

# Sequence Viewer



## Tutorial



# Contents

<b>PART I: GETTING STARTED</b>		<b>1</b>
1. <i>The main sequence window</i>		1
1.1. Sequence Viewer files	1	
1.2. Sequences and events	1	
1.3. Event texts and event codes	2	
1.4. Onset time	2	
1.5. Link states	2	
1.6. More information	3	
2. <i>Event variables</i>		4
2.1. Adding event variables	4	
2.2. Getting help	4	
2.3. Computing event variables	5	
2.4. Other properties of event variables	6	
3. <i>Code variables</i>		8
3.1. About coding schemes	8	
3.2. Defining code variables	8	
3.3. Analysing event codes	9	
3.4. Sequence definition	10	
3.5. Changing codes	12	
4. <i>Extensions of the main sequence window</i>		13
4.1. Sequence variables	13	
4.2. Information fields	13	
5. <i>Adding new data</i>		14
5.1. Adding a new sequence	14	
5.2. Linking a sound file	14	
5.3. Transcribing text, using a sound file	14	
5.4. Coding event texts	15	
5.5. Adding event times	15	
5.6. Changing sequence variables	16	
5.7. Changing info fields	16	
5.8. Closing and saving	16	
<b>PART II: Selected topics</b>		<b>17</b>
6. <i>Creating a sequence file</i>		17
6.1. Creating the new file	17	
6.2. File settings	17	
6.3. Linking movies	18	
6.4. Entering start times	19	
7. <i>Importing files and analyses</i>		20
7.1. The life events file	20	
7.2. Some analyses with the life events file	22	
Optimal matching	22	
Transition matrix	23	
Selecting sequences	24	
7.3. The Bush interview	25	
Sentences as sequences	25	
Assigning information from codes to sequence variables	27	
8. <i>Working with keys</i>		30
8.1. The Bush interview revisited	30	
Adding a new text key	30	
Editing keys	31	
Automatically adding text keys	32	
8.2. Time keys	33	
Adding a new time key	33	
Other differences with text keys	34	
9. <i>Word links</i>		35
10. <i>Execution scripts</i>		36
10.1. Creating a script with the Command Editor	36	
10.2. Auto recording scripts	38	
10.3. Auto running saved scripts	38	
11. <i>Waveforms</i>		40
11.1. The waveform window	40	
11.2. Sound analyses	41	
Loudness analysis	41	
Silence analysis	41	
Pitch analysis	42	

## Introduction

SEQUENCE VIEWER is a complex program and if you are not acquainted with it, your main problem is quite probably just 'how to start'. The goal of the first part, 'Getting Started' is to solve this problem. It will take you about two hours. The  icon tells you what you are expected to do. When you're finished, you are ready to create your own data files and to understand more fully the meaning of the online help texts in SEQUENCE VIEWER.

The second part treats a number of different topics. Chapter 6 tells you in more detail how to create your own files and how to link audio or video files to your sequences. The next chapter treats how you can import files with sequential data. Probably you already have a file with sequential information. This chapter shows how many kinds of files can be imported, even in case they don't look like SEQUENCE VIEWER files. In Chapter 8 we will treat the keys in SEQUENCE VIEWER. Keys are especially useful for more qualitative analyses and for developing a coding scheme. The same is true for the word links, discussed in chapter 9. In chapter 10 we show you how you can record SEQUENCE VIEWER commands, to execute the same analysis more than once, for example on different data files. Chapter 11 is devoted to waveforms.

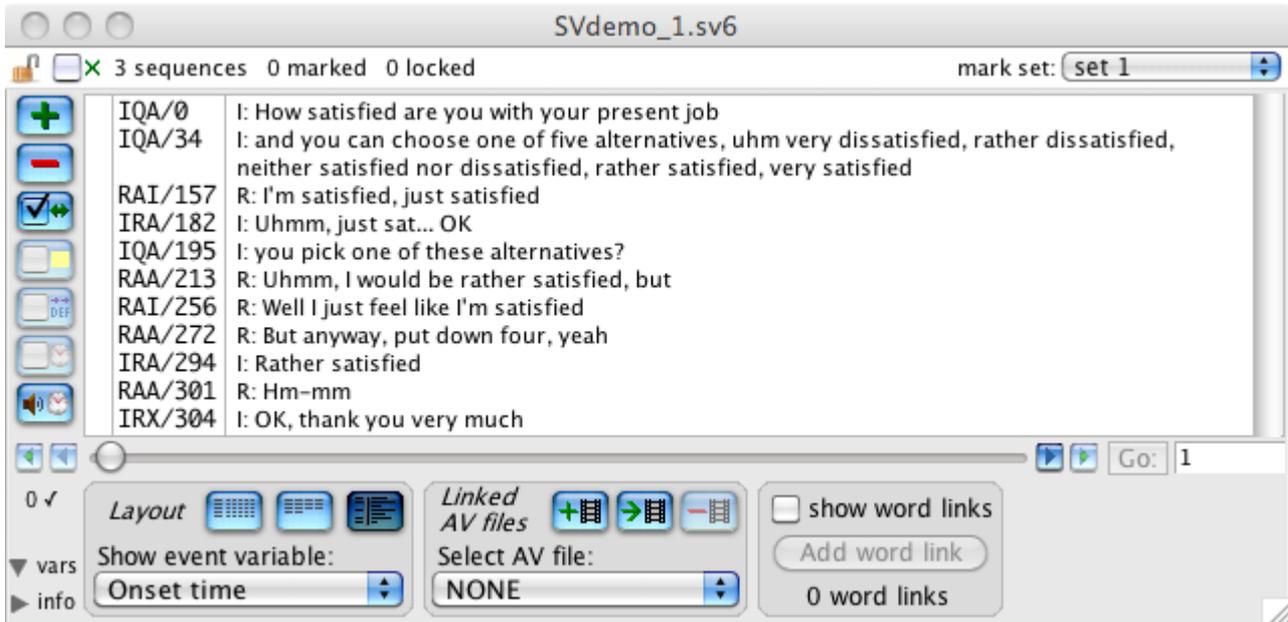
It is assumed that you already installed SEQUENCE VIEWER on your hard disk. If not, do this first. Then personalize SEQUENCE VIEWER (double click the program and fill in the empty fields). You should click somewhere on the opening screen that appears after you have personalized SEQUENCE VIEWER, to hide it.

# PART I: GETTING STARTED

## 1. The main sequence window

### 1.1. Sequence Viewer files

- ▶ Double click the SEQUENCE VIEWER icon, click the opening screen away and open the file 'SVdemo\_1.sv6' using **Open...** from the **File** menu (be sure to not open 'SVdemo\_2.sv6'). The file is located in the 'Tutorial folder'. A window, called the *main sequence window*, appears (see below).



If it looks differently:

- Be sure that the popup menu in the lower left corner shows 'Onset time'. If not, select this menu item.
- Be sure that the rightmost of the three *Layout* buttons  in the layout box is highlighted. If not, click this button. If there are only two buttons, or the window still looks different, someone apparently made significant changes to the demo file, making it unusable for this tutorial. Find the original demo file as came on the disk image, or [download](#) the Tutorial folder again from the SEQUENCE VIEWER website.

The *main sequence window* shows a large field with text and a smaller one with codes and numbers. All this information, the text and the codes, as well as a lot of other information, belongs to a *sequence*.

A SEQUENCE VIEWER file can consist of any number of sequences (unless the total size of the data file exceeds about 4 GB). The demo file only contains three sequences. These sequences concern transcribed Question-Answer sequences from a survey interview. To go to another sequence, use the horizontal scroll bar or the arrows on the left and right of it:



- ▶ Click  to go the next sequence and  to go to the previous one. Or move the slider  with the mouse to go immediately to a particular sequence; the number of the sequence is shown in the field to the right of the scroll bar. You can also type the number of a sequence in this field and click 'Go' (or press return)

### 1.2. Sequences and events

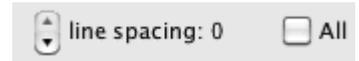
- ▶ Go back to the first sequence.

The field with the transcribed utterances of interviewer and respondent is called the *text field*. The text itself is called the *sequence text*. An utterance by the interviewer is preceded here by "I: ", and an utterance by the respondent by "R: ", just to make it easy to know who is speaking.

The transcripts (or whatever kind of texts you use) are divided into parts, each belonging to a different *event*. A sequence consists of events that are ordered in time or space. That means that one can say that one event in a sequence precedes or follows another event. Different events in the text field are divided by a carriage return. Hence, each paragraph in the text belongs to a different event.

### 1.3. Event texts and event codes

Such a text paragraph is called the *event text*. To make it easier to distinguish between the different event texts, you can use the *line spacing* stepper control located in the toolbar, in the upper left corner of the screen. If you cannot find the line spacing stepper control in the toolbar, probably this part of the toolbar is hidden. To show it, select **Edit** from the menu bar, and then **Toolbar > Show text layout**.



▶ Be sure the link button  is checked.

▶ Click a number of times at the down arrow of the line spacing stepper control. As you can observe, the line spacing between the event texts is increased.

The setting of the line spacing only applies to the present sequence. If you however check 'All', the setting is applied to all sequences in you file.

Left to the text field is a field with codes: the *code field*. A string of codes (before the slash) in the code field is called an *event code*. When you use the line spacing stepper control, each event code in the code field moves together with the event text: the event texts and the event codes are *linked*.

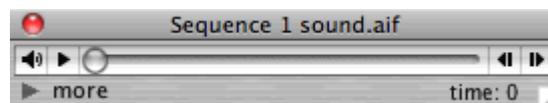
Thus, an event consists of an event text and an event code. Each event text (an utterance in this data file) is coded in a particular way. SEQUENCE VIEWER offers a lot of analyses on the sequences of coded events.

### 1.4. Onset time

Next to the event code in the code field, you see a slash and a number. This number represents the *onset time* of the event. That is, it represents the time the utterance starts, given the sound file that contained the original recorded interview. The onset time can be used to hear the sound of a particular event text.

▶ Hold the mouse down at the *Select AV file* popup menu (in the *Linked AV files* section of the main sequence window and select 'sequence 1 sound.aif'.

Probably a small dialog box appears telling you "Cannot find file 'sequence 1 sound.aif'". In that case, click at 'Look up file' in this dialog box, and select the file 'sequence 1 sound.aif' (it should be located in the 'tutorial files' folder; select 'Any file' as format). A Quick Time controller with the name of the file appears:



▶ Click at some event text in the text field. (You may note that the cursor changes into a loud speaker, when in the text field.)

Now the sound starts playing from the start of the event you clicked.

▶ Click once again to stop the sound.

You may wonder what the values of the onset time represent: what does the onset time of '34' of the second event mean? First, the meaning of '34' depends on the chosen *time unit*. In the present file this time unit is set to one tenth of a second; hence '34' means 3.4 seconds. Second, the onset time of an event refers to the very start of the *linked sound file* (the file 'sequence 1 sound.aif'). The onset time of the second event ("I: and you can choose...") is 3.4 seconds from the beginning of this sound file. The program uses the onset time of an event to play the proper part of the sound file if you click at the event text.

Settings like the time unit of a SEQUENCE VIEWER file and the reference of the onset time can be changed with **File settings...** from the **File** menu.

### 1.5. Link states

Of course clicking on an event text to hear its sound, only works if the event texts and the event codes with onset times are actually *linked*: each event text (utterance) belongs to one particular event code with onset time. Texts and codes will not be linked if, for example, the number of event codes does not equal to the number of event texts, that is, the number of paragraphs: an event text always consists of exactly one paragraph. Whether or not event texts and event codes of a sequence are linked, can be seen from the link button: 

▶ Click at the link button and observe that it changes to 

In addition, the vertical line separating the codes and the sequence text, is now coloured red; event codes and event texts are unlinked. This enables you to edit the text field.

▶ Delete the one but last sentence ("R: Hm-mm"). Click at the link button and it will change to  (note the red slash through the double arrow).

This means that text and code cannot be linked (because the number of paragraphs in the text field is does not equal the number of event codes). This is called a *bad link*. Bad links are important to identify, because analyses on both codes and texts at the same time, cannot be performed on sequences with bad links (or on sequences where text and codes are not linked at all).

Before you continue, you should first restore the original sequence.

- ▶ Click at the link button and to unlink event codes and event texts 
- ▶ Position the cursor after "I: Rather satisfied", press return and type "R: Hm-mm". Then click the link button once again, to change it to its linked state. Alternatively, you may select **Revert** from the **File** menu, to restore all original data at once.

### **1.6. More information**

For more information about the elements of the main sequence window, consult the *Reference SV6* document.

## 2. Event variables

The onset time is called an *event variable*: a variable that belongs to an event. Thus an event may have different types of information: event text, event code and event variables. Another event variable is the offset time of an event. 'Onset time' and 'Offset time' are special event variables that always occur together; you cannot have only one of them. They are jointly called the event time variables and have fixed names (you cannot change their names).

▶ Open the file 'SVdemo\_1.sv6' (if not open yet).

▶ To show the offset times, go to the *Show event variable* popup menu in the layout section of the main sequence window and select 'Offset time'.

The onset times after the slashes in the code field are now replaced by the offset times of the events. The offset time of an event does not necessarily have the same value as the onset time of the next event: there may be a pause between two utterances. You can also select 'NONE'; in that case no event variable is shown.

From the onset and offset times of each event, you can infer the duration of the utterance. SEQUENCE VIEWER can do this for you. Let's look how. First we will add a new event variable to the file.

### 2.1. Adding event variables

▶ Select **Add event variable...** from the **Event variables** menu.

The ADD EVENT VARIABLE setup window is opened. Don't click the **Add** button until you are told!

After 'Variable name:' you can type the name of the new variable.

▶ Change the variable name to 'duration'.

Variable names may be as long as 12 characters and spaces are allowed. Some characters are not allowed, for example a slash or a plus sign; you cannot type those characters. A variable name should not begin or end with a space and should not consist of digits only.

You have the choice between three types of event variables: numbers, dates, and time stamps. Dates and time stamps are actually numbers too, but have a special meaning. A 'Date' is the number of days since January 1, 1970, whereas a time stamp is the number of seconds, since January 1, 1970 midnight.

▶ Select 'Number'.

The number of decimal digits determines the precision of real (or floating point) variables. In this case, our 'duration' variable will be calculated by subtracting the onset time from the offset time, yielding whole numbers. However, these numbers are measured in units of one tenth of a second, and it may be more convenient to measure the duration in seconds, by dividing the difference by 10. That means that if the onset time is for example 84 and the offset time 126, the duration will become 4.2 seconds. Hence, if we want the duration in seconds, we need 1 decimal digit to prevent that we loose information.

▶ Select '1' with the popup menu after 'Decimal digits'.

You can safely set the minimum value to zero, as durations cannot be negative.

▶ Change to minimum value to 0.

The minimum and maximum values, are the lowest, respectively highest number that are possible in SEQUENCE VIEWER, given the number of decimal digits; these limits are different for different numbers of decimal digits. When you changed the decimal digits to 1, you probably observed that these lowest and highest possible values are automatically adjusted. Leave the maximum value as it is for the moment.

When the variable is created, all events obtain an initial value on the new variable. This value can be any number that fulfils the decimal digits (but you may type an initial value with less decimal digits) and the minimum and maximum values. The letter 'M' represents a missing value and in most cases this is the best initial value.

