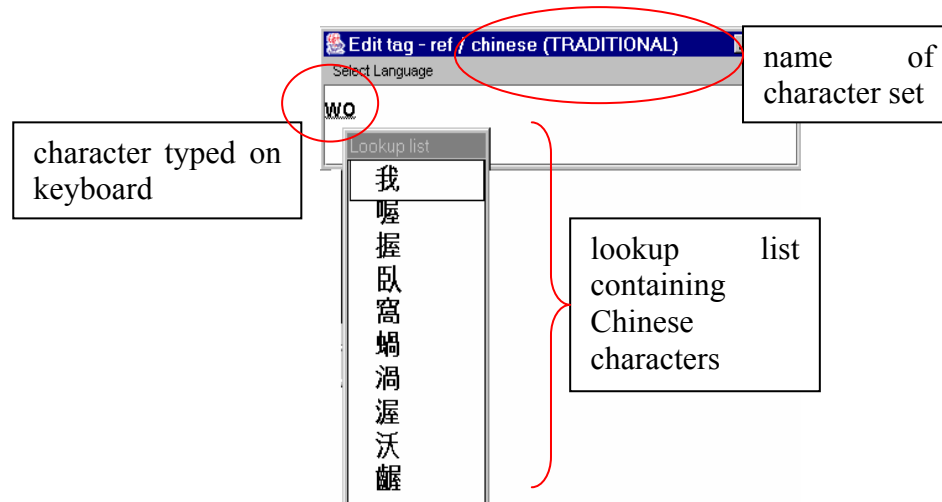


On Windows 2000, if you use an international keyboard, you have to type SPACE after typing the quotation mark (“) or one of the accents (‘, `, ^).

It is possible to stack diacritics, but there may be problems displaying them.

### (c) Chinese characters

Chinese characters, both traditional and simplified, are entered using the Pinyin method. Characters are selected by starting to type Roman characters. Candidates are shown in a lookup window while the user types along. The desired character is selected with the UP and DOWN arrow keys, e.g.:



Use the following procedure:

- Enter the pinyin word with the keyboard. For each pinyin word, a list of Han symbols is shown in a popup window.
- Navigate to a Han symbol with the UP and DOWN arrow keys.
- Page through the list with the PAGE UP and PAGE DOWN keys.
- Select a Han symbol with the SPACE or the ENTER key.

Do not use the mouse within the lookup window. If you do, the window will disappear.

On Windows 98, you cannot display both Chinese and IPA characters.

### (d) Keyboard tools

If the options above don't fulfill your needs (e.g. the character set is not supported or you don't want to use the on-screen display for a large amount of annotations) you might want to look for a third-party solution. Such a tool provides a mean to remap your keyboard to the desired input character set. For details, we refer to the following programs:

Keyman (Windows), <http://www.tavultesoft.com/>

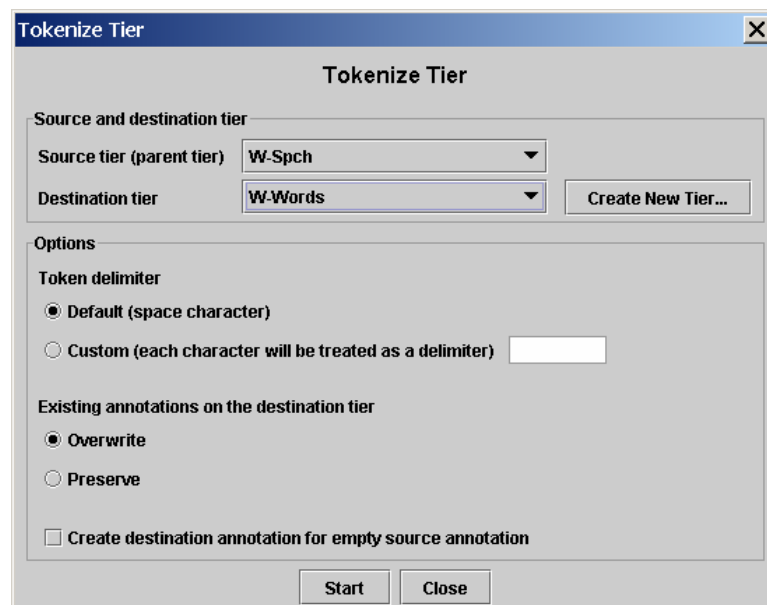
Ukelele (Mac), <http://scripts.sil.org/ukelele>



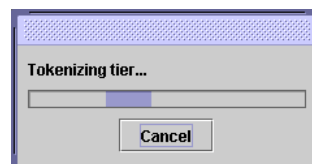
## 4.7 How to create annotation units on dependent tiers

One of the main advantages of using a hierarchical structured tiers is the possibility to split the content of an annotation unit on a parent tier automatically into smaller parts on a child tier. E.g. the words on an utterance tier could be split into separate words. This is called tokenizing in Elan. These steps will guide you through this process:

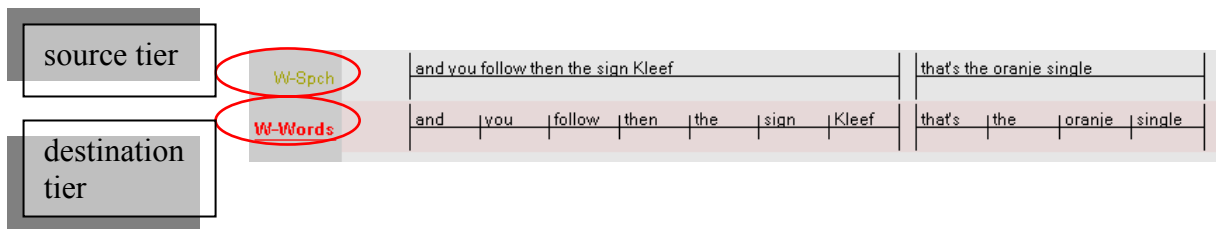
1. Go to **Edit > Tokenize tier**
2. Select a Source (= parent) and Destination (= child) tier
3. Optionally create a new destination tier by selecting **Create New Tier...**
4. Select a delimiter. The default is a space, but other choices are possible (e.g. “-” for morpheme breaks).
5. If the destination tier already contains annotation units, choose between overwriting or preserving them. If it’s still empty you can ignore this option.
6. Select **Create destination annotation for empty source annotation** if you want to create for every source annotation a destination annotation, even if it’s empty.



7. Click on **Start**, the tokenizing will begin:



8. When it is finished, you will see that every annotation unit from the source tier has been tokenized on the destination tier:



Even when tokenizing to a tier from the type time subdivision, all tokens (words in this example) on the destination tier have the same size (i.e. duration). In that case you probably want to adjust their length, as described in section **Error! Reference source not found.**

For symbolic associations, there is no need to use the tokenizer. Instead, go to the grid viewer and make sure the checkbox next to the dropdown menu is selected. Now you can fill in the annotations of the symbolic associations in their column of the grid. See also section 3.3.6. If you want to copy or filter the contents from 1 tier to another symbolic associated tier, have a look at the next section.

Tabs and newline characters are always treated as delimiters.

## 4.8 How to filter and copy tiers

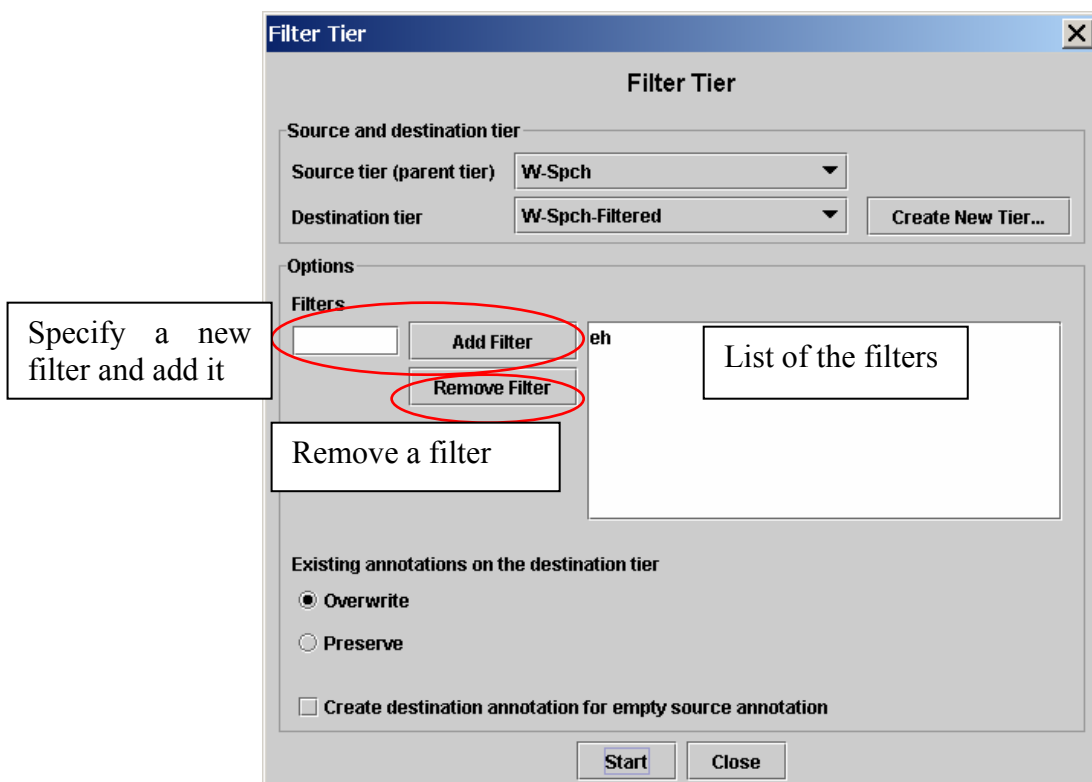
### 4.8.1 Filtering tiers

Filtering a tier (**Edit > Filter Tier...**) works in a very similar way as the previously described tokenizer. The main difference is that filtering is meant for transfer of annotation information between tiers that are symbolic associations.