If the checkbox 'Keep open' is checked, the window remains open when you click the **Add** button. This is handy in case you want to add a number of event variables in succession.

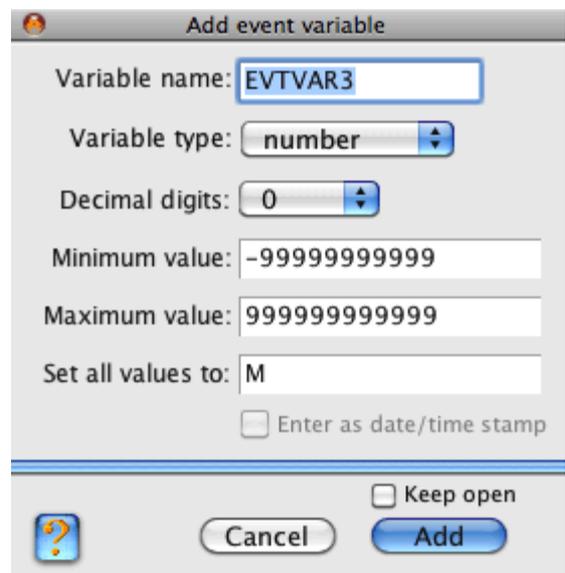
Don't click the **Add** button yet. First we will take a closer look at the ADD EVENT VARIABLE window.

### 2.2. Getting help

In the lower left corner, you see the *help* button: 

▶ Click the help button.

A window called SEQUENCE VIEWER HELP is opened with information about adding event variables. In the left field you see



a list of topics. You can click on a name to go to that topic. The help text also contain links to other parts of the help file. Links look like: [This is a link](#). In the field at the bottom of the Sequence Viewer help window, you can type words you wish to find.

Click the **Close** button in lower left corner of the help window (to prevent that your screen becomes cluttered with windows).

Move the mouse over the fields in the ADD EVENT VARIABLE WINDOW'.

You will see short descriptions of these fields in the help field of the toolbar (located in the upper left corner of the screen). If you move the mouse over the help button, a short description of the analysis or procedure at hand is given.

Moving the mouse over the help button, also has another function, you will especially appreciate with more complex setup windows. To illustrate this function, do the following.

In the field after 'Variable name', you typed 'duration'. Now delete this name (empty the field). You may observe that the **Add** button becomes disabled. If the setup in a setup window is not correct, the execution button (the **Add** button in this case) will be disabled.

Move the mouse over the help button. The help field in the toolbar now tells you (in red) that you should enter a name for the new variable.

Now type "SEQSIZE" as the name for the new variable. When you move the mouse over the help button, the help field now informs you that this is an illegal name.

Especially with complex setups, it will often not immediately be clear what's wrong with the setup (why the execution button is disabled). Moving the mouse over the help button will tell you the cause.

Type 'duration' again as the name for the new variable to obtain a correct setup.

### 2.3. Computing event variables

Now we can actually add the event variable to our file.

Click the **Add** button.

The output window gives you an overview of your changes. In this way you can always keep track of your changes to the data file. You can safely click the output window away with the close box; its content is not lost.

In the main sequence window you can check that the new variable is actually added:

Select 'duration' with the *Show event variable* popup menu in the layout section.

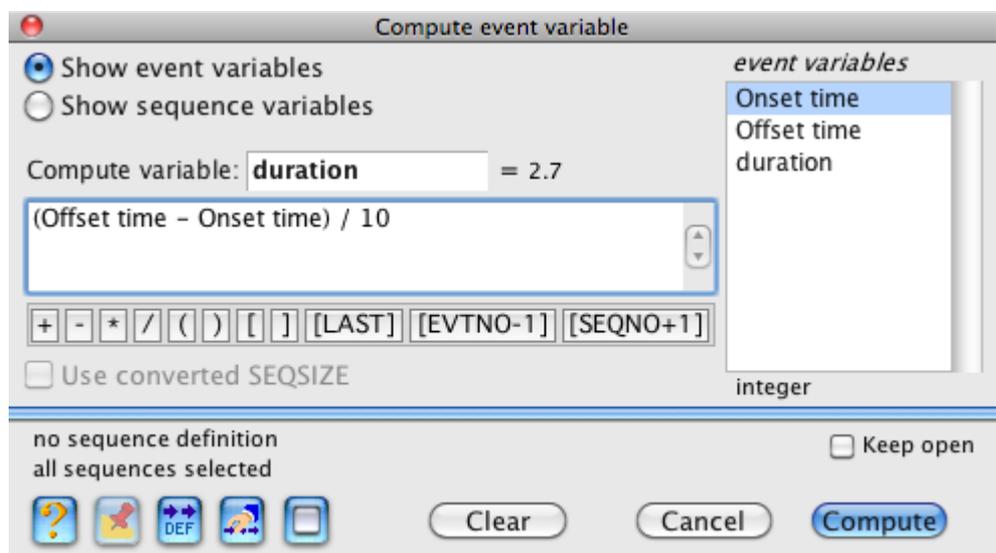
After the slashes in the code field you will see M's (for missing values). You may also observe that the code field has become larger and the text field smaller. The reason is that space is provided to account for the maximum value of 999999999.9 on 'duration'.

Now we want to assign the duration of each event to the new event variable. The duration is defined as the difference between the offset time of an event and the onset time of that event.

Select **Compute event variable...** from the **Event variables** menu to open the COMPUTE EVENT VARIABLE setup window. Be sure the 'Show event variables' radio button (in the upper left corner of the window) is selected. Click at the field after 'Compute variable' to select this field (it will be surrounded by a blue border; it is 'focused'; the colour may depend on your settings) and then click at 'duration' in the list with names of event variables. The name of the variable appears in the field after 'Compute variable'. You can also type the name of the variable in the field after 'Compute variable'; after typing the 'd', the program will try to complete the name.

Click at the large field below 'Compute variable' to select this field.

Here we must enter the calculation for the duration. It should look like in the screen shot at the right. Instead of typing variables names like 'Offset time', it is



faster to click on a name in the list with variable names (you can also click at the buttons like  instead of typing a minus sign).

- ▶ Click the **Compute** button.
- ▶ In the main sequence window, click the 'codes and event variables'  button in the layout section, and check that the values represent the difference between offset and onset times, divided by ten.
- ▶ Pull down the 'Edit' menu of the menu bar.

You will see a menu item 'Undo compute event variable'. If you select this menu item, the analysis will be undone and all values of 'duration' will be reset to their values before the analysis; in this case 'M' for missing value.

- ▶ Select this menu item and check that this is actually the case.

Not only are the missing values restored, but also the layout mode is reset to the mode at the moment you performed the 'Compute event variable' analysis. Actually, in case of undoing an analysis, everything is reset to the situation at the moment the command that is undone, was executed. So any changes to the file you made after a command, are lost too, when that command is undone.

You now have to perform the compute event analysis again (no 'redo' option is available), but as we will see, this can be done very quickly.

- ▶ Open the COMPUTE EVENT VARIABLE window again (select **Compute event variable...** from the **Event variables** menu).
- ▶ To the right of the help button you see the 'previous setup' button: . Click this button and the window shows the previous setup.
- ▶ Click the **Compute** button.

You're done. The 'previous setup' button is available in most windows, and can be of great help in case of minor changes to a setup you performed earlier.

#### 2.4. Other properties of event variables

Except the number of decimal digits and maximum and minimum values, more information can be stored with an event variable.

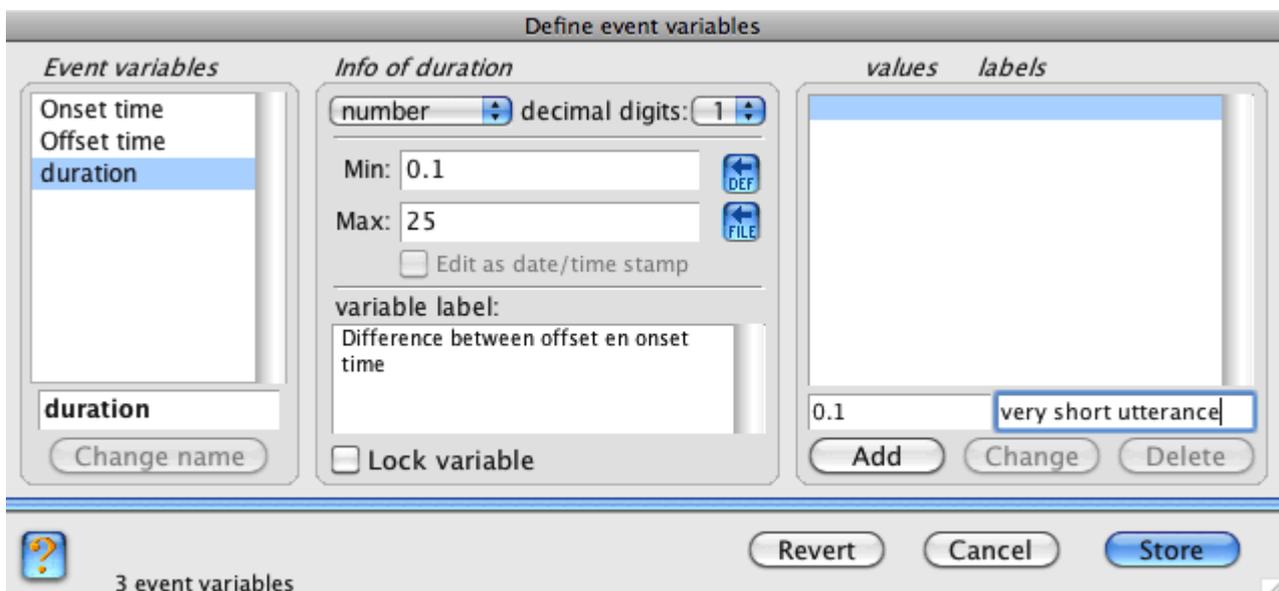
- ▶ Select **Define event variables...** from the **Event variables** menu. The DEFINE EVENT VARIABLES window is opened.
- ▶ Click at 'duration' in the field with the names of event variables.

First we are going to change the maximum value of 'duration'. We can use as a maximum value the actual maximum as appears in the file.

- ▶ Click at the 'File'  button (located to the right of the maximum value field, after 'Max:').

Both the actual minimum (0.3) and maximum (11.9) are put into their respective fields. You can change these values, in case you want to add new sequences with durations that may exceed these limits. For example, you may set the minimum value to 0.1 (as a duration of 0 does not make much sense and is quite unlikely). A reasonable maximum for duration might be 25.

- ▶ Change the minimum and maximum values to 0.1 and 25.



Further you can add a variable label. For example, you may simply type "Difference between offset and onset time" in the field below 'variable label:'.

You can also change the name of the variable, if you wish. After you have changed the name, you should click the  button; SEQUENCE VIEWER has to check whether or not such a name already exists, or whether the new name is a legal one, before you can change characteristics of other event variables.

Finally you can check the 'Lock variable' check box. A locked variable cannot be changed anymore (until you unlock it again by unchecking the check box). If you are sure that a particular event variable will not be changed, it prevents that you change it inadvertently.

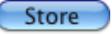
In the right part of the DEFINE EVENT VARIABLES window, you can add value labels to an event variable.

 Type the text as is shown in the picture at the previous page and click the  button.

The value and its label now appear in the upper field. You can continue to add value labels (although this does not make much sense for this variable, as its values are self explaining). To delete a value label, click on it and click the  button.

To change the properties of other event variables, click at the name of an event variable and change the properties. If you change the name of an event variable, be sure to click 'Change name'. You cannot change all characteristics of the onset and offset time variables: their names, decimal digits and variable labels are fixed.

When you are done with making changes to the event variables, you should click the  button to store the changes.

 Click the  button.

As soon as you save the file, all changes are stored within the file.

 Back in the main sequence window, select 'duration' with the *Show event variable* popup menu in the layout section.

You may note that the space between the event codes and the text field is much smaller now. SEQUENCE VIEWER adjusts this space according to maximum length of values, given the user defined minimum and maximum values.

## 3. Code variables

### 3.1. About coding schemes

Left to the text field in the main sequence window you will see the code field (with or without one of the event variables). Event codes play a very important role in SEQUENCE VIEWER. Analyses can be performed on event codes to find dependencies between successive event codes, to discover systematic patterns in the whole sequence of event codes, or to find similarities between different sequences of event codes.

In our example of a question posed by an interviewer to a respondent, one could discern between a number of different kinds of utterances, for example:

- A: Interviewer poses question as scripted
- B: Interviewer changes wording of question
- C: Respondent answers adequately (selects one of the response alternatives)
- D: Respondent answers inadequately
- E: Respondent asks for repetition
- F: Interviewer repeats answer

Essentially, a coding scheme consists of a number of codes (like 'A', 'B', or 'C' in the example above) and a description of the kind of text that determines when a particular code is applicable. In the ideal case, a coding scheme is exhaustive (for each kind of text a code should be available) and mutually exclusive (only one code should be applicable).

In developing and applying coding schemes, the number of codes tends to grow; researchers want to make more and more finer distinctions between utterances. Generally, the more codes a coding scheme has, the more information is retained, the more difficult (and less reliable) coding will be, and the more difficult it becomes to analyze the codes in a meaningful way.

SEQUENCE VIEWER offers you a way to distinguish between a very large number of different codes on the one hand, but to perform analyses on limited subsets of these codes in a very systematic way, on the other hand. This is done by using so-called *code variables*: each piece of the sequence text is coded on a number of different aspects, each representing a different *code variable*.

In the example file SVdemo\_1.sv6 we use three such code variables. The first one is simply the speaker (called 'actor') and can be 'I' or 'R', for interviewer and respondent, respectively. The second one is called 'exchange' and can be 'Q' (for question), 'A' (for answer) and R (for other reactions). Finally, the code variable 'adequacy' discerns between 'A' (adequate), 'I' (inadequate) and 'X' (adequacy not applicable). This, of course is only a very rudimentary coding scheme, and only used for illustrative purposes. When you develop your own coding scheme, you will quite probably use more code variables, and you should define codes more carefully (like what exactly is 'adequate' and 'inadequate').

Each code variable corresponds to a particular position in the fully coded event (or *event code*). For example, event code 'IQA' means that the interviewer ('I') poses a question ('Q') as worded in the questionnaire (hence it is adequate, or 'A'). 'RAI' means that the respondent ('R') answers a question ('A'), but does not use one of the response alternatives ('I' for inadequate).

The advantage of this approach is that quite a lot of unusual utterances can be simply coded, without adding new codes to the coding scheme. For example, if the respondent poses a question to the interviewer, it might be coded as 'RQX'. Moreover, if the researcher for some particular analysis is not interested in whether an utterance is adequate or not, one can very easily perform an analysis using only the first two code variables. Further, it is possible to investigate the relations between the different code variables, that is, the different aspects of a coded event. For example, who poses relatively more questions, the interviewer or the respondent? Although a trivial research question for the present kind of data (an interview), it is not for other types of data; for example if one analyses group discussions. There are a lot of other advantages to this approach (for example, it is easy to add a new code variable and use it for recoding, without affecting the original codes).

Of course you can confine yourself to only one code variable, especially if your data only need a limited number of different categories to distinguish meaningfully between different pieces of the event texts. On the other hand, in analyzing complex interactions (like an interview or group discussions) you may need both more code variables and more categories for each code variables than in our simple scheme. The maximum number of code variables is nine.

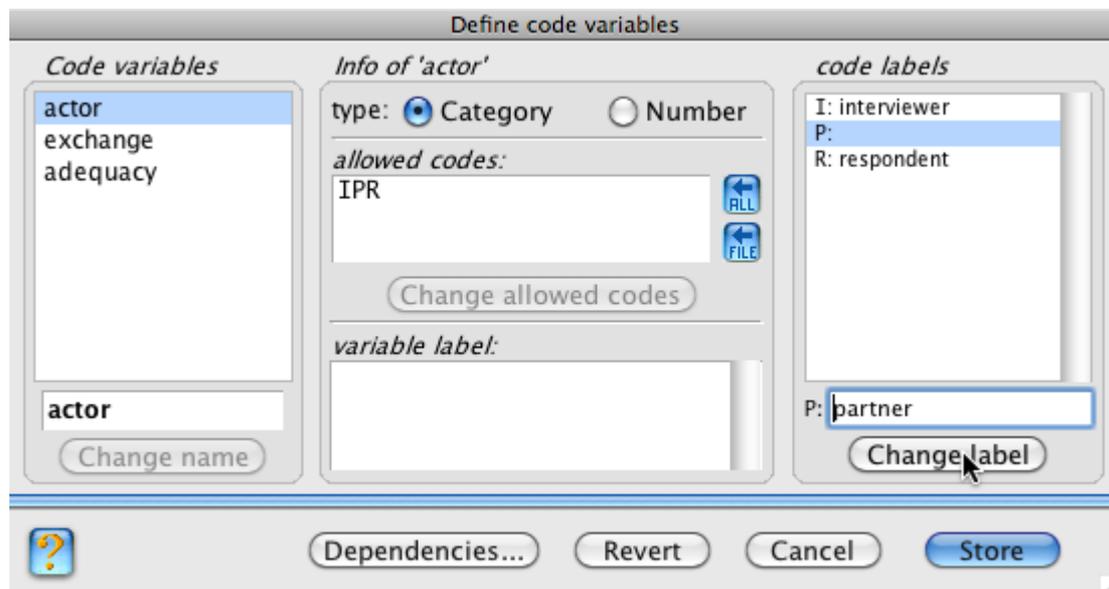
### 3.2. Defining code variables

Like event variables, you can add information to code variables, change their names and add a variable label. Moreover you can define which codes for a particular code variable are allowed. Let's illustrate how it works.

 First open the file SVdemo\_1.sv6'.

Suppose you observed that it often occurs that the partner of the respondent becomes involved in the interview. Let's add a code to the code variable 'actor'.

 Select **Define event codes...** from the **Codes** menu.



The DEFINE CODE VARIABLES setup window appears. Be sure the code variable 'actor' is selected. Now we have to type a character signifying the partner into the field with allowed codes ('I' and 'R' are already there). You can type any character between 'a' and 'z', 'A' and 'Z' and '0' and '9'. Codes are case sensitive; 'R' is different from 'r'. To identify a partner, 'P' seems the most appropriate character here.

- ▶ Type a 'P' after 'IR' in the 'allowed codes' field.
- ▶ Click **Change allowed codes**.

You may note that codes are automatically alphabetized (so in case of many codes, it is easy to find one). You may also try to type an already existing code, e.g. 'I'; it will not be accepted.

In the right field, with 'code labels', the code 'P' is added. You can add a code label to it, for example 'partner'.

- ▶ Click at the 'P' in this field and type "partner" in the field just above 'Change label'.
- ▶ Click the **Change label** button.
- ▶ Click **Store** to accept all changes.

### 3.3. Analysing event codes

A simple analysis is just a frequency distribution: how many different event codes appear in your file.

- ▶ Select **Frequencies...** from the **Codes** menu.

The FREQUENCIES OF EVENT CODES setup window is shown. Don't make any changes to the setup for now.

- ▶ Click **OK**.

The output window shows a frequency distribution of all event codes in your file.

- ▶ Open the FREQUENCIES OF EVENT CODES setup window once again.

The window has a number of options that are similar to other windows. We will refine our analysis, and at the same time show these common options.

First you may want to have a frequency distribution of the coded respondent utterances only.

- ▶ Click at the first question mark after 'enter:' and type 'R' (if you type a lower case 'r', it will be automatically converted to 'R', because 'r' does not belong to the allowed codes). Instead of typing a 'R', you can also click at **R** in the field below 'allowed codes:'.

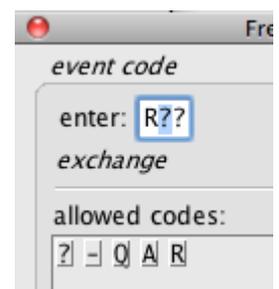
This part of the setup window should now look like this:

- ▶ Click **OK**.

The frequency distribution now only concerns event codes starting with 'R'. A question mark in the field after 'code:' means 'any code'. Thus 'R??' in the FREQUENCIES OF EVENT CODES window means all event codes with 'R' at the first position, that is, with code 'R' on code variable 'actor'.

A more interesting analysis would be to compare interviewer and respondent utterances.

- ▶ Open the FREQUENCIES OF EVENT CODES window once again. Select 'code variable' with the popup menu below 'by variable:'.



A list appears with the names of the code variables.

- ▶ Select 'actor'; be sure that the field after 'enter:' contains only question marks!
- ▶ Click .

You obtain a cross table, showing that the codes after the first code (the 'actor') are very different for both actors. As could be expected, interviewers tend to pose questions (?QA and ?QI), whereas respondents tend to provide answers (?RA and ?RI). To make this difference even more clear, we can apply a *sequence definition*.

### 3.4. Sequence definition

- ▶ Once again, open the FREQUENCIES OF EVENT CODES window.  
Locate the 'apply sequence definition' button, at the bottom of the window: .

- ▶ Hold the option-key down and click the 'apply sequence definition' button (alternatively, you can select **Define sequence...** from the **Sequences** menu).

A new window, the SEQUENCE DEFINITION setup window appears. No sequence definition is selected yet, as is shown by the popup menu after 'select:'.

- ▶ Select 'Sequence definition 1' with this popup menu.

Because (for this example) we are not interested in whether questions and answers of interviewer and respondents are adequate or not, we want to neglect the third code variable, 'adequacy'.

- ▶ Check 'Mask events'.

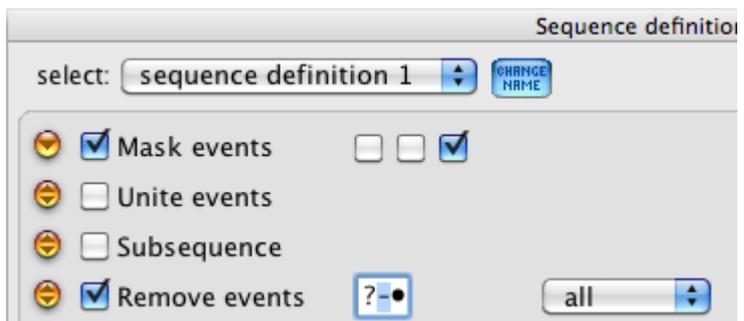
Three checkboxes appear, each corresponding to a code variable.

- ▶ Check the third one. The third code variable, 'adequacy', becomes 'masked'.

From the output of the event code frequencies we can see that the event code 'I--' occurs five times and the event code 'R--' two times. The code '-' is the general code for 'not coded yet'. Hence it makes sense to get rid of these event codes for our analysis.

- ▶ Check 'Remove events'. A field with '??●' and a popup menu appear. The '●' means that the code variable at this position is masked.

- ▶ Change the second position to '-'. You can type a dash, or click the dash in the field below 'allowed codes'. The relevant part of the window should now look like the screen shot at the right.



What does this mean? First, that the codes on the third code variable are neglected, second, that all event codes with '-' at the second position are removed from the sequence for the analysis.

- ▶ Click  to activate the sequence definition (leave the SEQUENCE DEFINITION setup window open).

We can look at the effect of such a sequence definition on our sequences.

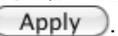
- ▶ Go to the second sequence in the main sequence window.
- ▶ Click at the first of the three layout buttons: . This changes the mode of the main sequence window to *codes and event variables*. The event codes with all event variables, but without event texts are shown now.
- ▶ Check the 'show sequence definition' button  at the left of the code field in the main sequence window.

The field now shows the effect of the sequence definition on the sequence. Leave the window as it is and go back to the SEQUENCE DEFINITION window to apply one more sequence transformation.

- ▶ Check 'Unite events' to take similar succeeding event codes together (neglect the popup menu that appears, for the moment).

- ▶ To observe this effect on sequence number 2, first click the  button in the SEQUENCE DEFINITION window.

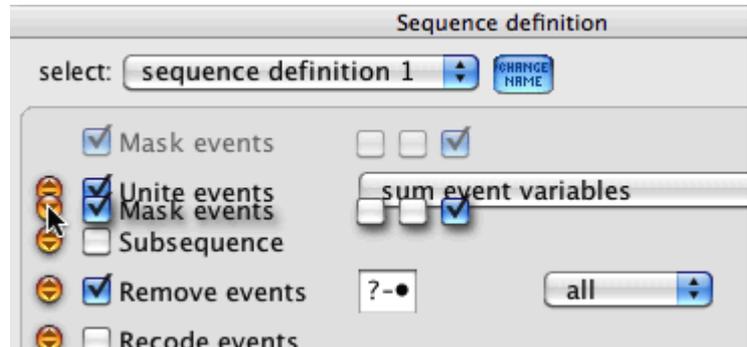
It can easily be seen that the first two event codes (that are similar because the third code variable is neglected) are taken together. You can also observe that the onset time of the 'united' event is 0, being the onset time of the original first event. The offset time of the united event becomes 120: the offset time of the original second event. Which is as it should be, of course. However, this does not hold for our 'duration' variable. The duration of the united event equals the duration of the original first event. We can partially correct this.

- ▶ In the SEQUENCE DEFINITION setup window, select 'sum event variables' with the popup menu after 'Unite events' and click .

First you should observe that this does not affect the offset and onset times, because these two variables have strictly defined meanings in SEQUENCE VIEWER. It does affect the 'duration' variable however (and all other event variables, if available). The durations of both original events are summed, yielding 11.5. Unfortunately, this is not quite the difference between the offset and the onset time of the 'united' event. The reason is that there is a delay of 0.5 seconds between the offset time of the first original event and the onset time of the second one (see note 1 at the end of section 3.4 for a solution).

The original event codes of the first two events were 'IQA' and 'IQI'. After masking they changed to 'IQ●' and 'IQ●', which is why they are similar. Apparently, the event codes are *first* masked and *then* united. We can easily change the order of the transformations.

- Go back to the SEQUENCE DEFINITION setup window.
- Hold the mouse down at the  icon (before the 'Mask events' checkbox) and move the mouse down. The whole 'Mask' transformation line moves down, as a 'disabled' line, leaving a 'copy' behind.
- Leave the mouse as soon as the transformation line is just below the 'Unite events' transformation line, and becomes 'enabled' (see screen shot at the right).



The transformation lines 'Mask events' and 'Unite events' are now reversed, and so is the order of the transformations. Hence, if this sequence definition is applied, the first two events of sequence 2 will *not* be united as they are still different. It will be clear that the order of the transformation can strongly affect how the transformed sequence will look like.

Back to our frequencies of event codes example.

- Reverse the order of the transformations again (first mask, then unite).
- Click **Done** in the DEFINE SEQUENCE window to activate the sequence definition and close the window.
- Click at the 'apply sequence definition' button in the FREQUENCIES OF EVENT CODES setup window. The arrows become blue again and no sequence definition will be applied.
- Click at the 'apply sequence definition' button. The button changes a bit: the arrows become red. This tells you that the selected sequence definition will be applied for the analysis. The arrows become blue again and no sequence definition will be applied. The text above the row of buttons tells you the name of the sequence definition.

You may also observe that in the FREQUENCIES OF EVENT CODES setup window, the code for the third code variable is changed to a ● to tell you that this code variable is masked.

- Click at the 'apply sequence definition' button once again. The arrows become blue again and no sequence definition will be applied.

In this way, you can easily switch between applying the default sequence definition or not.

- Click the button once again to apply the sequence definition for our analysis.
- Click **OK**.

The cross table now clearly shows the differences between interviewer and respondent utterances with respect to the exchange code variable, irrespective the adequacy of the utterance .

#### note 1

The problem can be solved by computing the duration in the same way as in the example above, but in addition applying the sequence definition in the COMPUTE EVENT VARIABLE window. Add another event variable ('duration2') and open the COMPUTE EVENT VARIABLE window. Click the 'previous setup'  button and change the 'compute variable' to 'duration2'. Click at the 'apply sequence definition' button to activate it. Click **Compute**. The correct duration of the united events is now assigned to the *first* original event. For example, in sequence 2, the duration is calculated as the difference between the offset time of the first two 'joint' (united) events (120) and the onset time of these joint events (0). Divided by 10 yields 12.0 in stead of 11.5. In the SEQUENCE DEFINITION setup window, select 'take event variables of first event' from the popup menu after 'Unite events' and click **Apply**. The duration of the 'united' event now has the correct duration.

#### note 2

The *codes and event variables* mode is also suited to manually change the values of event variables. Just click at a value and edit the value.

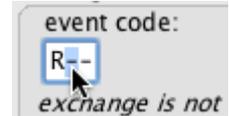
### 3.5. Changing codes

Be sure the main sequence window is in the *codes and text* mode: click at the third one of the layout buttons. The second sequence has four event codes that are not fully coded, as appears from the '-'. The code field cannot be edited directly. To code the events, we should use the EDIT SEQUENCE window.

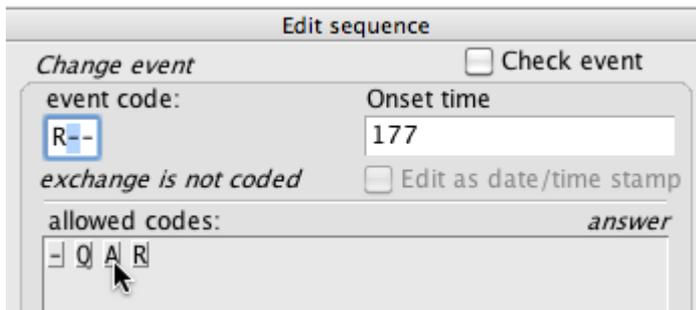
▶ Select **Edit sequence...** from the **Edit** menu.

The first event code that is not fully coded ('R--') becomes automatically selected (this depends on the file settings; if this event code does not become selected because the file settings are changed, click on event code 'R--' in the main sequence window). In the EDIT SEQUENCE window the selected event code is shown below 'event code:'.

▶ Click at the first '-' in the EDIT SEQUENCE window to select this code, if it is not selected yet.



You can now either type the correct code, or click at a code in the field below 'allowed codes'. We will use the latter method.



▶ Move the mouse over the codes in this field.