A step-by-step approach to filtering:

1. Select a source tier (from which the information will be copied)
2. Choose a destination tier. If necessary create a new tier (with the **Create new tier...** button)
3. Optionally specify a filter. If a filter expression is found, it will be removed from the destination annotation. Without any filter, the complete source tier is copied to the destination tier.
4. If the destination tier already contains annotation units, choose between overwriting or preserving them. If it's still empty you can ignore this option.
5. Select **Create destination annotation for empty source annotation** if you want to create for every source annotation a destination annotation, even if it's empty.
6. Click on **Start** to begin the filter operation or **Close** to go back to ELAN's main screen.





#### 4.8.2 Copying tiers

Like you can change the parent of a tier, it also is possible to copy the complete content of a tier. This option differs in that it does not put any constraints on the designated parent tier and it leaves you the choice whether you want to copy the tier's children.

This process can be started via the **Edit > Copy tier** menu. Follow the steps below:

- (1) Choose a tier to copy. If you also want to create a copy its dependent tiers, check the **Copy dependent tiers as well** box.
- (2) Specify the parent tier for the copy. To make it independent, select **Transcription (no parent)**
- (3) By default, the linguistic type will be kept. If you want to change it, select another one from the dialog window and click on **Finish**.
- (4) Now the tier (and optionally its children) will be copied. “-cp” will be added to the names in order to prevent confusion with the original tier.

Note that this is similar to the change parent tier functionality (see section 4.3.5). However it differs in 2 aspects:

- It is not mandatory to copy the child tiers.
- The parent for the copy can any be tier in the transcription, including the tier itself.

## 4.9 How to undo an action

Most actions in Elan can be undone via the **Edit > Undo** (or CTRL+Z) menu. A redo (go back to the state before the undo action) is available to, via **Edit > Redo** (or CTRL+Y).

The following commands support undo/redo:

- Add new tier / linguistic type
- Copy tier / change parent tier
- Insert annotation before / after
- Change tier attributes / linguistic type
- Clear selection (through the icon or keyboard shortcut)
- Delete annotation / tier / linguistic type
- Modify annotation / annotation time
- Create new annotation
- Play around selection time setting
- The tokenizer
- Add / Change / Edit controlled vocabularies

## 5 Searching

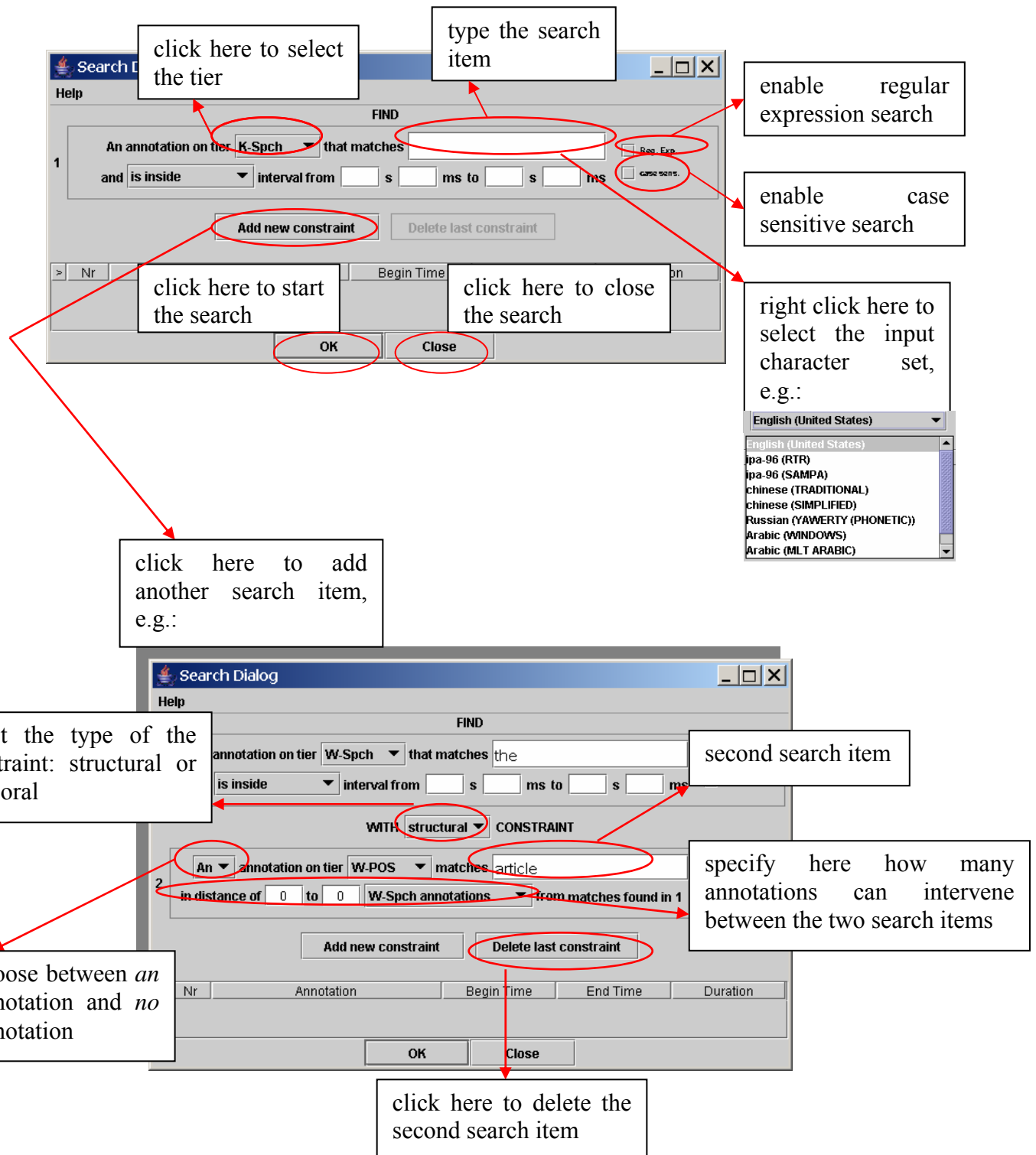
The ELAN tool allows you:

- to search annotations within one document (see section 5.1)
- to display the results (see section 5.2)
- to jump to the corresponding annotation the ELAN window (see section 5.3)
- to search through multiple annotation files (see section 5.4)

## 5.1 Searching in a single annotation file

To search for text, do the following:

- (1) Click on Search menu.
- (2) Go to Find (And Replace)... (alternatively you can press CTRL+F) The following dialog window is displayed:



(3) The following search options are available:

- (1) Go to **Annotation on tier** and, from the pull-down menu, select the tier to be searched.
- (2) Go to **matches** and type in the item to be searched.

You can always make use of regular expressions to conduct your searches when “Reg. Exp.” is checked. (see

Appendix: REGULAR EXPRESSION SEARCH for the regular expression syntax).

By default the search is not case sensitive. To change this, select the “case sens.” checkbox.

Optionally, specify the interval to search in (from ... s ...ms to ... s ... ms). Make a choice between searching within a time interval and finding annotations that overlap with a certain interval. Click on **Add new constraint** to add a second tier and search item. Up to 10 constraints can be used. There exist 2 kinds of them:

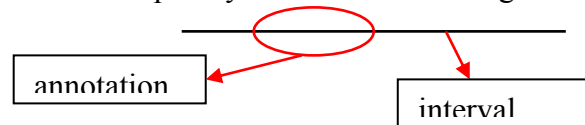
1. Constraints based on **structural distance**. (Annotation units “around” a certain annotation entity). This option is only available for tiers that are symbolically associated to (or are a symbolical subdivision of) the tier mentioned in the first search box.