You will see the label of the code, to assist you in selecting the correct one ('answer' in the screen shot). The utterance ("Ehm, rather satisfied") is clearly an answer, hence the code 'A' is the correct one.

▶ Click at the 'A'.

Code 'A' replaces the '-'. The next '-' becomes automatically selected. The correct code for the third code variable is 'A' because the answer is adequate.

▶ Again click at 'A' (or type 'A').

Because this is the last code variable of the event, the event is now fully coded. The codes automatically replace the 'R- -' in the code field of the main sequence window (provided that 'Replace if last code of event is typed' is checked in the 'change event codes' box). The next event code becomes automatically selected (if 'Move to next event after replace' is checked). If not, you should click at the event code you wish to edit in the main sequence window.

#### *note*

Behaviour like automatically selecting an event code that is not yet fully coded, is governed by the **File settings** (the **Coding** tab) and can be different for different files. File settings are stored within your file when you save the file.

## 4. Extensions of the main sequence window

### 4.1. Sequence variables

A sequence consists of an ordered series of events. Events may consist of event texts, event codes and event variables. Sequences themselves may also have characteristics. For example, in the demo file the number of the question in the questionnaire, the number of the interviewer and the number of the respondent are characteristics of the whole sequence. A general characteristic of a sequence is its size, that is the number of event codes. The first sequence, for example, has 11 event codes. Such information about the sequence is stored in so-called *sequence variables*.

- ▶ Open the file 'SVdemo\_1.sv6'. The values of sequences variables can be seen in a drawer at the left part of the main sequence window (or the right part if there is not enough room at the left side).
- ▶ Click at the ▼ triangle in the lower left corner (left to 'vars') to show the sequence variables drawer. The triangle will change to ◀.

In the drawer you can see the sequence variables (if you inadvertently click on a name, a small window appears, allowing you to change values; click it away with the close box or the **Cancel** button).

The first sequence variable is always the variable SEQSIZE, the number of event codes of a sequence. It can only be changed by adding or deleting events from a sequence. The second sequence variable is always a variable called STARTTIME. This is the time the sequence (the first event) starts, given the linked sound or video file. This variable is always available, also if there are no linked files at all. The other four sequence variables are self-defined variables, with number of interviewer, respondent and question, and the eventual score as entered by the interviewer.

You can add as many sequence variables as you want. You can change their properties like variable labels, value labels, etc. in very much the same way as with event variables, so this will not be explained here (you can use **Define variables...** from **Sequence variables** menu, or click the **Define** button in the sequence variables drawer). Like with event variables, 'M' means a missing value.

Like event variables, there are three types of sequence variables: number, date and time stamp. Dates are stored as the number of days since January 1, 1970 and time stamps as the number of seconds since January 1, 1970, midnight. The advantage of the date (and time stamp) type is that you can enter dates in the form of 'Jan 7, 2005' (and time stamps as 'Jan 12, 2013 08:32:14'). In calculations dates and time stamps are just treated as an integer. So you can easily calculate the difference between two dates.

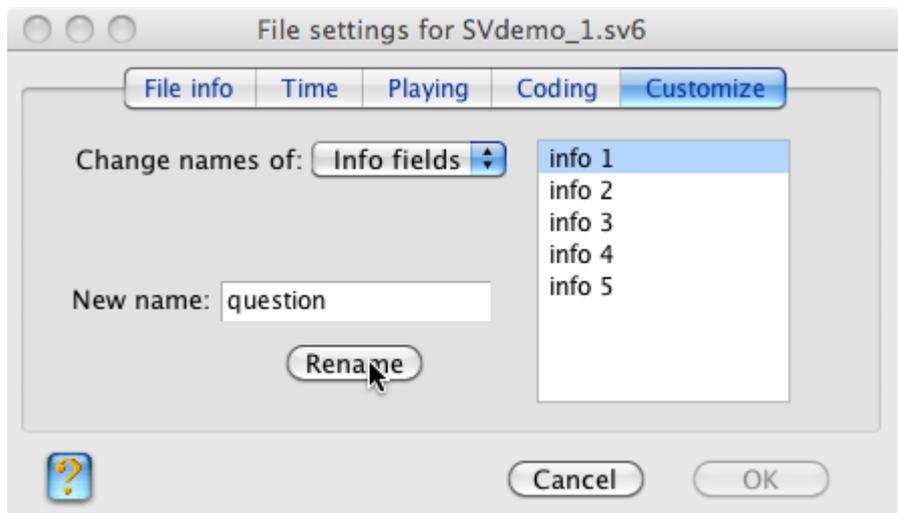
You can edit values of sequence variables manually, by clicking on a value, and changing the value. To hide the sequence variables drawer, click the ◀ triangle.

### 4.2. Information fields

The main sequence window has five information fields available, where you can enter any kind of information pertaining to the sequence at hand. To show these fields, click at the ▶ triangle (left to 'info'). In the demo file the first field is used for the scripted question from the questionnaire and the response alternatives. The different fields become available by clicking at one of the tabs labelled 'info 1' to 'info 5'.

You can change the default names of the five information fields to more meaningful names. Let's change the name of 'info 1' to 'question'.

- ▶ Select **File settings...** from the **File** menu and click the **Customize** tab.
- ▶ Select **Info fields** from the popup menu after 'Change names of:'. Be sure 'info 1' is selected in the field with the names of the five information fields.
- ▶ Type 'question' in the field after 'New name:'. The **Rename** button becomes enabled (because the new name is different from the old one).
- ▶ Click the **Rename** button. In the names field, 'info 1' is replaced by 'question', and the **OK** button becomes enabled.
- ▶ Click the **OK** button. The name of the first information field is changed to 'question'.



You can change the names of mark sets in a similar way. Customized names are stored together with your file, as soon as you save the file.

## 5. Adding new data

### 5.1. Adding a new sequence

Usually SEQUENCE VIEWER files have many sequences. We will show how to add a new sequence to the demo file and how to 'fill' it with information.

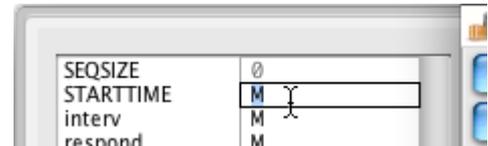
- ▶ Open the file SVdemo\_1.sv6'.
- ▶ Go to the last (third) sequence.
- ▶ Click the 'add sequence' button  located in the upper left corner of the main sequence window. A new sequence is added after sequence 3.

### 5.2. Linking a sound file

In the demo file each sequence has its own linked file; the sound file linked to sequence 1, is another file than the one linked to sequence 2. In such a case, the start of the sequence (text), coincides with the beginning of the sound file. That need not be the case. For example, a sound file may consist of a whole interview, and a sequence concerns the utterances belonging to a particular question. In such a case the start of a sequence does not coincide with the beginning of the sound file. Instead, the start of a sequence coincides with a particular time location on the sound file. This is called the STARTTIME of a sequence. The STARTTIME is a sequence variable, and simply indicates how many time units from the beginning of the sound file, a particular sequence starts. In the example of an interview, this may be the moment the interviewer poses a new question from the questionnaire.

Because in the demo file each sequence has its own sound file, the STARTTIME of each sequence should be 0 (zero).

- ▶ Open the Sequence variables drawer (see section 4.1) if it is not open.
- ▶ Click at the 'M' (for missing value) after STARTTIME in the sequence variables drawer. The 'M' becomes surrounded by a rectangle. Change the 'M' to '0' (zero).



- ▶ Click the 'link new AV file' button  in the 'linked AV files' box.
- ▶ A common 'Open file' dialog box appears. Select the file 'Sequence 4 sound.aif' in the 'tutorial files' folder.

A window titled 'Sequence 4 sound.aif' with a QuickTime controller is opened. This window is called the PLAYER window. We will use this sound file to transcribe the question-answer sequence.

### 5.3. Transcribing text, using a sound file

- ▶ Click at the triangle  in the lower left corner of the PLAYER window (before 'more') to extend the window.

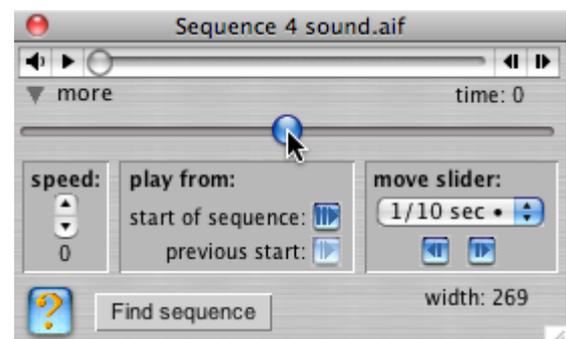
Click at  in the QuickTime controller (the uppermost slider) to play the sound. It will change to . Click  to stop the sound after a little while and type the text in the text field of the main sequence window. Continue transcribing in this way. Precede an interviewer utterance with "I: " and a respondent utterance with "R: ". Although not required by SEQUENCE VIEWER, we will see that this is quite useful. Type a carriage return after each utterance.

Quite probably you want to move the play head a bit back during transcribing. There are a number of options to fine tune the position of the play head.

First, you can use the precision slider to position the play head within a range between -25 and +25 time units (that is minus or plus 2.5 seconds for this file) from the present position of the play head (see screen shot at the right).

Second, and probably even more useful, you can use the , , and  and  buttons. Here is what they mean.

-  Moves the play head to the beginning of the sequence and starts playing from there.
-  Moves the play head to the position the last time you started playing. If the linked file is already playing, it stops the sound or movie. In this way you can repeatedly hear exactly the same fragment.
-  and  Moves the play head a number of time units back, respectively forward. How many time units depends on the popup menu below 'move slider'. Here too, if the sound is already playing, it stops.
- ▶ Try these buttons and watch the time counter in the upper right corner of the PLAYER window.
- ▶ Complete the transcription.



#### 5.4. Coding event texts

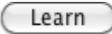
SEQUENCE VIEWER can assist you in the often tedious task of coding. To show you how, we first go back to sequence number 3.

➤ Go back to sequence 3.

The first event text reads "I: How satisfied are you with your present job", which is coded as 'IQA'.

➤ Shift-click at 'IQA'.

This opens the EDIT SEQUENCE window, whereas 'IQA' becomes automatically selected.

➤ In the EDIT SEQUENCE window, click at the  button (in the 'Suggestions' box).

The program has now 'learned' that an utterance like "I: How satisfied are you with your present job" will quite probably be coded as 'IQA'.

➤ Go to sequence 4 again.

Be sure that the different utterances are preceded by "I: " or "R: " and that after each utterance a carriage return is typed. The first event text should read something like: "I: How satisfied are you with your present job".

➤ Click at the  button in the EDIT SEQUENCE window (below the blue separator line, to the right of the 'Help' button).

The code field is now filled with event codes, one for each event text (paragraph). The very first code of each event code corresponds to the first character of the event text. These characters should be 'I' or 'R', provided that you preceded each utterance with 'I' or 'R' for interviewer and respondent. The other codes consist of '-' for 'not coded yet'. The first event code is selected and appears also in the edit field below 'Code:' in the EDIT SEQUENCE WINDOW.

➤ Click at the  button in the EDIT SEQUENCE window.

The edit field in the EDIT SEQUENCE window should now contain 'IQA'. Because the program learned that the event string "I: How satisfied are you with your present job" is associated with 'IQA', this code is suggested. At the bottom of the window you can see a measure indicating how much the event text to be coded is similar to the learned text string.

➤ Click at the  button in the 'Change event codes' box of the EDIT SEQUENCE window.

'IQA' replaces the 'I--' in the code field of the main sequence window. The next event code becomes selected. You should enter the codes manually as you earlier did in section 'Changing codes' in Chapter 4, because no more suggestions are available. Of course you can add as many suggestions as you wish. Suggestions (associations between text strings and codes) are stored within your file as soon as you save the file.

Of course, suggestions only make sense if your sequences contain a lot of similar event texts that should obtain the same event code. In such cases the 'suggestion' procedure can considerably speed up coding. The event text and the learned text string need not be exactly the same in order to suggest a code. For more information about 'suggestions', see the help file in SEQUENCE VIEWER (see 'Suggestions' of section 'Edit sequences' of chapter 'Edit').

➤ Continue coding. Close the EDIT SEQUENCE window when you are done.

#### 5.5. Adding event times

There are a number of ways to add event times (onset and offset times) to each event. The procedure described here shows you how it can be done relatively fast and easy. Be sure that you are still at sequence 4 and the 'sequence 4 sound.aif' file is open.

➤ Select 'Offset time' from the 'event variables' popup menu in the layout box.

➤ Click at the 'assign event time' button  to check it.

➤ Click at the first event text to play the sound. Don't click in the code field!. Be sure you have changed the STARTTIME to 0 (see section 5.2), else the sound will not play.

➤ At the moment the utterance ends, click the event once again to stop the sound.

➤ Click at the corresponding event code in the code field (the cursor changes into a clock as soon as you enter the code field).

The offset time automatically replaces the 'M after the slash. Don't worry if the offset time is not exact yet; leave it as it is.

➤ Click at the second event text and repeat the procedure.

➤ Continue until you are done.

➤ Press the caps lock button.

➤ Click at the first event

The caps lock button causes that only the last 10 time units of the event are played, whereas the sound stops at the moment the offset time is reached. In this way it is easy to hear if the offset time should be adjusted.

- ▶ Press the arrow up key to decrease the offset time with one time unit, or the arrow down button to increase it with one time unit, depending on whether the offset time should be earlier or later.
- ▶ Click once again on the first event text (still with the caps lock down), to check if the sound stops exactly at the moment the interviewer stops speaking.
- ▶ Continue until you are done.
- ▶ Unlock the caps lock button.
- ▶ **Uncheck the 'assign event time' button.** This is important, in order to prevent that you inadvertently click at the code field, which will change the offset time.

To enter the onset times, you can use a similar procedure. It may save time however, to first assign the offset time of each event to the onset time of each successive event. This procedure is described in detail in the help file of SEQUENCE VIEWER (see section 'Edit event times' of chapter 'General').

### 5.6. Changing sequence variables

The values of the sequence variables of our new sequence have the value 'M', except SEQSIZE (which was automatically adjusted when you clicked the 'Fill' button in the EDIT SEQUENCE window, and STARTTIME, which we adjusted earlier (in the section 5.2, 'Linking a sound file').

We can change the other sequence variables in the same way as we changed STARTTIME. Click at a 'M' and type a value, e.g. '2' for 'interv' and 'respond', '1' for question and '5' for score.

### 5.7. Changing info fields

You can copy the information from info field 1 of sequence 3 and copy it to the same field of sequence 4 using 'Copy' and 'Paste' from the 'Edit' menu. We will show you a different way, that is especially useful if you want to copy information to a number of sequences at once.

- ▶ Go back to sequence 1.
- ▶ Select **Paste sequences...** from the **Edit** menu.
- ▶ Click at the radio button 'Copy part of sequence' in the PASTE SEQUENCES setup window.
- ▶ Check 'Copy info field 'question' (if you did not change the name of this info field in section 4.2, it will read 'info 1' instead of 'question').
- ▶ Check 'destination field is empty' below 'Only paste if:'.
- ▶ Be sure that all other checkboxes are unchecked, that the line at the top of the window reads "Copy information from sequence 1", and click .
- ▶ Go to sequence 4 and check that the text is pasted indeed.

### 5.8. Closing and saving

You're prepared now to create your own SEQUENCE VIEWER file. This is treated in chapter 6.

- ▶ Close the demo file (select **Close** from the **File** menu or click the close box in the upper left corner).

A window appears telling you about the changes you made to the file. This, of course, may help you to decide whether or not you want to save the changes. It is important to know that SEQUENCE VIEWER does not report all changes you made to the file. Generally, SEQUENCE VIEWER reports all changes with respect to event texts, event codes, event variables, sequence variables, keys and word links. It does not report changes in suggestions, sequence definitions and the like, as such information is not conceived as 'primary data'. Of course, any change in your file, no matter how small, will cause SEQUENCE VIEWER you to ask whether or not to save the file, and such changes are stored with your file when you save it. For the moment, you can decide yourself whether or not you save the file, but if you do, be sure you have the original file available.

# PART II: Selected topics

## 6. Creating a sequence file

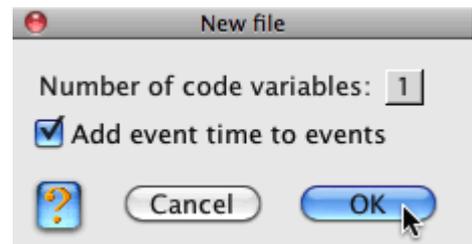
SEQUENCE VIEWER has a number of options that helps you to create a new file and to link audio or video files in a quick and easy way. Which of these options are most useful, depend on the type of your data. For the present purpose we assume that you have available three movie files, "movie1", "movie2" and "movie3". Moreover, "movie1" consists of ten episodes, each corresponding to a sequence. Similarly, "movie2" concerns eight sequences and "movie3" 14 sequences. You want to create one data file, with  $10 + 8 + 14 = 32$  sequences, each sequence linked to the correct episode of the appropriate movie. Each sequence (episode on the movie) consists of a number of fragments, corresponding to events. You also plan to assign onset and offset times to each fragment (event). We will describe the easiest way to create such a file. If you understand the basic ideas, you can easily adapt them to your own situation.

### 6.1. Creating the new file

▶ Select **New** from the **File** menu.

A small window is opened asking you how many code variables you want and whether or not you want event time variables (the offset and onset times).

Because you want to have onset and offset times, check 'Add event time to events'. If you already know how your coding scheme will look like, you should select the correct number of code variables. If not, for example because you will develop the coding scheme after inspection of the data, select '1'. You can always add (or delete) code variables afterwards.

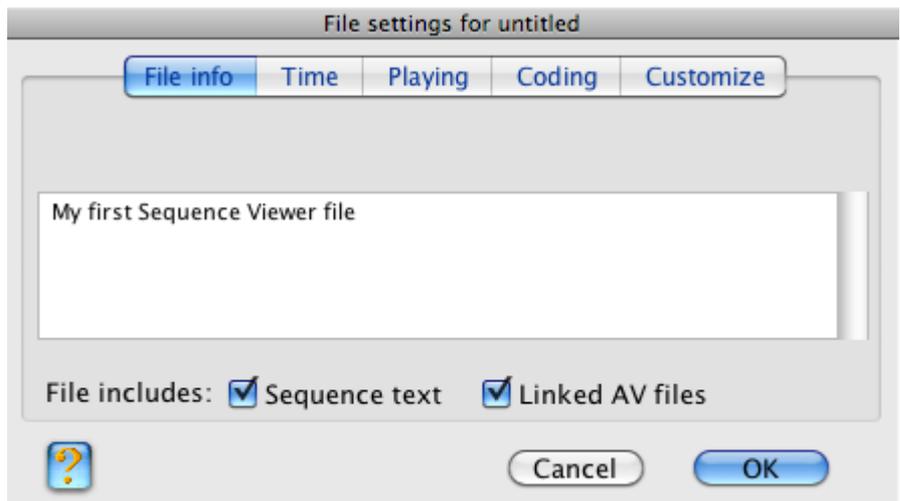


▶ Click **OK**. The main sequence window is opened with one (empty) sequence.

### 6.2. File settings

▶ Select **File settings...** from the **File** menu. The **File info** tab should be selected in the FILE SETTINGS window.

You should at least decide upon some basic characteristics of your new Sequence Viewer file. If you use sequence texts (like transcripts), you should check 'Sequence text'. In our example, you will also use video, hence check 'Linked AV files'. In the large text field you can enter any descriptive information about your file.

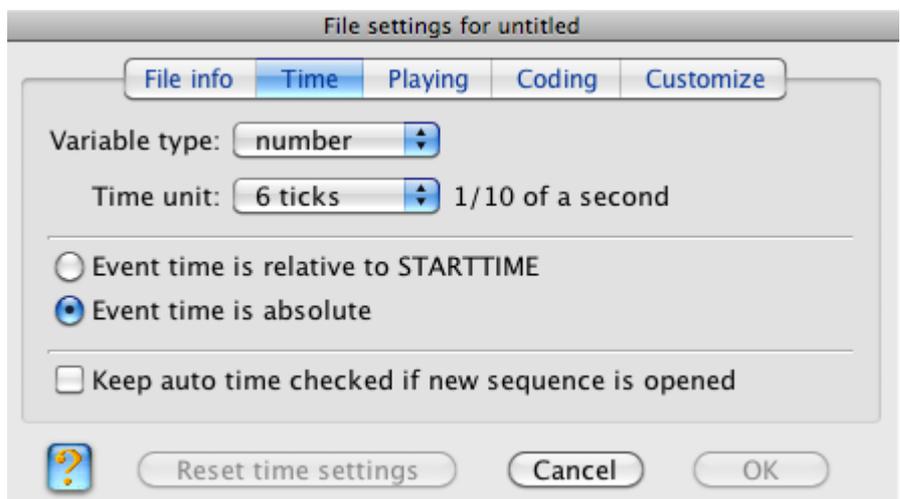


Next you should select an appropriate time unit.

▶ Select the **Time** tab in the File settings.

In case of speech, and calculating information like latencies between utterances, we recommend to select '6 ticks' (or one tenth of a second). You can change the time unit afterwards, for example from 10 ticks to 6 ticks, but in that case you will lose some precision because of rounding errors. In case of video, it may be useful to let the time unit correspond with the number of frames per second; for example 30 fps corresponds to 2 ticks.

You also have to decide between absolute and relative time. Absolute



means that the onset and offset times of each event refer to the very start of the AV file. Relative time means that onset and offset times refer to the time the sequence starts (the STARTTIME). For example, suppose that sequence number 3 of "movie2" starts at time 1198, that is, 1198 time units (e.g. of 0.1 second) from the beginning of "movie2". Hence the STARTTIME of sequence 3 will be 1198, and the absolute onset time of the first event of sequence 3 will be 1198 too. The relative onset time however is 0 (as it starts 0 time units from time point 1198 of the AV file). Suppose that the second event of this sequence starts 34 time units later. The relative onset time of this event is 34 time units, whereas the absolute onset time will be  $1198 + 34 = 1232$  time units.

Thus, relative onset (and offset) times are relative to the start of the sequence: the STARTTIME. If you plan to use relative times, the STARTTIME of each sequence should have a meaningful value, viz. the time the sequence starts. In most cases the best choice is 'absolute', but you can always change the absolute into the relative time and vice versa; the program will adjust the onset and offset times automatically.

- ▶ Select 'Event time is absolute'.
- ▶ Select the **Playing** tab in the FILE SETTINGS.
- ▶ Check 'Stop sound at end of event' and 'Play last ten units of event' below 'If the capslock is down'. In section 5.5 about adding event times, we saw the effects when the caps lock is down.
- ▶ Check 'Open linked file if new sequence is opened'. This means that if you go to a next sequence, the linked file will automatically be opened; as we will see, this saves you time when linking the movie files.
- ▶ Select the **Coding** tab in the File settings. You may check 'Use first character of event text as first code'. As we saw in section 5.4 about coding, the **Auto fill** button in the EDIT SEQUENCE window caused that the first code of each event code equalled the first character of the event text ('I' or 'R'). If you want to make use of this option, check this checkbox.
- ▶ When you are done, click **OK**.

### 6.3. Linking movies

Now we are ready to link 'movie1' to the first sequence. You can use any movie you have available on your hard disk for this exercise. You don't have to rename it to 'movie 1'; we just use this name to refer to the movie you are going to link. Movies (or audio files) cannot be changed in SEQUENCE VIEWER, so there is no risk in linking a file. Linking an audio or video file just means that the file path of the file is stored in your sequence viewer file, not the audio or video file itself. You can link up to four different files to one sequence.

- ▶ Click the 'link new file' button  located in the 'Linked AV files' section of the main sequence window.
- ▶ Select the file 'movie1' (that is, the movie file you want to link) and click **OK**. The movie file is now linked to the first sequence and will appear in a PLAYER window. Now we need to enter the start time of the movie. Probably this is 0 (zero), if the start of the movie and the onset time of the first event of the sequence coincide.
- ▶ Click the ▼ triangle in the lower left corner of the main sequence window to show the sequence variables drawer.
- ▶ Click the 'M' after STARTTIME and type a '0' (see section 5.2).

Our first movie covers ten sequences. Hence we should add 9 sequences to our file (there is already one sequence). Of course you can click the 'add sequence' button  located in the upper left corner of the main sequence window nine times, but this becomes a bit boring in case you want to add 1000 sequences. Fortunately there is a better way.

- ▶ Open the ADD SEQUENCES window with **add sequences...** from the **Sequences** menu.
- ▶ Check 'Copy information'.

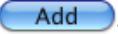
A number of check boxes become enabled, and you can specify the sequence number from which you want to copy the information. We only have one sequence, so leave the '1' in the edit field after 'Copy information from sequence:'.



- ▶ Check 'Copy linked files'.

This means that the linked files from sequence 1, are also linked to the new sequences. This of course saves you a lot of time; you don't have to link the same movie file to each sequence separately. And this is why we first linked "movie1" to the first sequence before we add new sequences. Because sequence 1 does not contain other information, it doesn't make sense to check more check boxes.

- ▶ Type '9' in the edit field between 'add' and 'sequences' and select 'after sequence 1' with the popup menu below 'Add where'.

- ▶ Be sure the window looks like the screen shot on the previous page, and click .

Now your file should contain 10 sequences. The file "movie1" is linked to all ten sequences. The program goes to the second sequence. The movie remains visible, thanks to the fact that you checked 'Open linked file if new sequence is opened' in the FILE SETTINGS window.

#### 6.4. Entering start times

Now you should enter the start times for each sequence; that is, enter the time an episode starts that corresponds with a new sequence. To this end, you should move the play head to the start of the second episode:

- ▶ Click at the triangle ▶ before 'more' in the lower left corner of the PLAYER window to extend the window.

- ▶ Use the slider  to roughly position the play head.

- ▶ Use the precision slider  to position the play head more precisely.

- ▶ To enter the time you can click the 'M' after STARTTIME as described in section 5.2, and enter the time as appears after 'time:' in the PLAYER window. Alternatively, you can hold both the option and command keys down, and click at the word 'time:' (see screen shot below). The time is automatically entered in the STARTTIME field. A third method is pressing a 't' with the option and command keys down.



- ▶ Go to sequence number 3 and proceed entering start times until sequence number 10.

You may now go back one or two sequences, and observe that the movie automatically goes to the correct position, that is, the start of the episode corresponding with the sequence.

To add the second movie, go to the last sequence (number ten), click the 'add sequence' button , link file "movie2", enter the start time (zero), add seven sequences (this movie has eight episodes) and proceed as described above. Finally you can add the sequences for file "movie3".

## 7. Importing files and analyses

SEQUENCE VIEWER has a number of import options. First you can create a new SEQUENCE VIEWER file with codes from a file with sequential data. Secondly, you can add information to an existing SEQUENCE VIEWER file, for example adding SEQUENCE VIEWER files to the present file (that is, combining files), adding sequence or event variables to an existing file, and adding sequence texts to an existing file. In this section we will discuss the first options. In addition we will discuss some common analyses.

You may have available a file that contains data that can be viewed as 'sequential' in one way or another. Our problem is how we can create a SEQUENCE VIEWER file with codes from such a file. Our first example concerns a file with life events: each sequence represents a person and the events represent different stages in their life with respect to living and education. This file also illustrates that SEQUENCE VIEWER files are not confined to (interviewer-respondent) interactions. The second file is just a plain text file, consisting of an interview with President Bush, conducted by Juan Williams on Monday January 29, 2007 for NPR.

### 7.1. The life events file

The life events file was created by SPSS, using data from a survey among young adults about their careers with respect to living, education, and job. So at a particular moment in time, the respondent's position on these three variables could be described in a three-letter code. These three event variables could take the following values (codes):

**living:** H (at home with parents), S (Single, living alone), P (with Partner, not married), M (with partner, Married), 0 (zero; other, e.g. student dormitory)

**education:** F (Full-time), P (Part-time), 0 (no education)

**job:** F (Full-time, 4 days or more), P (Part-time, 1 to 4 days), 0 (no job)

For example, 'HPP' means that the respondent lives at home, (H), follows a part-time education (P), and has a part-time job (P). As another example, 'M0F' means that the respondent lives with his/her wife/husband (M), does not follow an education (0), and works four or five days a week (F). A change in either one of these variables, is viewed as a new event. For example 'P0F' followed by 'M0F' means that the respondent got married. For each respondent a sequence was created (from within SPSS), consisting of a sequence of such coded events, delimited by a tab. Two files were created, one for men (250 cases) and one for women (244 cases). Separate cases (respondents) were delimited by a carriage return. The sequences (life courses) start at age 6.

Because such a file closely resembles the structure of Sequence Viewer files with respect to the event codes, it is very easy to import such a file. First we import the data from the male respondents.

▶ Open the SEQUENCE VIEWER program. If SEQUENCE VIEWER is already open, close any open file.

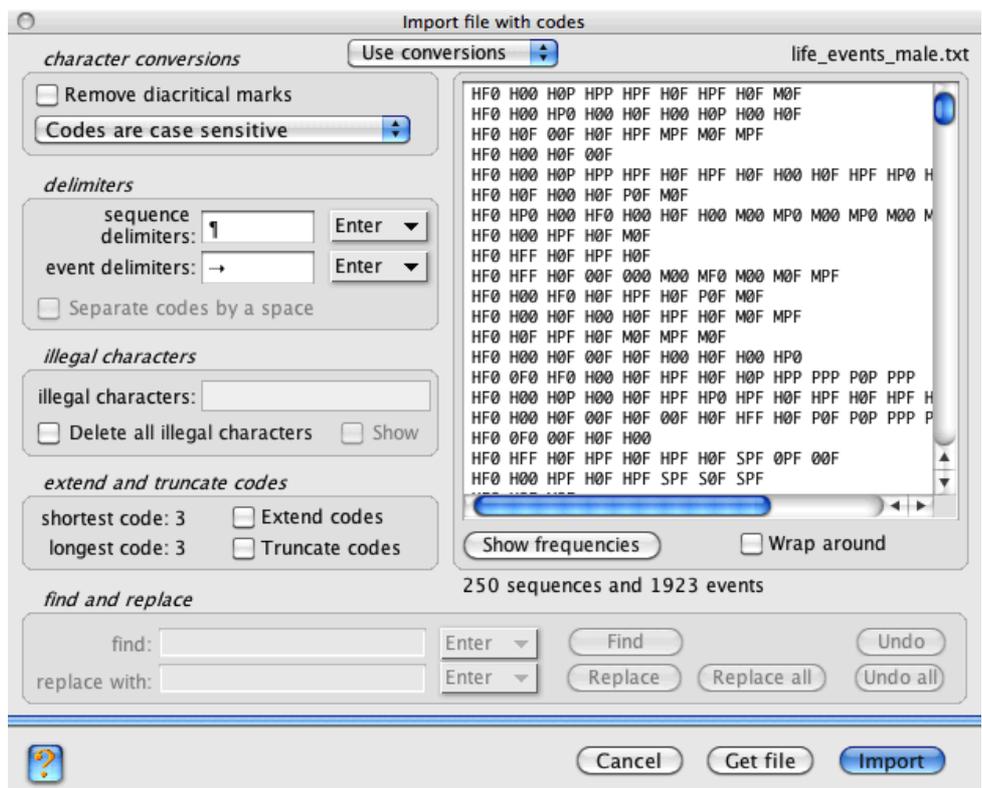
▶ Select **Import...** from the **File** menu.