For example: annotations contained in a structural distance of  $-1$  to  $2$  *tx*-annotations from *trees* on the tier *tx* are **sees, trees, and, flowers**.

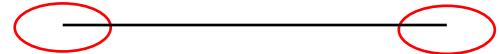
tier	annotations							
st (sentence)	He sees trees and flowers.							
tx (word)	he	sees		trees		and	flowers	
mb (morpheme break)	...	see	-s	tree	-s	...	flower	-s
ps (part of speech)	...	V	SUF	N	SUF	...	N	SUF

2. Constraints based on **temporal distance**. This means search results are restricted on the basis of the temporal relation between two intervals:

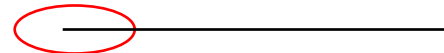
- *is inside*: the annotation is completely contained within a given interval



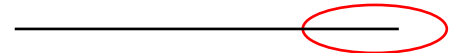
- *overlaps*: at least a part of the annotation is contained within the given interval



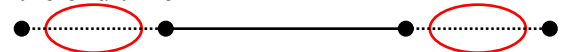
- *overlaps only begin time of*: the annotation only has its end part in common with the given interval



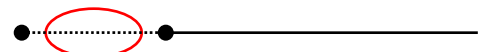
- *overlaps only end time of*: the annotation only has its begin part in common with the given interval



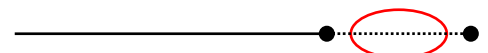
- *is within ... around*: the annotation is contained in an interval around either the begin time or the end time



- *is within ... around begin time of*: the annotation is contained in an interval around the begin time



- *is within ... around end time of*: the annotation is contained in an interval around the begin time



It is possible to search on different tiers within one annotation. For example, the search parameters illustrated below search for all annotations on the tier tx,

which contain “-s” in one of their morpheme breaks and “N” in one of their parts of speech. (Both “-s” and “N” are in distance of “0 words”, i.e., they occur within the same word as specified on the tier tx.) I.e., these parameters would find “trees” and “flowers” in the above example, but not “sees”.

Search Dialog

Help

FIND

1 An annotation on tier tx that matches \w ☒ Reg. Exp. ☐ case sens.

and is inside interval from s ms to s ms

WITH structural CONSTRAINT

2 An annotation on tier ps matches ^N\$ ☒ Reg. Exp. ☐ case sens.

in distance of 0 to 0 tx annotations from matches found in 1

WITH structural CONSTRAINT

3 An annotation on tier mb matches -s ☐ Reg. Exp. ☐ case sens.

in distance of 0 to 0 tx annotations from matches found in 1

Add new constraint Delete last constraint

>	Nr	Annotation	Begin Time	End Time	Duration
0 annotations found					

OK Close

tx matches \w (any non-empty annotation)

ps matches N

mb matches -s

Another option is searching for sequences of utterances, words or other annotations on the same tier, e.g.:

Search Dialog

Help

FIND

1 An annotation on tier tx that matches trees ☐ Reg. Exp. ☐ case sens.

and is inside interval from s ms to s ms

WITH structural CONSTRAINT

2 An annotation on tier ps matches flowers ☒ Reg. Exp. ☐ case sens.

in distance of 1 to 3 tx annotations from matches found in 1

Add new constraint Delete last constraint

>	Nr	Annotation	Begin Time	End Time	Duration
0 annotations found					

OK Close

on the tier tx, the annotation “flowers” follows 1 to 3 annotations after the annotation “trees”

You can delete the second (or third) search item. Click on Delete last constraint to delete it.



- (3) Right click in a text box to change the input character set and select the suitable language from the pull-down menu.

You only need this option if you want to select a non-default character set. The box automatically displays the default set of the selected tier (see section **Error! Reference source not found.**).

After you have specified your search parameters, click **Ok** to start the search process.

Make sure the box next to **Reg. Exp.** is checked when you search for “special” characters (i.e. all characters that are not plain letters or digits) like diacritic characters.

### 5.1.1 Advanced searching: an example

Suppose we are investigating turn taking and we want to find all switches from speaker W to speaker K that don’t overlap, with gaps of at most 2 seconds. In order to find this, we fill in the search form as follows:

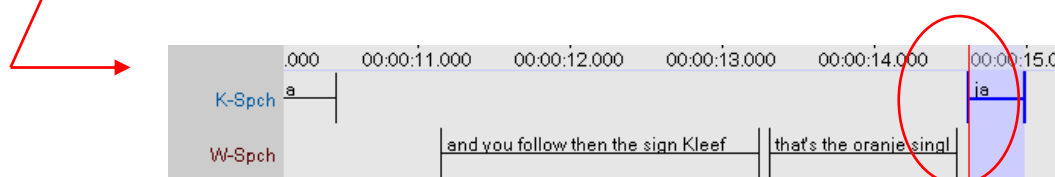
The Search Dialog window is shown with the following settings:

- FIND**
  - 1. An annotation on tier **K-Spch** that matches [ ] ☐ Reg. Exp. ☐ case sens.
  - and **is inside** interval from [ ] s [ ] ms to [ ] s [ ] ms
- WITH temporal CONSTRAINT**
  - 2. **No** annotation on tier **W-Spch** matches [ ] ☐ Reg. Exp. ☐ case sens.
  - and **overlaps** matches found in 1
- WITH temporal CONSTRAINT**
  - 3. **An** annotation on tier **W-Spch** matches [ ] ☐ Reg. Exp. ☐ case sens.
  - and **is within ... around begin time of** [ 2 ] s [ 0 ] ms matches found in 1

Buttons: Add new constraint, Delete last constraint

Nr	Annotation	Begin Time	End Time	Duration
1	ja	00:00:14.620	00:00:14.990	00:00:00.370
2	rotunda	00:00:27.030	00:00:27.990	00:00:00.960

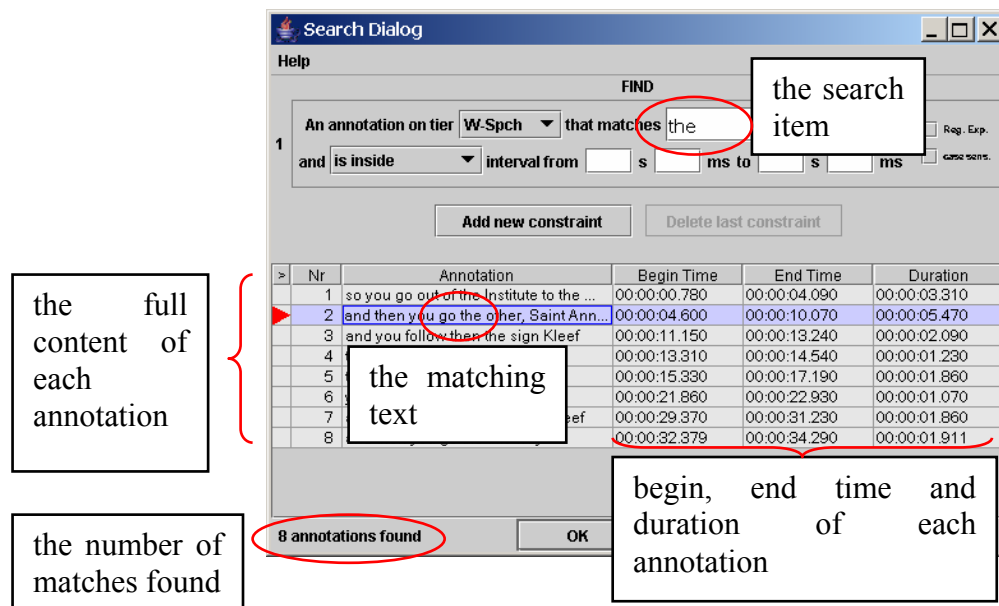
2 annotations found [OK] [Close]



## 5.2 Displaying search results

After the Search is completed, ELAN displays the following search results:

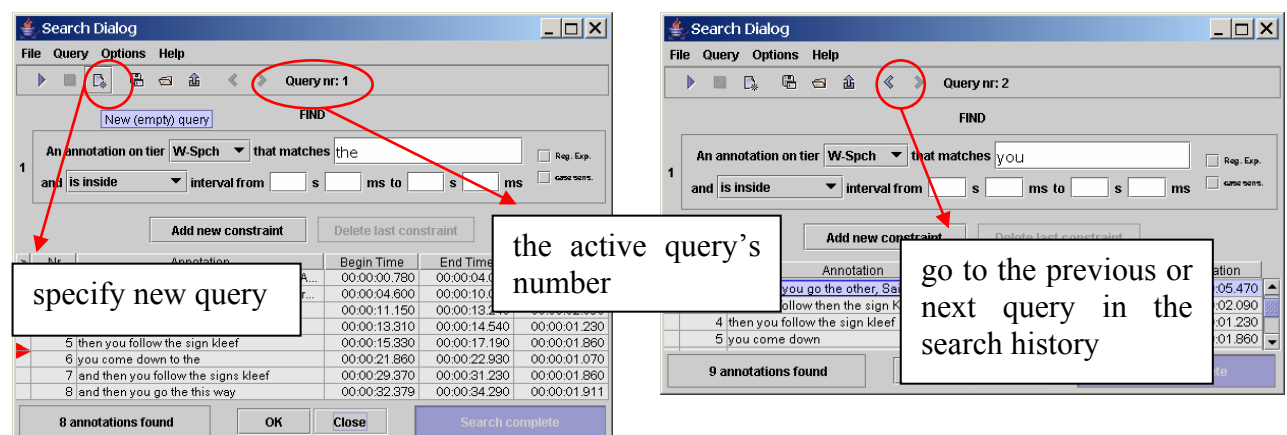
- The number of annotations where the search item was found.
- The full content of each annotation where the search item was found.
- The begin, end time and duration of each annotation where the search item was found. To see more information per match, right click in the results and check the other desired fields (file – tier – before – after ).