▶ From the IMPORT FILE window, select 'Import codes' and click **OK**.

▶ In the IMPORT FILE WITH CODES setup window, click **Get file** (in the lower right corner). Select the 'life\_events\_male.txt' file, located in the 'tutorial files' folder and open it. The contents of the file appear in the field at the right side of the window.

▶ At the top of the window, is a popup menu. Select 'Use conversions'. The window should now look like the screen shot.

From the contents of the file, SEQUENCE VIEWER guesses which characters serve as sequence delimiters (carriage return in this case, shown as ¶)



and events delimiters (the tab, shown as →). There is no need to make further adjustments.

Click **Import**.

The Sequence Viewer file is created. It consists of just codes. Next we want to add the sequences for the female respondents. To know which sequences concern male and which sequences concern female respondents, we first add a sequence variable.

Select **Add variable...** from the **Sequence variables** menu. Rename 'SEQVAR3' to 'gender', set the variable type to 'number', the decimal digits to 0, the minimum value to 0, the maximum value to 1 and the initial value to 1 (see screen shot).

Click **Add**.

To add the file with the sequences of the female respondents, you have to close and save the file with data from male respondents. You can first choose **Save** and then **Close** from the **File** menu. For illustrative purposes, just choose **Close**, without previous **Save** (or click the close box of the main sequence window). Because you have made changes to your data file, you are asked if you want to save the changes before closing. Moreover, you are informed about the kind of changes you made:

Click **Save**. A common 'Save file' dialog appears. Save the file as 'life\_events\_male sv6' in the 'tutorial files' folder.

To create the file with sequences of female respondents, select **Import...** from the **File** menu.

From the IMPORT FILE setup window, select 'Import codes' and click **OK**.

Click **Get file**, select the 'life\_events\_female.txt' file, select 'Use conversions' and click **Import**.

Select **Add variable...** from the **Sequence variables** menu. Rename 'SEQVAR3' to 'gender' (be sure the name is exactly the same as in the 'male' file), set the variable type to 'number', the decimal digits to 0, the minimum value to 0, and the maximum value to 1. Type 0 as the initial value.

Click **Add**.

Now we have to combine both files. We can either add the 'females' file to the 'males' file, or vice versa. Because the 'female' file is already open, it's easier to add the 'male' file to the 'female' file.

Select **Import...** from the **File** menu.

From the IMPORT FILE window, select 'Import Sequence Viewer file(s)' and click **OK**.

The 'Add Sequence Viewer files' setup window appears, with some information about the 'female' file.

Click the **Get file** button at the bottom of the window.

Select the 'life\_events\_male sv6' file you just saved and open it.

The name of the file to be imported appears in the field in the 'files to add' box. In the 'file info' box some information about this file is displayed.

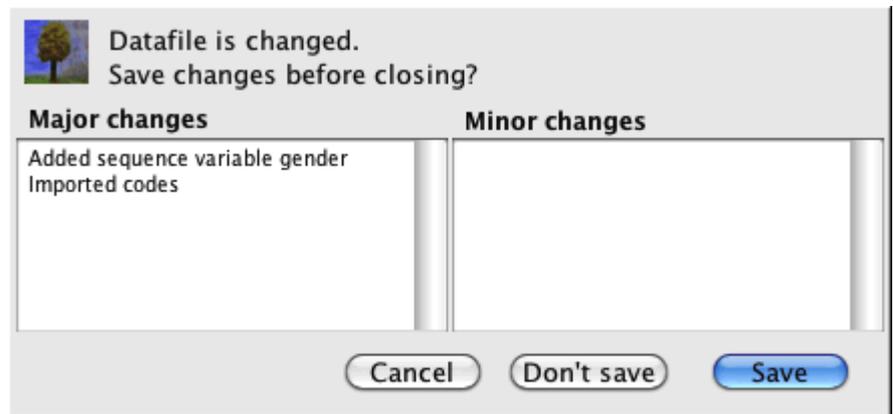
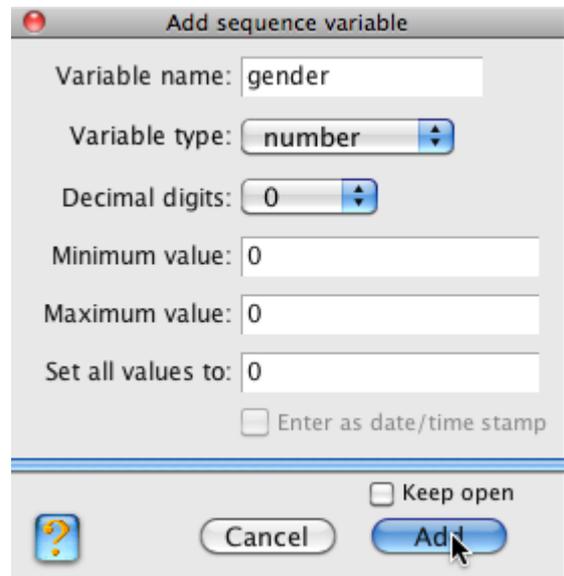
Click **Add**.

You may observe that the file now has 494 sequences: both sets of data are combined. Save the file as 'life\_events sv5' (**File** menu, **Save as...**).

Next, to remember that '0' means female and '1' means male respondents, you may add labels to these values.

Select **Define variables...** from the **Sequence variables** menu. Select 'gender' (click at 'gender' in the list of sequence variables).

Click in the text box, above the (disabled) **Add** button. Type a '0' (the value for women).



- Click in the text box to the right of the values text box. Type a label, e.g. 'female'.
- Click at the **Add** button (which now has become enabled). In the field with values and labels, you can now see '0 female'.
- Enter 'male' as the label for value '1' and click **Add**. Click **Store** to store all changes you have made to the properties of your sequence variables.



Finally, we also add labels to the different codes. In addition we will give the code variables appropriate names, and define the allowed codes.

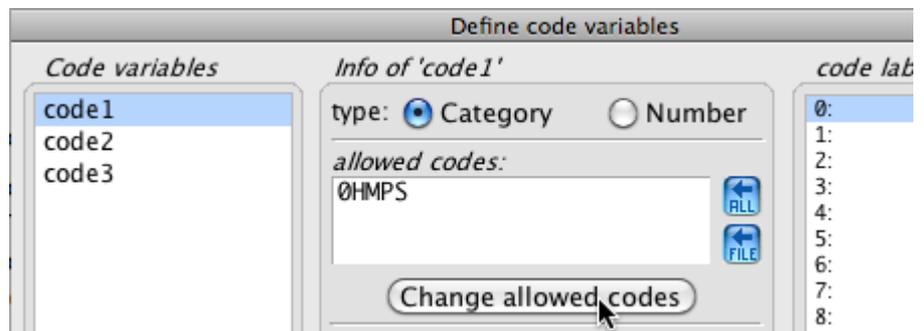
- Select **Define event codes...** from the **Codes** menu.
- Click at 'code 1', if it is not selected yet.

The field below 'allowed codes' show all codes that are allowed for this particular code variable. As you can see, all available codes (0..9, a..z, A..Z) are allowed. Actually however, this code variable can only take the codes H, S, P, M, 0 (see the beginning of section 7.1). To prevent typing errors we will change the allowed codes. You can type the allowed codes, but you can also click at the 'File' button.

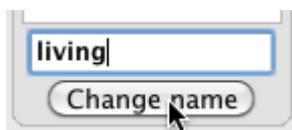
- Click the 'File' button.

The program now looks into the file which codes actually occur for this code variable, and puts these codes in the 'allowed codes' field. In addition the **Change allowed codes** becomes enabled.

- Click the **Change allowed codes** button.



You can now add meaningful labels to the codes, e.g. 'home' for 'H', etcetera, as was shown in section 3.2. To change the name of code variable 'code 1', into a more meaningful name, just type a name, e.g. 'living' in the edit field below the field with the names of code variables, and click **Change name**. If you are done with the first code variable, click at 'code 2' in the code variables field, and change the properties of this code variable in a similar way. When you are done with all three code variables, click **Store**.

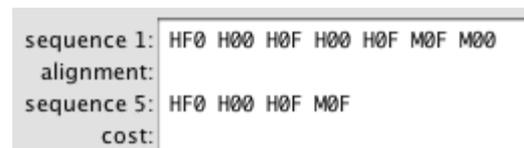


## 7.2. Some analyses with the life events file

### Optimal matching

Sequential analysis often concerns how much one sequence resembles another sequence. One of the methods to obtain a measure of the agreement between two sequences is called optimal matching or alignment. Essentially it is based on the number of transformations (deleting an event, inserting an event) that are needed to make both sequences similar. SEQUENCE VIEWER provides the optimal matching algorithm (in addition to some other measures of agreement). SEQUENCE VIEWER also shows the eventual aligned sequences (just read on, in a moment it will become clear what this all means).

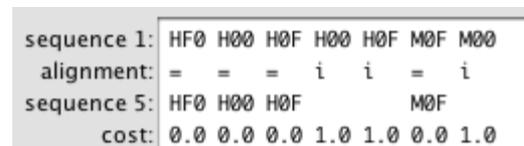
- Select **Define agreement...** from the **Codes** menu.
- Select 'OM unweighted' from the **Agreement measure** popup menu.
- Type '1' after the prompt **Sequence 1 =**.
- Type '5' after the prompt **Sequence 2 =**.



The field now shows both sequences. You can see that both sequences have the first three event codes in common. The fourth event code of the second sequence ('MØF') also occurs in the first sequence (the sixth one). So we can make both sequences alike by deleting event code 4, 5 and 7 from the first sequence. Or insert these three events at the appropriate positions in the second sequence. Making such modifications is called 'alignment'.

- Click **Align** to align both sequences.

The 'i' on the line after 'alignment' means that 'HØØ', 'HØF' and 'MØØ' from the first sequence should be inserted at these positions in the second sequence, to make both sequences similar. Other possible operations are deletions ('d') and substitutions ('s'). In the present example the number of operations (only



insertions here) is three, hence the total *cost* is 3. The maximum number of necessary operations to make two sequences similar equals the number of events of the longest sequence, seven in this case. Hence the agreement is  $1 - (3/7) = 0.571$ .

What can we do with such agreement measures? A research question might be how life events of men and women differ. Such a question can be answered in several ways. For illustrative purposes, we will compare how much the life course of men and women differ from a traditional life course. A traditional life course may look like (remember that the life course starts at age 6):

HF0 (at home, full time education, no job)

H00 (at home, no education, no job)

H0F (at home, no education, full time job)

M0F (married, no education, full time job)

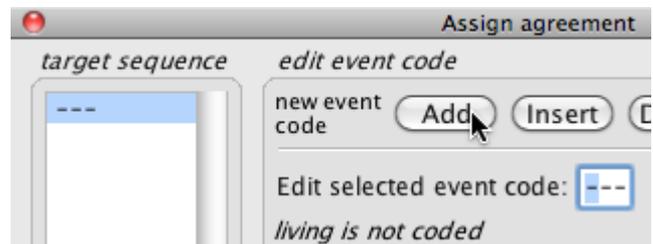
How do the actual life courses compare to this 'traditional' life course, and how do men and women differ in this respect?

➤ Leave the settings in the 'Define agreement' window as they are (OM unweighted, Substitution/indel cost = 1.0), and click **Define**.

➤ First, select **Add variable...** from the **Sequence variables** menu to create a new sequence variable. Call it 'TradAgree' (from 'agreement with traditional life course'), and set the minimum value to 0, the maximum value to 1 and the number of decimal digits to 4. Leave the initial value as 'M'. Click **Add**.

➤ Select **Assign agreement...** from the **codes** menu. The 'Assign agreement' setup window pops up.

In the field below 'target sequence', we have to enter the sequence that we want to compare with the sequences in the file: HF0 H00 H0F M0F.



➤ Click the **Add** button. This adds an 'empty' event code (with just dashes) to the field.

➤ Type 'HF0' in the edit field after 'Code:'. Each code you type immediately replaces the corresponding dash in the target sequence field.

➤ Add another event with **Add** and type 'H00'. Proceed with the third (H0F) and fourth (M0F) event codes.

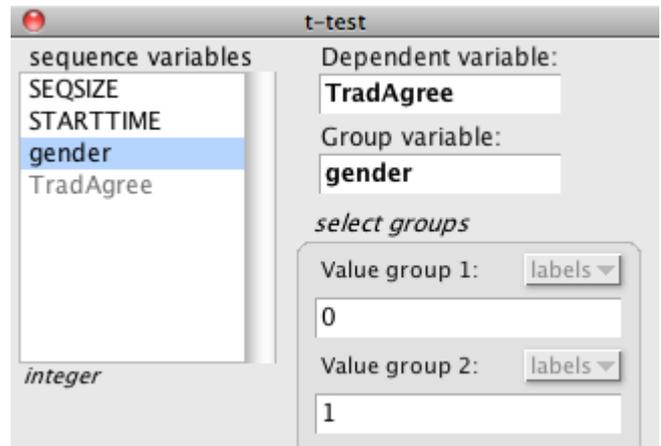
➤ Select 'TradAgree' in the field with sequence variables. The selected variable name appears in the edit field below the field with variable names. You can also type a 't' in the edit field; SEQUENCE VIEWER selects the first sequence variable that starts with a 't'.

➤ Click **Assign**.

The agreement between the HF0 H00 H0F M0F and each sequence in the file, is assigned to the sequence variable TradAgree. We can now perform a t-test on the difference between men and women with respect to this variable.

➤ Select **T-test...** from the **Sequence variables** menu.

➤ Be sure the setup looks like the screen shot at the right, and click **OK**.



The output tells you that although the agreement for men is a bit higher, the difference is not statistically significant.

Results also dependent on the selected agreement measure in **Define agreement...** from the **Codes** menu. This window is also directly available with the  button at the bottom of the 'Assign agreement' setup window.

*note*

A more sophisticated (and more appropriate) analysis to answer our research question is a cluster analysis (also available in SEQUENCE VIEWER).