### 5.2.1 The search history

When performing multiple search actions one after another, a search history is created. With this feature it becomes possible to browse through all the entered queries:

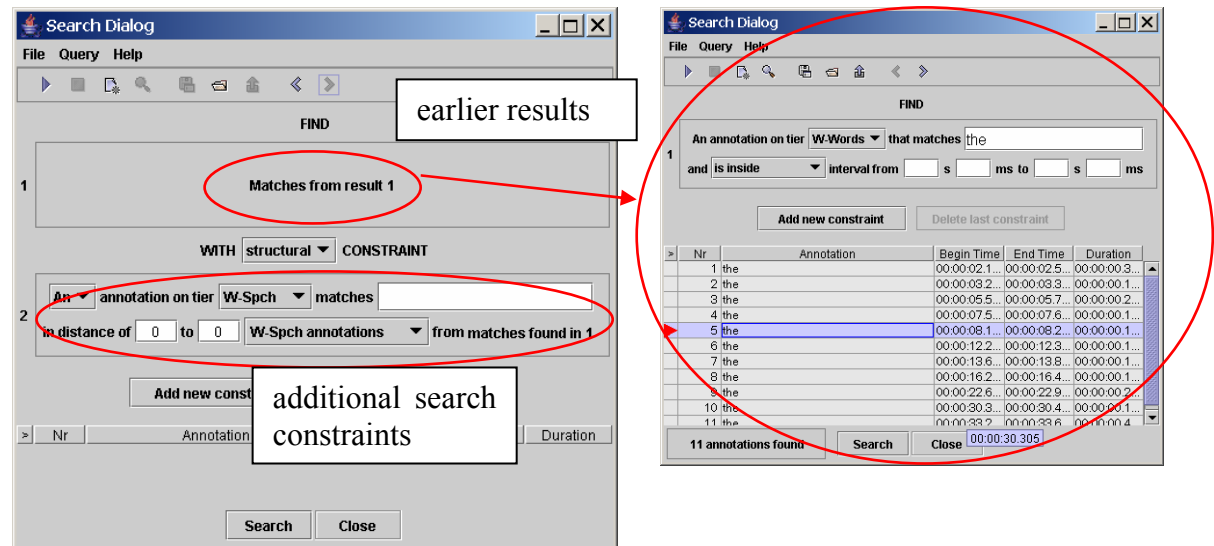
- (1) Enter a query and click OK to activate it.
- (2) Click on the New Query icon
- (3) Enter another query and choose OK
- (4) Now you can use the back and forward icons to browse through the search history.



! Note: when closing the search dialog, the query history is removed. If nonetheless you want to save a certain search command, have a look at section 5.2.3

### 5.2.2 Refining a search

After a search operation, one can perform another search within the results of the previous action. To achieve this, select **Search on current result** from the **Query** menu in the Search dialog window. This will get you the following window:



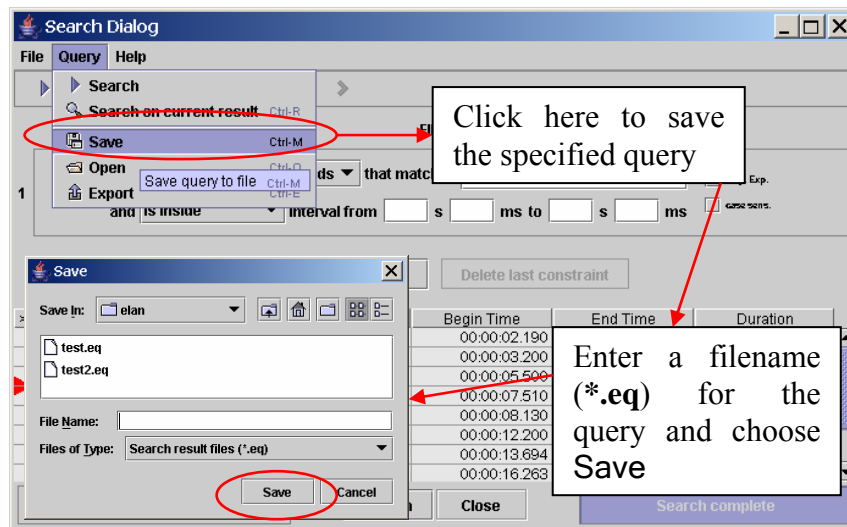
By specifying extra search constraints, you can narrow down the results. This is similar to the addition of an extra search constraint.

### 5.2.3 Saving and loading queries

When using complex queries, or one is repeating the same search on different annotation files, it is convenient to reuse a query. This is possible in ELAN thanks to the save and load query functionality.

#### Saving a query

- (1) Specify a query in the search dialog.
- (2) Either choose **Query > Save** or click on the save icon in the toolbar.
- (3) Enter a filename for the active query.
- (4) Choose save.



## Loading a query

- (1) Either choose **Query > Open** or click on the Open icon in the toolbar.
- (2) Select an existing query from the file dialog.
- (3) Choose Open.

### 5.2.4 Exporting the search results

To export the results of a query to a tab-separated file, go through these steps:

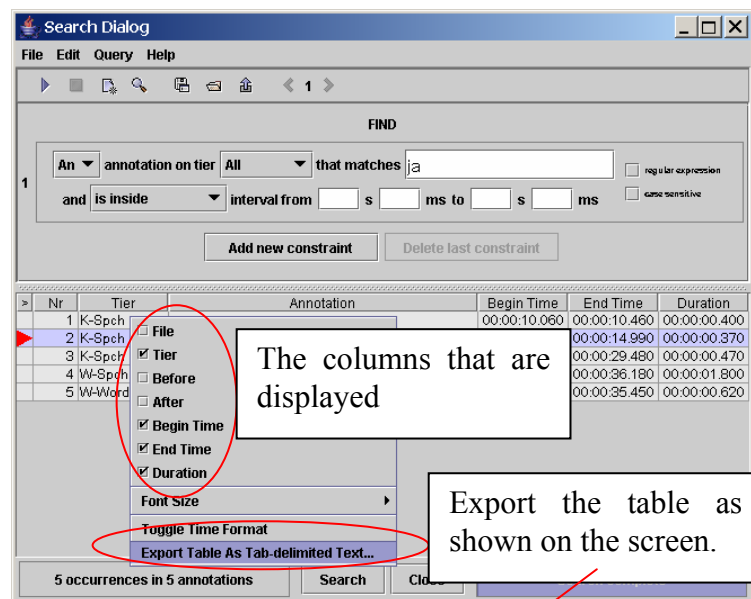
- (1) In the query dialog, select **Query > Export matches**
- (2) Specify a file name
- (3) Click on save

An exported file looks as follows:

<div>the end time of the annotation (both in hh:mm:ss.mmm format and seconds)</div>							
W-Words	00:00:02.190	2.19	00:00:02.570	2.57	00:00:00.380	0.38	the
W-Words	00:00:03.200	3.2	00:00:03.340	3.34	00:00:00.140	0.14	the
W-Words	00:00:05.500	5.5	00:00:05.700	5.7	00:00:00.200	0.2	the
Tier name		the begin time of the annotation (both in hh:mm:ss.mmm format and seconds)		the duration of 1 annotation unit		query result	

If you right click in the table containing the search results, a popup menu appears. Tick the checkboxes to show or hide columns that are related to the found

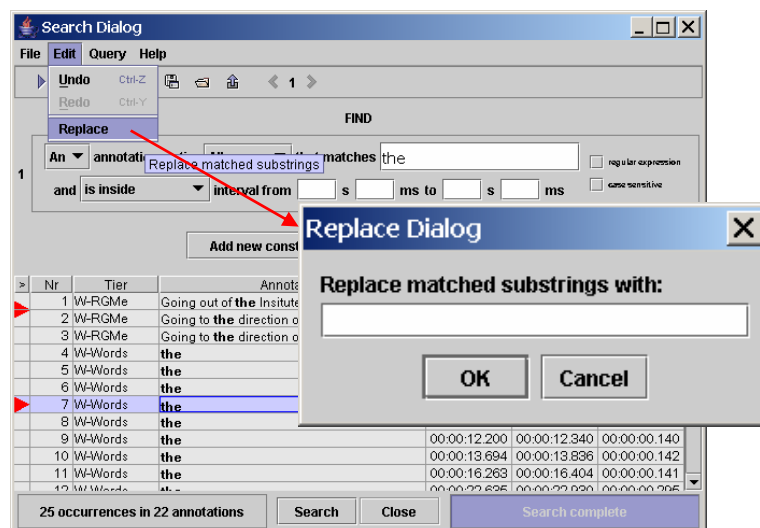
annotations. In the same popup menu you will find the option **Export Table as tab-delimited text**, which literally saves the displayed result table to a text file, as shown in the example below.



Nr	Annotation	Begin Time	End Time	Duration
1	the	00:00:02.190	00:00:02.570	00:00:00.380
2	the	00:00:03.200	00:00:03.340	00:00:00.140
3	the	00:00:05.500	00:00:05.700	00:00:00.200

### 5.2.5 Search and replace

ELAN offers the possibility to replace all of the found search results. Choose **Edit > Replace** in the Search Dialog:



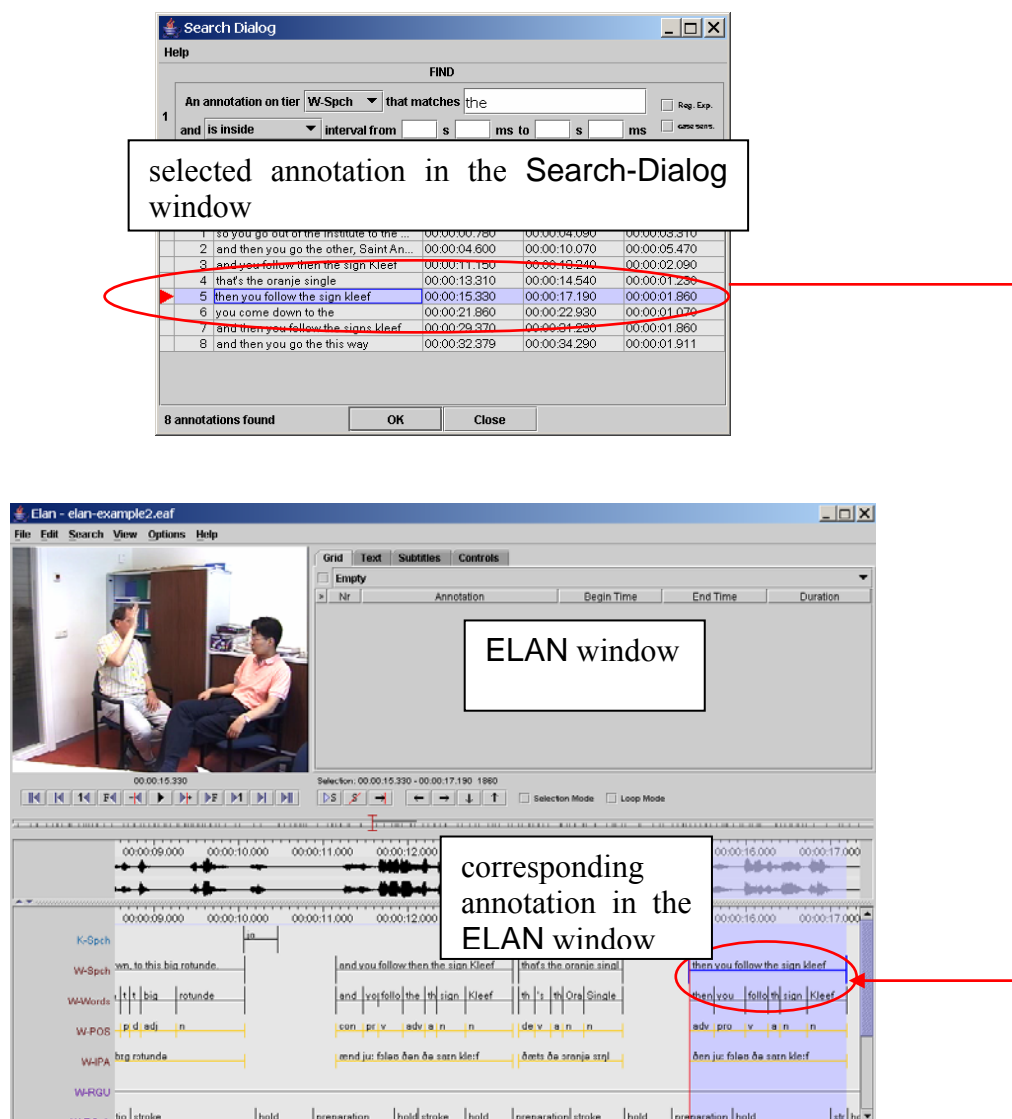
Then enter the text that should replace the found results and choose **OK**

### 5.3 Jumping from the search results to annotations

You can jump from the annotations displayed in the Search-Dialog window to the corresponding annotations in the ELAN window. This option allows you to access other types of information that are linked to the annotations found: to listen to the audio, to watch the video, to view the waveform or to read the annotations on different tiers.

Do the following: In the Search-Dialog window, click on the annotation that you want to jump to. It will be highlighted in blue color. In the ELAN window, the corresponding annotation is automatically accessed.

The screen display should look similar to the following illustration:



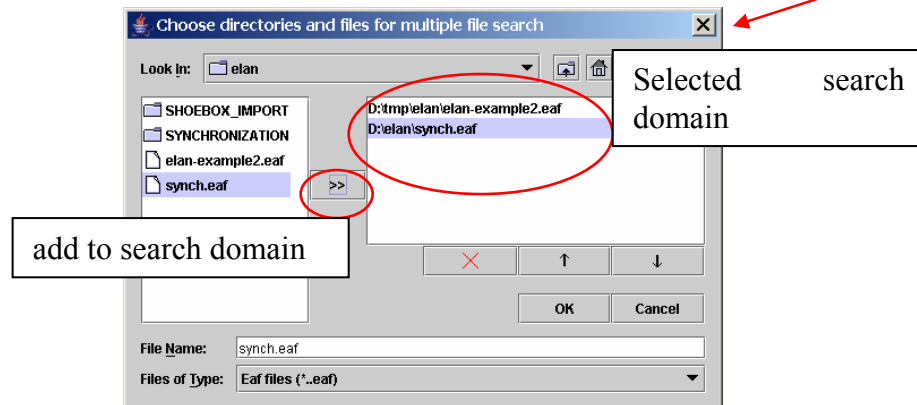
### 5.4 Searching through multiple annotation files

Elan offers an option to search for an expression through multiple files. To access it, go to Search > Search multiple eaf... This will open the following dialog box:

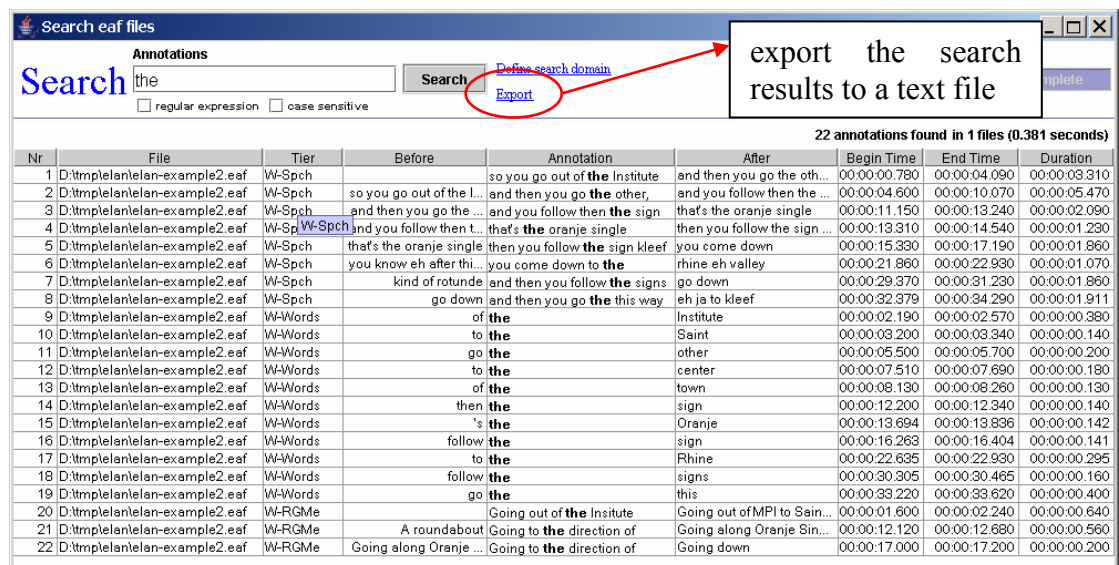


In order to perform a multiple file search, go through the following steps:

1. Click on **define search domain**
2. Add the files and/or directories you want to be searched in the dialog window



3. Click on **OK** to confirm the selection
4. Enter a search expression, and optionally enable a regular expression and/or case sensitive search.
5. Click on **Search**. The result screen will appear:



The result window contains the following fields for every found annotation:

- **Nr:** the order in which the results appear
- **File:** the file that contains the result
- **Before / After:** the annotation as found before and after the annotation that matches the search expression
- **Annotation:** the search result itself
- **Begin time, end time, duration:** of the annotation unit that was found

It is not possible to restrict the search results to a certain tier or to specify extra structural or temporal constraints.