*Transition matrix*

Another common sequential analysis concerns which events (*given events*) are followed by which other events (*target events*). SEQUENCE VIEWER provides a number of different *transition matrices*, telling you which events tend to be followed by which other events.

➤ Select **Matrix...** from the **Codes** menu.

▶ Check 'frequencies' (in the 'output' box) and 'general dependency test' Leave all other settings as they are.

▶ Click **OK**.

The output shows a frequency transition matrix: each cell contains the number of times a particular event code is followed by another event code. The dependency test tells you that in general the occurrence of a particular event code is highly dependent on the preceding event code. The transition matrix itself however is not very informative about particular dependencies. First, the frequency does not tell you much about the relation between two successive events. A high frequency does not necessarily mean that both events are related in a statistical sense: one may expect that two codes that occur often, will also have a high transition frequency, whereas two event codes that occur seldom, will have a low transition frequency. This can be solved by the other types of transitions matrices that SEQUENCE VIEWER provides.

Another drawback is the size of the matrix: it is quite difficult to find a pattern in the transitions. A solution for this is the *transition graph*.

▶ Select **Transition graph...** from the **Codes** menu.

▶ Click **Calculate**.

A graph is shown, with only those event codes that have the highest transition frequencies. In each circle you will find the frequency of the event as given event below the event code, and the frequency as target event above the event code.

Let's first adjust the graph a bit. Just hold the mouse down on a circle and drag the circle to another position. Try to position the circles in such a way that you obtain a nice looking graph. In general, to obtain an easy to interpret graph:

- position circles with high given event frequencies, and low target event frequencies more to the left of the graph, and circles with low given event frequencies, and high target event frequencies more to the right of the graph;
- single arrows point from left to right;
- prevent that arrows overlap.

For example, you can read the text in the circle of event code 'HF0' as:

64 → HF0 → 555,

meaning that HF0 is the target event of (is preceded by) 64 other events, whereas HF0 itself serves as given event for 555 target events. Of these 555 target events, 325 events are coded as 'H00': the frequency in the middle of the arrow between HF0 and H00.

You can use the **Options** button to vary the number of event codes in the graph (by changing the minimum transition frequency). You can also let the arrows reflect the size of the transition frequency, or colour the circles, depending on a particular code.

You can send the graph to the output window, save the graph, or copy the graph to the clipboard to paste it for example in a MS Word document, using the icon buttons at the bottom of the window.

### Selecting sequences

You may wonder whether the pattern is different for men and women. To this end we should analyze only those sequences belonging to men, respectively women. There are several ways to accomplish this. Here is the easiest one for the present example.

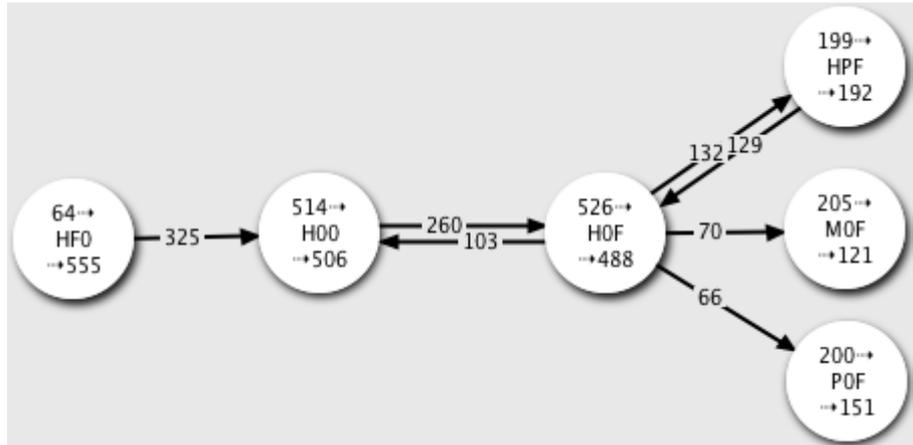
▶ Select **Find value of sequence variable...** from the **Sequence variables** menu.

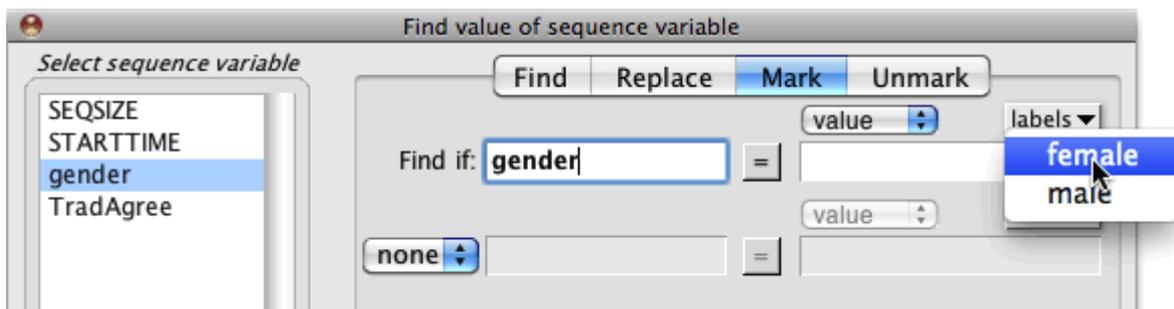
▶ Select the 'Mark' tab.

▶ Click in the text box after 'Find if:'.

▶ Click at 'gender' in the Sequence variables list.

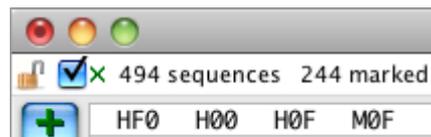
▶ Select 'female' from the labels drop down menu (see screen shot at the next page).





The value (0) is now automatically entered in the values text box. Of course you can also enter a zero directly into this box, but generally it is more easy to remember labels than values. The labels drop down menu is disabled if you don't have labels for a particular sequence variable; in that case you have to type the value.

- ▶ Check 'Unmark all sequences first' at the bottom of the window.
- ▶ Click **Mark**. 244 sequences are found and marked, as can be seen from the top left corner of the main sequence window.
- ▶ Close the FIND VALUE OF SEQUENCE VARIABLE window.
- ▶ Open the 'Transition graph' window once again (if you closed it).
- ▶ Check the marked sequences  icon at the bottom of the window.



The lower part of the window now looks like this:  
You may note that you are told how many sequences are selected.



- ▶ Click **Calculate**.
- ▶ Use the **Options** button to decrease the minimal frequency (and hence increase the number of relations and circles with event codes). You may also try the other options. Move the circles until you have an interesting graph.
- ▶ Click the 'output' button to send the graph to the output window.

You can now use **Find value of sequence variable...** from the **Sequence variables** menu to mark the sequences of the male respondents. It's faster however in this case to unmark the marked sequences and mark the unmarked ones.

- ▶ Go to the **Sequences** menu and select **Mark sequences → Toggle marks**.
- ▶ Create a graph of the life events of the male respondents. Use the graph you sent to the output window to compare both graphs.

### 7.3. The Bush interview

#### Sentences as sequences

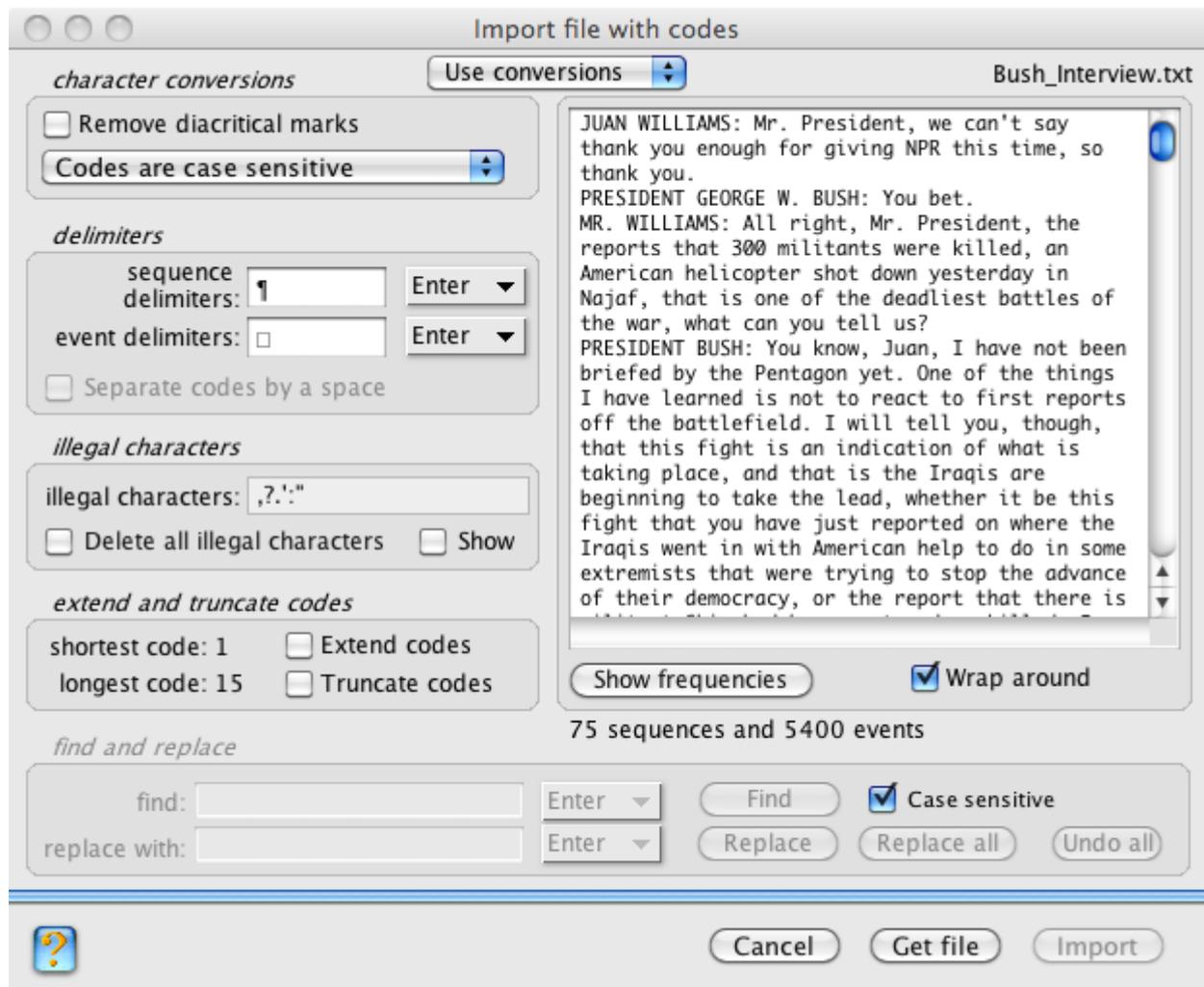
This exercise is merely to discuss the options in the IMPORT FILE WITH CODES setup window, and to illustrate that you can import nearly any kind of file, that is sequential in one way or another. The eventual file is not very meaningful for analyses (although linguists may find applications for it). The exercise is quite meaningful however as it shows you some interesting data manipulations. In a next section we will import the same file in a different way, that makes more sense for qualitative analyses. For the present purpose, we will view sentences as sequences, and words as events ('event codes').

- ▶ Open the SEQUENCE VIEWER program. Close any open Sequence Viewer file.
- ▶ Select **Import...** from the **File** menu.
- ▶ From the IMPORT FILE window, select 'Import codes' and click **OK**.
- ▶ In the IMPORT FILE WITH CODES setup window, click **Get file**. Select the 'Bush\_Interview.txt' file, located in the 'tutorial files' folder and open it. The contents of the file appear in the field at the right side of the window.

Of course the 'event codes' in this file (the words in the interview) are very different from eligible SEQUENCE VIEWER event codes. First, the 'event codes' contain characters that are not allowed as codes, for example question marks and comma's. Second, 'event codes' differ with respect to length, the shortest being only one character long, whereas the longest is 15 characters. The maximum length of an event code in SEQUENCE VIEWER is nine, and all event codes should be of equal length. We have to adjust the file to make it suited for SEQUENCE VIEWER. Select 'Use conversions' from the popup menu at the top of the window and check 'Wrap around' (see screen shot at the next page).

Let's first take a look at the illegal characters: characters that are not allowed as codes. You can find these illegal characters, as far as they occur in the file, after 'illegal characters'.

- ▶ Check 'Show' to colour all illegal characters red.



In the first sentence you can see that a single quote (which is an illegal character) occurs in the word “can’t”. If we delete the single quote, the word becomes “cant”. In stead, we can also replace “can’t” by “cannot”.

- ▶ Select 'Edit text' with the popup menu at the top of the window again.
- ▶ Enter “can’t” in the 'Find' edit box, “cannot” in the 'Replace' edit box, and click **Replace all**
- ▶ Also change “JUAN WILLIAMS:” and “MR. WILLIAMS:” into “INTJUAN”. The space between JUAN and WILLIAMS is shown as a square. Be sure to include the colon. We choose “INTJUAN” because this word does not appear in our file.
- ▶ Change “PRESIDENT GEORGE W. BUSH:” and “PRESIDENT BUSH:” into “RESPBUSH”.
- ▶ Select 'Use conversions' again.

You may observe that the single quote and the colon are no longer among the illegal characters; apparently they occurred only in the strings we have changed.123

Now let's have a look which characters are still in our file.

- ▶ Click **Show frequencies**.

You obtain a table with frequencies of the different 'event codes'. Scroll to the bottom. There you find an overview of both the legal and illegal 'codes'. The first 'legal code' is a dash. Although this is a legal code (for 'not coded yet'), it is unlikely that the dash in this file has this meaning.

- ▶ Select 'Edit text'. Type “-” in the 'Find' edit box and click the **Find** button. The dash appears to occur in the word “al-Qaida”. Without losing information we can simply delete this character. We can do this manually, because there is only one dash in the whole file.

We want to view sentences as sequences. That is, periods and question marks (and a carriage return) signify the end of a sequence. However, a period also appears in “Mr.”.

- ▶ Change “Mr.” into “Mister”.
- ▶ Select 'Use conversions'.

A first adjustment may be removing diacritical marks. Diacritical marks are not very common in English, but that is different for other languages. If 'Remove diacritical marks' is checked, a word like 'señor' is replaced by 'senor'.

- ▶ Because we view the period and the question mark as sequence delimiters (in addition to the carriage return), type both characters in the field with sequence delimiters, next to ¶, a symbol that has the meaning of 'carriage return'. The field with 'codes' is immediately adjusted.
- ▶ Spaces are viewed as delimiters between event codes (for different files this may be for example a tab). A space is shown as a . Be sure this symbol is in the 'event delimiters' field.

The period and the question mark are now sequence delimiters, and no longer appear as illegal characters. However, we want to make a distinction between 'sequences' (sentences) with and without a question mark. If we click the **Show frequencies** button, we can observe (near the end of the table) that the capital character 'Z' does not occur in our file. We can use this to indicate a question mark.

- ▶ Select 'Edit text' to go back to the Edit mode. Type a question mark in the 'find' field and "Z?" in the 'replace' field.

We are now ready to make our final adjustments.

- ▶ Select 'Use conversions'.
- ▶ The quote and the comma remain as illegal characters. We can check 'Delete all illegal characters' to get rid of them.
- ▶ Our final problem is the length of the 'event codes'. This is easily solved by checking 'Extend codes' and 'Truncate codes'.