The displayed search results can be exported to a tab-separated text file as well. The exported files are very similar to that described in section 5.2.4.

- Just click on **Export** next to the **Search** button.
- Enter a filename.
- Select the **Save** button.

If you click on one of the listed annotations, a new Elan window will be opened, and that annotation unit will be selected. When clicking on another result, the newly opened window is reused.



## **PART III: REFERENCE GUIDE**

---

In this reference guide, you find concise descriptions of the mouse options (1), the menu item (2), and the shortcut keys (3). In addition, a brief definition of key concepts is provided (4), and the new features of version 1.4 (compared to earlier versions) are listed (5).

### **5.5 The Mouse Options**

#### **5.5.1 Resizing Viewers**

- Click on the up/down arrows in the split-pane to increase/decrease the size of the corresponding Viewer.
- Go with the mouse to the split-pane and move it up/down to increase/decrease the size of the corresponding Viewer.

#### **5.5.2 Rearranging the order of tiers**

- Drag a tier label from its place in the Timeline or Interlinear Viewer to any other place in the Timeline Viewer.

#### **5.5.3 Make a tier the active tier**

- Double click on a tier label.
- Right click in the Timeline or Interlinear Viewer and choose Active Tier.

#### **5.5.4 Accessing points in time**

- Click on any place in the Timeline, Waveform or Annotation Density Viewer to jump to this point in time.
- Click on one of the media controls to go forward/backward in time.
- Click on the timecode box and enter a timecode to jump to this point in time.

#### **5.5.5 Playing a document**

- Click on the **Play/Pause** icon to start or pause the playback.
- Use the Rate slider (in the Controls tab) to increase/decrease the playback rate.

#### **5.5.6 Making a selection**

- Click with the mouse on any point in the Timeline or Waveform Viewer, and drag it to another point to select a time interval. Or click on an annotation in the Grid Viewer and drag the mouse to another annotation.
- Click on an annotation in the Timeline or Interlinear Viewer to select it.

#### **5.5.7 Adding/modifying an annotation**

- Double-click on an annotation or selection (at about the height of the tier where you want to enter an annotation). The **Inline Edit** box appears and you can enter the annotation.

## 5.6 The Menu Items

The starred options (\*) are also available whenever you right-click somewhere within the Timeline or Interlinear Viewer (at about the height of the relevant tier).

### 5.6.1 File menu

- **New:** Creates a new annotation file (\*.eaf) for an existing media file (\*.wav, \*.mpg/\*.mov).
- **Open:** Opens an existing annotation file (\*.eaf).
- **Close:** Closes the current ELAN document.
- **Save:** Saves changes to the annotation file (\*.eaf).
- **Automatic backup:** Creates an automatic backup copy.
- **Page Setup:** Changes the page settings of a printer
- **Print Preview:** Gives a preview of the page to be printed
- **Print:** Prints annotations
- **Export as > tab-delimited text ...:** Exports the document as a tab-delimited text file (\*.txt).
- **Export > as SMIL...:** Exports the document as a SMIL file (\*.smil).
- **Export > Quicktime text ...:** Exports the document as a quicktime subtitle file (\*.txt).
- **Export > as Shoebox file ...:** Exports the document as a Shoebox file (\*.txt).
- **Export > as CHAT file ...:** Exports the document as a CHAT file (\*.cha).
- **Export > as Traditional Transcript file ...:** Exports the document as a text file (\*.txt).
- **Export > as Media clip ...:** Exports a part of the video as an MPEG file. (\*.mpg \*.mpeg)
- **Export > as Image from Elan window ...:** Exports a screenshot (\*.jpg \*.png).
- **Import > Shoebox file ...:** Imports a Shoebox file (\*.txt).
- **Import > CHAT file ...:** Imports a Chat file (\*.cha).
- **Import > Transcriber file ...:** Imports a transcriber file (\*.trs).
- **Quit:** Quits the ELAN window.

### 5.6.2 Edit menu

- **Undo:** Go back to the previous state, see also section 4.9.
- **Redo:** Go back to the previous state, before the undo was performed, see also section 4.9.
- **(\*) New annotation here:** Creates a new annotation on the active tier. The option is only available if a time interval or an annotation is selected.
- **(\*) New annotation before:** Subdivides an existing annotation (and adds the new part to the left of the old annotation). The option is only available if an annotation is selected and if the corresponding tier is assigned to either the stereotype Time Subdivision or Symbolic Subdivision.
- **(\*) New annotation after:** Subdivides an existing annotation (and adds the new part to the right of the old annotation). The option is only available if an annotation is selected and if the corresponding tier is assigned to either the stereotype Time Subdivision or Symbolic Subdivision.
- **(\*) Modify annotation value:** Modifies the content of an annotation. The option is only available if an annotation is selected.
- **(\*) Delete annotation:** Deletes an existing annotation. The option is only available if an annotation is selected.
- **Shift all annotations...:** Moves all annotations to the right or to the left
- **Tokenize tier:** Auto-segments a tier into parts on another tier.
- **Filter tier:** Copy annotations from one tier to another and optionally filter them.
- **Add tier:** Displays a dialog window that allows you to add a new tier.
- **(\*) Change tier attributes:** Changes the attributes of an existing tier.
- **(\*) Delete tier:** Deletes an existing tier.
- **Add linguistic type:** Displays a dialog window that allows you to add a new linguistic type.
- **Change linguistic type:** Changes the attributes of an existing linguistic type.
- **Delete linguistic type:** Deletes an existing linguistic type.
- **Edit Controlled Vocabularies:** Add, delete or change Controlled Vocabularies as used by Linguistic Types.
- **Linked files...:** Manage the references to media files

### 5.6.3 Search menu

- **Search ...:** Allows you to search for text within an ELAN document.
- **Search multiple EAF files...:** Performs a search action through multiple annotation files.
- **Go To ...:** Go to a certain point of time.

#### 5.6.4 View menu

- Tier Dependencies: Shows the hierarchical relations between tiers.
- Shortcuts: Displays a list of shortcut keys.

#### 5.6.5 Options menu

- Propagate time changes > Normal mode: activates the Bulldozer mode.
- Propagate time changes > Bulldozer mode: activates the Bulldozer mode.
- Propagate time changes > Shift mode: activates the shift mode.
- Annotation / Media Synchronisation Mode: toggle between these 2 modes
- Play around selection: Plays a few (milli-)seconds before and after the current selection.
- Rate and volume toggle: enter playback rate and volume values to switch to with a keyboard shortcut
- Video Standard: make a choice between PAL and NTSC
- Language: choose the interface language

#### 5.6.6 Help menu

- About EUDICO Linguistic Annotator: Displays the version and copyright information.

### 5.7 The Shortcut Keys

#### 5.7.1 File options

CTRL+S	<u>S</u> aves the current project
CTRL+O	<u>O</u> pen a document
CTRL+P	<u>P</u> rints the current document
CTRL+N	Create a new document

### 5.7.2 Playing a document & making selections

CTRL+SPACE	Starts or stops the Media Player
CTRL+B	Go to the beginning of the video/audio fragment
CTRL+PAGE UP	Go to the previous scroll view
SHIFT+LEFT	Go back one second
CTRL+LEFT	Go back one frame
CTRL+SHIFT+LEFT	Go back one “pixel” on the annotation density viewer
CTRL+SHIFT+RIGHT	Go to the next “pixel” on the annotation density viewer
CTRL+RIGHT	Go to the next frame
SHIFT+RIGHT	Go to the next second
CTRL+PAGE DOWN	Go to the next scroll view
CTRL+E	Go to the end of the media fragment
CTRL+ALT+R	Switch to predefined playback rate
CTRL+ALT+V	Switch to predefined volume
Ctrl+G	Go to time

### 5.7.3 Working with annotations

ALT+N	Opens the <b>Inline Edit</b> box for inserting a <u>n</u> ew annotation (only available if a time interval or annotation is selected)
ALT+M	Opens the <b>Inline Edit</b> box for <u>m</u> odifying an annotation (only available if an annotation is selected)
ALT+D	<u>D</u> eletes an annotation (only available if an annotation is selected)
ALT+LEFT	Go to previous Annotation
ALT+RIGHT	Go to next Annotation
ALT+UP	Go annotation upward
ALT+DOWN	Go annotation downward
ALT+T	Add new tier
ALT+R	Delete tier
CTRL+UP	Set previous tier active
CTRL+DOWN	Set next tier active
CTRL+ALT+SHIFT	Show empty slots in multitier viewer
CTRL+ENTER	Saves the changes made to the <b>Inline Edit</b> box or the <b>Edit annotation</b> box
ESC	(from within the <b>Inline Edit</b> box or the <b>Edit annotation</b> box): exits the box without saving the changes
ESC	(in all other contexts): deselects a selection

SHIFT+ENTER	Allows you to switch between the Inline Edit box and the Edit annotation box
-------------	--

#### 5.7.4 Working with selections

SHIFT+SPACE	Play selection
CTRL+SHIFT+SPACE	Play around selection
CTRL+C	Clear selection
CTRL+/	Move the crosshair to the right or left of the selection
CTRL+K	Toggle selection mode
CTRL+L	Toggle loop mode

#### 5.7.5 Searching

CTRL+F	Find
CTRL+SHIFT+F	Search in multiple eaf files

#### 5.7.6 General

CTRL+Z	Undo
CTRL+Y	Redo

## 5.8 Key concepts of ELAN

### 5.8.1 Annotation file (\*.eaf)

An annotation file is the document that contains all the information about tiers (their attributes and dependency relations), annotations and time alignments.

### 5.8.2 Media file (\*.mpg/\*.mov, \*.wav)

A media file contains the digitized video/audio (e.g. **\*.mpg**) or only the audio (**\*.wav**) data. It determines the time axis to which the annotation file (**\*.eaf**) is linked.

### 5.8.3 Annotation

An annotation is any type of text (e.g. a transcription, a translation, coding, etc.) that is entered on a tier. It is assigned to a selected time interval of the video/audio file (e.g., to the time interval corresponding to the utterance of a speaker) or to an annotation on another tier (e.g., a translation is assigned to an orthographic transcription).

### 5.8.4 Tier

A tier is a set of annotations that share the same characteristics, e.g., one tier containing the orthographic transcription, or another tier containing the free translation.

A tier can be ‘independent’ and ‘time-alignable’, in which case it is directly linked to a time interval of the media file (e.g., the ‘orthographic transcription’ tier). Or it can

be ‘referring’, in which case it is linked to another tier, its so-called parent tier (e.g., the ‘orthographic transcription’ tier is a parent tier to the ‘free translation’ tier). The referring tier shares its time alignment with its parent tier. Some referring tiers can be assigned to the time axis, but only to an interval that is contained within the interval of their parent annotation.

It is possible to build nested hierarchies, e.g., the ‘orthographic transcription’ tier is the parent tier to a ‘word’ tier, and the ‘word’ tier is the parent tier to a ‘morpheme break’ tier.

Tiers are assigned to linguistic types, which specify certain constraints. The following constraints exist: None (independent, time-alignable tiers), Time Subdivision (the annotation on the referring tier can be subdivided and linked to the time axis), Symbolic Subdivision (the annotation on the referring tier can be subdivided, but not linked to the time axis), Symbolic Association (one annotation on the referring tier corresponds to exactly one annotation on the parent tier).

## APPENDIX: REGULAR EXPRESSION SEARCH

---

### Brief Background

A regular expression consists of a character string where some characters are given special meaning with regard to pattern matching. Regular expressions have been in use from the early days of computing, and provide a powerful and efficient way to parse, interpret and search and replace text within an application.

### Supported Syntax<sup>10</sup>

#### Characters

<i>x</i>	The character <i>x</i>
<code>\\</code>	The backslash character
<code>\0n</code>	The character with octal value <i>0n</i> ( $0 \leq n \leq 7$ )
<code>\0nn</code>	The character with octal value <i>0nn</i> ( $0 \leq n \leq 7$ )
<code>\0mnn</code>	The character with octal value <i>0mnn</i> ( $0 \leq m \leq 3, 0 \leq n \leq 7$ )
<code>\xhh</code>	The character with hexadecimal value <i>0xhh</i>
<code>\uhhhh</code>	The character with hexadecimal value <i>0xhhhh</i>
<code>\t</code>	The tab character (' <code>\u0009</code> ')
<code>\n</code>	The newline (line feed) character (' <code>\u000A</code> ')
<code>\r</code>	The carriage-return character (' <code>\u000D</code> ')
<code>\f</code>	The form-feed character (' <code>\u000C</code> ')
<code>\a</code>	The alert (bell) character (' <code>\u0007</code> ')
<code>\e</code>	The escape character (' <code>\u001B</code> ')
<code>\cx</code>	The control character corresponding to <i>x</i>

#### Character classes

<code>[abc]</code>	a, b, or c (simple class)
<code>[^abc]</code>	Any character except a, b, or c (negation)
<code>[a-zA-Z]</code>	a through z or A through Z, inclusive (range)
<code>[a-d[m-p]]</code>	a through d, or m through p: <code>[a-dm-p]</code> (union)
<code>[a-z&amp;&amp;[def]]</code>	d, e, or f (intersection)
<code>[a-z&amp;&amp;[^bc]]</code>	a through z, except for b and c: <code>[ad-z]</code> (subtraction)
<code>[a-z&amp;&amp;[^m-p]]</code>	a through z, and not m through p: <code>[a-lq-z]</code> (subtraction)

#### Predefined character classes

<code>.</code>	Any character (may or may not match <a href="#">line terminators</a> )
<code>\d</code>	A digit: <code>[0-9]</code>
<code>\D</code>	A non-digit: <code>[^0-9]</code>
<code>\s</code>	A whitespace character: <code>[\t\n\r\f]</code>
<code>\S</code>	A non-whitespace character: <code>[^\s]</code>
<code>\w</code>	A word character: <code>[a-zA-Z_0-9]</code>
<code>\W</code>	A non-word character: <code>[^\w]</code>

---

<sup>10</sup> Source: <http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html>



## POSIX character classes (US-ASCII only)

<code>\p{Lower}</code>	A lower-case alphabetic character: <code>[a-z]</code>
<code>\p{Upper}</code>	An upper-case alphabetic character: <code>[A-Z]</code>
<code>\p{ASCII}</code>	All ASCII: <code>[\x00-\x7F]</code>
<code>\p{Alpha}</code>	An alphabetic character: <code>[\p{Lower}\p{Upper}]</code>
<code>\p{Digit}</code>	A decimal digit: <code>[0-9]</code>
<code>\p{Alnum}</code>	An alphanumeric character: <code>[\p{Alpha}\p{Digit}]</code>
<code>\p{Punct}</code>	Punctuation: One of <code>!"#\$%&amp;'()*+,-./:;&lt;=&gt;?@[ \]^_`{ }~</code>
<code>\p{Graph}</code>	A visible character: <code>[\p{Alnum}\p{Punct}]</code>
<code>\p{Print}</code>	A printable character: <code>[\p{Graph}]</code>
<code>\p{Blank}</code>	A space or a tab: <code>[\t]</code>
<code>\p{Cntrl}</code>	A control character: <code>[\x00-\x1F\x7F]</code>
<code>\p{XDigit}</code>	A hexadecimal digit: <code>[0-9a-fA-F]</code>
<code>\p{Space}</code>	A whitespace character: <code>[\t\n\x0B\f\r]</code>

## Classes for Unicode blocks and categories

<code>\p{InGreek}</code>	A character in the Greek block (simple <a href="#">block</a> )
<code>\p{Lu}</code>	An uppercase letter (simple <a href="#">category</a> )
<code>\p{Sc}</code>	A currency symbol
<code>\P{InGreek}</code>	Any character except one in the Greek block (negation)
<code>[ \p{L} &amp; [ ^ \p{Lu} ] ]</code>	Any letter except an uppercase letter (subtraction)

## Boundary matchers

<code>^</code>	The beginning of a line
<code>\$</code>	The end of a line
<code>\b</code>	A word boundary
<code>\B</code>	A non-word boundary
<code>\A</code>	The beginning of the input
<code>\G</code>	The end of the previous match
<code>\Z</code>	The end of the input but for the final <a href="#">terminator</a> , if any
<code>\z</code>	The end of the input

## Greedy quantifiers

<code>X?</code>	<code>X</code> , once or not at all
<code>X*</code>	<code>X</code> , zero or more times
<code>X+</code>	<code>X</code> , one or more times
<code>X{n}</code>	<code>X</code> , exactly $n$ times
<code>X{n,}</code>	<code>X</code> , at least $n$ times
<code>X{n,m}</code>	<code>X</code> , at least $n$ but not more than $m$ times

## Reluctant quantifiers

<code>X??</code>	<code>X</code> , once or not at all
<code>X*?</code>	<code>X</code> , zero or more times
<code>X+?</code>	<code>X</code> , one or more times
<code>X{n}?</code>	<code>X</code> , exactly $n$ times

$X\{n,\}?$	$X$ , at least $n$ times
$X\{n,m\}?$	$X$ , at least $n$ but not more than $m$ times

### Possessive quantifiers

$X?+$	$X$ , once or not at all
$X^{*+}$	$X$ , zero or more times
$X^{++}$	$X$ , one or more times
$X\{n\}^{+}$	$X$ , exactly $n$ times
$X\{n,\}^{+}$	$X$ , at least $n$ times
$X\{n,m\}^{+}$	$X$ , at least $n$ but not more than $m$ times

### Logical operators

$XY$	$X$ followed by $Y$
$X Y$	Either $X$ or $Y$
$(X)$	$X$ , as a <a href="#">capturing group</a>

### Back references

$\backslash n$	Whatever the $n^{\text{th}}$ <a href="#">capturing group</a> matched
----------------	--

### Quotation

$\backslash$	Nothing, but quotes the following character
$\backslash Q$	Nothing, but quotes all characters until $\backslash E$
$\backslash E$	Nothing, but ends quoting started by $\backslash Q$

### Special constructs (non-capturing)

$(?:X)$	$X$ , as a non-capturing group
$(?idsux-idsux)$	Nothing, but turns match flags on - off
$(?idsux-idsux:X)$	$X$ , as a <a href="#">non-capturing group</a> with the given flags on - off
$(?=X)$	$X$ , via zero-width positive lookahead
$(?!X)$	$X$ , via zero-width negative lookahead
$(?<=X)$	$X$ , via zero-width positive lookbehind
$(?<!X)$	$X$ , via zero-width negative lookbehind
$(?>X)$	$X$ , as an independent, non-capturing group

### Backslashes, escapes, and quoting

The backslash character (`'\'`) serves to introduce escaped constructs, as defined in the table above, as well as to quote characters that otherwise would be interpreted as unescaped constructs. Thus the expression `\\` matches a single backslash and `\{` matches a left brace.

It is an error to use a backslash prior to any alphabetic character that does not denote an escaped construct; these are reserved for future extensions to the regular-expression language. A backslash may be used prior to a non-alphabetic character regardless of whether that character is part of an unescaped construct.

Backslashes within string literals in Java source code are interpreted as required by the [Java Language Specification](#) as either [Unicode escapes](#) or other [character escapes](#). It is therefore necessary to double backslashes in string literals that represent regular

expressions to protect them from interpretation by the Java bytecode compiler. The string literal `"\b"`, for example, matches a single backspace character when interpreted as a regular expression, while `"\\b"` matches a word boundary. The string literal `"\ (hello\)"` is illegal and leads to a compile-time error; in order to match the string `(hello)` the string literal `"\\ (hello\\)"` must be used.

## Character Classes

Character classes may appear within other character classes, and may be composed by the union operator (implicit) and the intersection operator (`&&`). The union operator denotes a class that contains every character that is in at least one of its operand classes. The intersection operator denotes a class that contains every character that is in both of its operand classes.

The precedence of character-class operators is as follows, from highest to lowest:

- |   |                |                                     |
|---|----------------|-------------------------------------|
| 1 | Literal escape | <code>\x</code>                     |
| 2 | Grouping       | <code>[...]</code>                  |
| 3 | Range          | <code>a-z</code>                    |
| 4 | Union          | <code>[a-e][i-u]</code>             |
| 5 | Intersection   | <code>[a-z&amp;&amp;[aeiou]]</code> |

Note that a different set of metacharacters are in effect inside a character class than outside a character class. For instance, the regular expression `.` loses its special meaning inside a character class, while the expression `-` becomes a range forming metacharacter.

## Line terminators

A *line terminator* is a one- or two-character sequence that marks the end of a line of the input character sequence. The following are recognized as line terminators:

- A newline (line feed) character (`'\n'`),
- A carriage-return character followed immediately by a newline character (`"\r\n"`),
- A standalone carriage-return character (`'\r'`),
- A next-line character (`'\u0085'`),
- A line-separator character (`'\u2028'`), or
- A paragraph-separator character (`'\u2029'`).

If [UNIX\\_LINES](#) mode is activated, then the only line terminators recognized are newline characters.

The regular expression `.` matches any character except a line terminator unless the [DOTALL](#) flag is specified.

By default, the regular expressions `^` and `$` ignore line terminators and only match at the beginning and the end, respectively, of the entire input sequence. If [MULTILINE](#) mode is activated then `^` matches at the beginning of input and after any line terminator except at the end of input. When in [MULTILINE](#) mode `$` matches just before a line terminator or the end of the input sequence.

## Groups and capturing

Capturing groups are numbered by counting their opening parentheses from left to right. In the expression `((A)(B(C)))`, for example, there are four such groups:

- 1 ((A)(B(C)))
- 2 (A)
- 3 (B(C))
- 4 (C)

Group zero always stands for the entire expression.

Capturing groups are so named because, during a match, each subsequence of the input sequence that matches such a group is saved. The captured subsequence may be used later in the expression, via a back reference, and may also be retrieved from the matcher once the match operation is complete.

The captured input associated with a group is always the subsequence that the group most recently matched. If a group is evaluated a second time because of quantification then its previously-captured value, if any, will be retained if the second evaluation fails. Matching the string "aba" against the expression `(a(b)?)+`, for example, leaves group two set to "b". All captured input is discarded at the beginning of each match.

Groups beginning with `(?` are pure, *non-capturing* groups that do not capture text and do not count towards the group total.

Unicode support

This class follows [Unicode Technical Report #18: Unicode Regular Expression Guidelines](#), implementing its second level of support though with a slightly different concrete syntax.

Unicode escape sequences such as `\u2014` in Java source code are processed as described in [23.3](#) of the Java Language Specification. Such escape sequences are also implemented directly by the regular-expression parser so that Unicode escapes can be used in expressions that are read from files or from the keyboard. Thus the strings `"\u2014"` and `"\\u2014"`, while not equal, compile into the same pattern, which matches the character with hexadecimal value `0x2014`.

Unicode blocks and categories are written with the `\p` and `\P` constructs as in Perl. `\p{prop}` matches if the input has the property *prop*, while `\P{prop}` does not match if the input has that property. Blocks are specified with the prefix `In`, as in `InMongolian`. Categories may be specified with the optional prefix `Is`: Both `\p{L}` and `\p{IsL}` denote the category of Unicode letters. Blocks and categories can be used both inside and outside of a character class.

The supported blocks and categories are those of [The Unicode Standard, Version 3.0](#). The block names are those defined in Chapter 14 and in the file [Blocks-3.txt](#) of the [Unicode Character Database](#) except that the spaces are removed; "Basic Latin", for example, becomes "BasicLatin". The category names are those defined in table 4-5 of the Standard (p. 88), both normative and informative.

## Comparison to Perl 5

Perl constructs not supported by this class:

The conditional constructs `(?{X})` and `(?(condition)X|Y)`,

The embedded code constructs `(?{code})` and `(??{code})`,

The embedded comment syntax `(?#comment)`, and

The preprocessing operations `\l \u, \L, and \U`.

Constructs supported by this class but not by Perl:

Possessive quantifiers, which greedily match as much as they can and do not back off, even when doing so would allow the overall match to succeed.

Character-class union and intersection as described [above](#).

Notable differences from Perl:

In Perl, `\1` through `\9` are always interpreted as back references; a backslash-escaped number greater than 9 is treated as a back reference if at least that many subexpressions exist, otherwise it is interpreted, if possible, as an octal escape. In this class octal escapes must always begin with a zero. In this class, `\1` through `\9` are always interpreted as back references, and a larger number is accepted as a back reference if at least that many subexpressions exist at that point in the regular expression, otherwise the parser will drop digits until the number is smaller or equal to the existing number of groups or it is one digit.

Perl uses the `g` flag to request a match that resumes where the last match left off. This functionality is provided implicitly by the [Matcher](#) class: Repeated invocations of the [find](#) method will resume where the last match left off, unless the matcher is reset.

In Perl, embedded flags at the top level of an expression affect the whole expression. In this class, embedded flags always take effect at the point at which they appear, whether they are at the top level or within a group; in the latter case, flags are restored at the end of the group just as in Perl.

Perl is forgiving about malformed matching constructs, as in the expression `*a`, as well as dangling brackets, as in the expression `abc]`, and treats them as literals. This class also accepts dangling brackets but is strict about dangling metacharacters like `+`, `?` and `*`, and will throw a [PatternSyntaxException](#) if it encounters them.

For a more precise description of the behavior of regular expression constructs, please see [\*Mastering Regular Expressions, 2nd Edition\*, Jeffrey E. F. Friedl, O'Reilly and Associates, 2002.](#)