All words longer than nine characters are truncated (fortunately we will not lose the "Z", because question marks only appear after shorter words in this file). All words shorter than nine characters are extended with dashes.

You may also observe that the **Import** button has become enabled. The file can be imported. Click this button. As you can see, although we have lost some information (especially by truncating the words), the sequences can be very well read. To finish the exercise we will now show you how you can use the question mark information and the speaker ('INTJUAN' and 'RESPBUSH') information.

#### Assigning information from codes to sequence variables

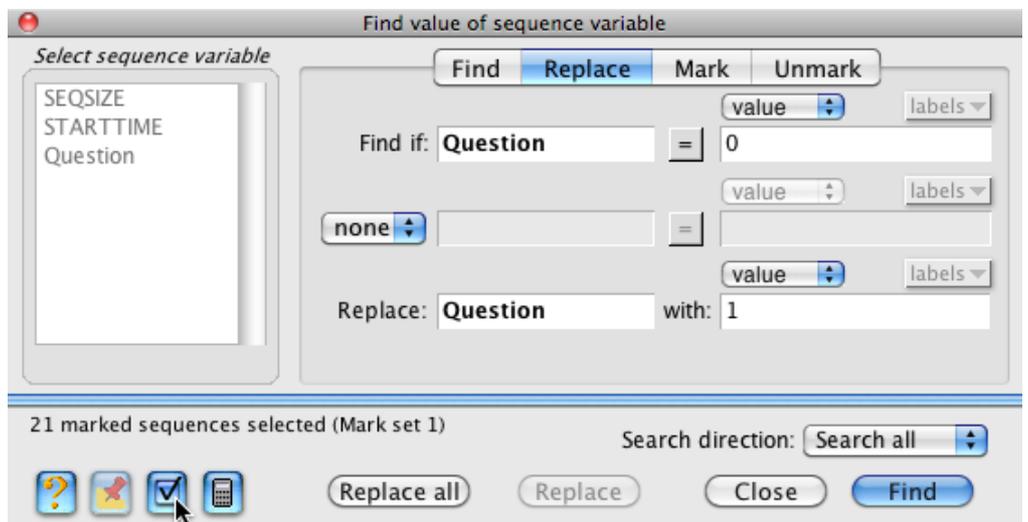
- ▶ In the 'Bush\_Interview' Sequence Viewer file, we first add a sequence variable with **Add variables...** from the **Sequence variables** menu. Call this variable "Question" and set all initial values to 0 (zero).
- ▶ Open the FIND CODES setup window with **Find event codes...** from the **Codes** menu and click the 'Mark' tab.
- ▶ Type an uppercase 'Z' at the position of the last question mark after 'Find:'. 
- ▶ Click the **Mark** button (one sequence should be marked).
- ▶ Type an uppercase 'Z' at the position of the one but last question mark after 'Find:', replace the 'Z' at the last position by a dash and click the **Mark** button. 
- ▶ Repeat this procedure until all sequences representing a question are marked. You should end up with 21 marked sequences. Close the FIND CODES window.

Now we are going to assign the value '1' to the newly added sequence variable 'Question' in case the sentence was a question.

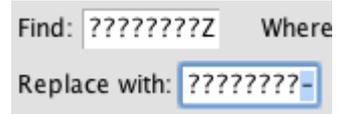
- ▶ Open the FIND VALUE OF SEQUENCE VARIABLE window with **Find value of sequence variable...** from the **Sequence variables** menu. Click the 'Replace' tab. Type 'Question' both in the 'Find if' field and the 'Replace' field. Type a '0' for the find value and a '1' for the replace value. And check the 'marked sequences'  icon (which is the crucial trick here).

- ▶ Click the **Replace all** button.

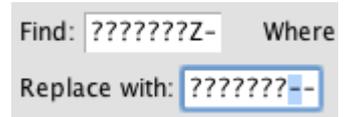
Now the sequence variable 'Question' has the value '1' if the sequence (sentence) is a question, and the value '0' otherwise. Next we have to get rid of the 'Z' code that told us whether the sentence was a question.



- Close the FIND VALUE OF SEQUENCE VARIABLE window and open the FIND CODES setup window with **Find event codes...** from the **Codes** menu once again and click the 'Replace' tab.
- Type an uppercase 'Z' at the position of the last question mark after 'Find:' and a dash at the position of the last question mark after 'Replace with:'.
- Click the **Replace all** button.

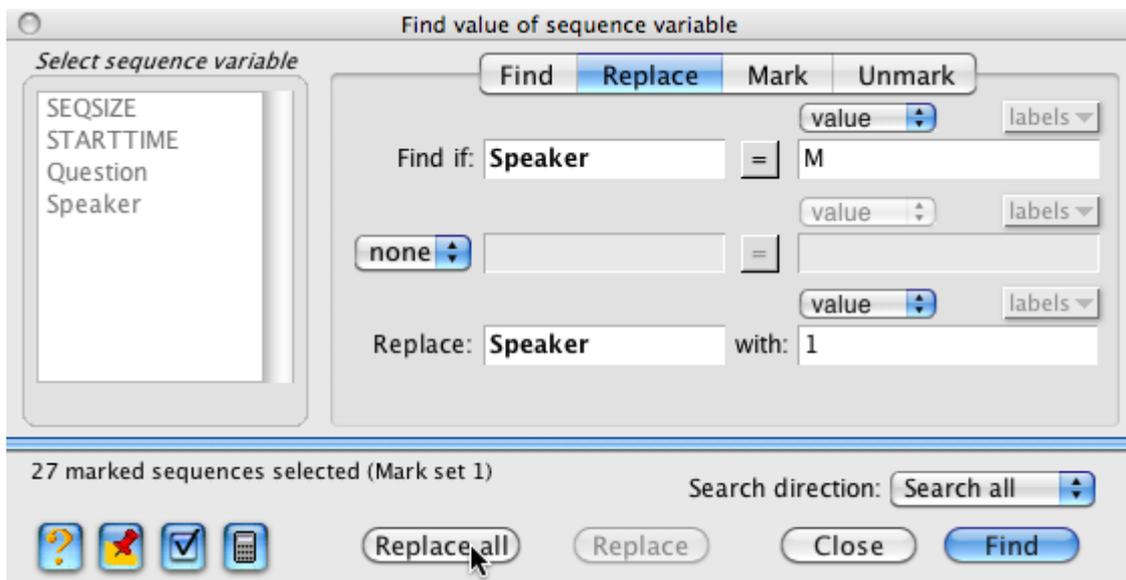


Repeat this procedure for the other positions of 'Z' in the event codes, e.g. for a 'Z' on the one but last position, see the screen shot at the right.



Finally we will add another sequence variable to indicate the speaker of a sequence.

- Add a sequence variable with **Add variables...** from the **Sequence variables** menu. Call this variable "Speaker" and set all initial values to M (indicating a missing value).
- Again, open the FIND CODES window and click the 'Mark' tab.
- Type 'INTJUAN??' as the event code that should be found. 
- Check 'Unmark all sequences first', because part of our sequences are already marked.
- Click the **Mark** button. You should have 27 marked sequences now.
- Use the FIND VALUE OF SEQUENCE VARIABLE window to assign the value '1' (designating the interviewer) to the sequence variable 'Speaker'. To this end we select 'Speaker' as the sequence variable to be found, and type the 'M' as the value to be found. Of course all sequences have the value 'M' on this variable, but we will replace this value with '1', only for the marked sequences. Hence the setup should look like the screen shot below.



- Click the **Replace all** button. 27 sequences are changed.

Next we repeat this procedure with 'RESPBUSH?' to assign the value '2' (designating the respondent) to the sequence variable 'Speaker'.

- Open the FIND CODES setup window and type 'RESPBUSH?' as the event code.
- Click the 'Mark' tab, check 'Unmark all sequences first' and click the **Mark** button. 26 sequences should be marked.
- Open the FIND VALUE OF SEQUENCE VARIABLE setup window and assign value '2' (designating the respondent) to the sequence variable 'Speaker'.

You can check the values of the sequence variables. In the main sequence window, click at the grey triangle (before 'vars', in the lower left corner) the get a drawer with those values. Move through your sequences with 'go next' (🔍). Sequence 5 still shows a 'M' for missing value on the variable 'Speaker': this sequence has no event code 'INTJUAN--' or 'RESPBUSH-'. But of course the speaker is the same as the preceding sequence. Hence we would like to assign the same value on 'Speaker' for a sequence as the *preceding* sequence, but *only* if the sequence has the value 'M'. So we first mark all sequences that should be changed: sequences with the value 'M' on 'Speaker'.

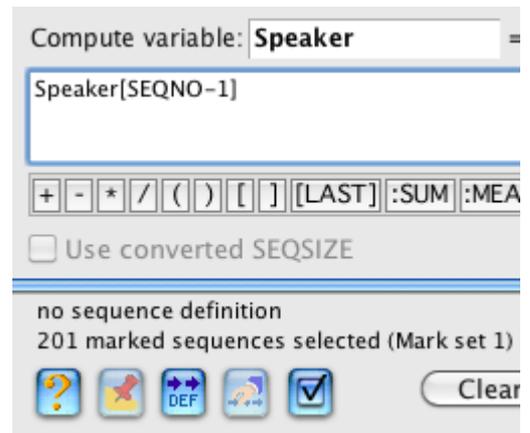
- Open the FIND VALUE OF SEQUENCE VARIABLE window and click the 'Mark' tab. Type 'Speaker' in the find field, a 'M' for the find value, check 'Unmark all sequences first' and click the **Mark** button. Now you should have 201 marked sequences.

▶ Open the COMPUTE SEQUENCE VARIABLE setup window with **Compute value...** from the **Sequence variables** menu. Type 'Speaker after 'Compute variable', and 'Speaker[SEQNO-1]' in the equation field. You should also check the 'Marked sequences' button!

▶ Click **Compute**.

The expression 'Speaker[SEQNO-1]' means that the program looks to the value on 'Speaker' of the preceding sequence ('SEQNO-1'). This value is assigned to the variable 'speaker' of each marked sequence. Which yields exactly the result we want. You may check this with the sequence variables drawer.

Finally you can delete all event codes 'INTJUAN--' and 'RESPBUSH-' from your file.



▶ Choose **Delete events...** from the **Sequences** menu. Enter 'INTJUAN??' after 'delete event' and click **Delete**. Do the same for 'RESPBUSH?'

You're done!

The primary purpose of this section was to show you a number of data manipulation options. But you can do all kinds of analyses on this file. An interesting analysis might be differences in the use of words by both speakers. You can perform such an analysis with **Frequencies...** from the **Codes** menu. Select 'Sequence variable' with the popup menu below 'by variable:' and select 'Speaker'. Also select 'percentages' instead of 'frequencies' to be better able to compare differences in word use. The output, although long, shows some interesting differences.

## 8. Working with keys

Keys are either (small) fragments of sequence texts (*text keys*), or of linked audio or video files (*time keys*). Text keys more or less resemble similar concepts in programs for qualitative analyses (e.g. 'quotations' in Atlas). We will first discuss the text keys in some detail, again using the Bush interview (but now in a more usual way).

### 8.1. The Bush interview revisited

- ▶ Select **New** from the **File** menu to create a new file. Set the number of code variables to 1, and don't add event times.
- ▶ Select **Import...** from the **File** menu. In the IMPORT FILE setup window, select 'Import sequence texts' and click **OK**.
- ▶ In the 'Import text' setup window, click the **Get file** button. Select the 'Bush\_interview.txt' file, located in the 'tutorial files' folder and open it.
- ▶ Leave the 'Import text' setup as is. The sequence delimiter field should be empty, because we will view the whole interview as only one sequence. 'Start import at first sequence' should be selected, because we will put the text of the interview in the text field of the first (and only) sequence. Click **Import**.

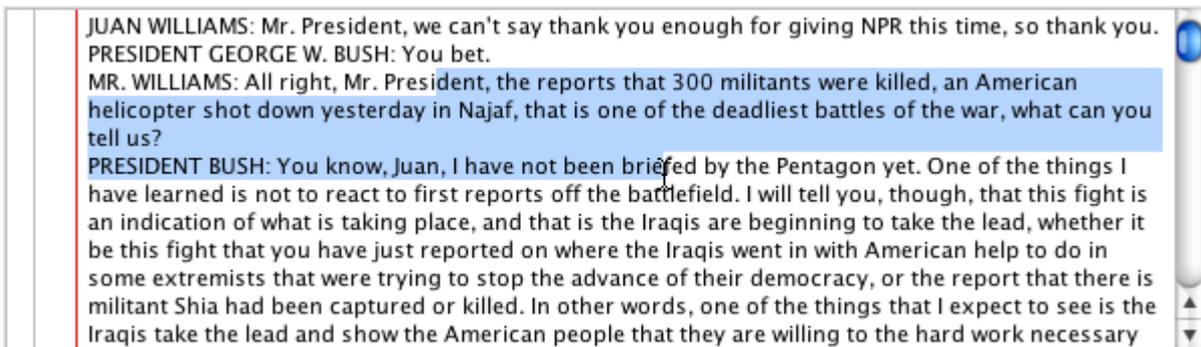
The text of the interview is now put into the text field of the main sequence window.

- ▶ Select **Text keys** from the **Keys** menu to open the TEXT KEYS window. In the upper part of the window you can see the *key field* slider. The position of the key field slider corresponds to a word in the text field. Move the slider, and observe that the word corresponding to the position of the slider is highlighted. In addition, the number in the upper left corner tells you the number of the word.

#### Adding a new text key

We will first show you how to add a new key to the file.

- ▶ In the TEXT KEYS window, click the **New** button, located at the left of the window.
- ▶ In the drawer that appears, type a name for the key, for example 'key1'. The name of a key is called a *keyword*. If you have already keys assigned, you can also select an existing keyword. Don't click **Add key** yet!
- ▶ Now select some arbitrary text in the text field of the main sequence window, for example:



As soon as you leave the mouse, you may observe that the selection is extended a bit: the whole word 'President' and the whole word 'briefed' are selected. Moreover, in the drawer where you entered the name, the number after 'start point' becomes 30, and the number after 'end point' becomes 70. This means that word 30 to 70 is selected. In the large white field of the TEXT KEYS window (the *key field*), you will see a small rectangle.

- ▶ Click the **Add key** button.



The small rectangle in the *key field* now becomes coloured. This rectangle represents the key. Keys with the same keyword will all have the same colour, whereas keys with different keywords are coloured differently. The width of this rectangle represents the number of words of the new key, and is proportional to the number of words of the whole sequence text. You may also note that keys are not confined to fragments within one paragraph (event): the present selection covers fragments of two subsequent paragraphs.

- ▶ Click the **New** button once again, to add another key. Call this key 'key2' and select a different part of text (but keys may overlap). Click the **Add key** button to add the key to your file.

As you can see, this key is coloured differently. SEQUENCE VIEWER adds arbitrary (random) colours to different keywords. Colours as well as some other properties of keywords can be changed. Keyword properties are saved with your file. To change the colour of a keyword, we have to open the DEFINE KEYWORDS window.

- ▶ Leave the TEXT KEYS window open and select **Define keys** from the **Keys** menu.

The DEFINE KEYWORDS window appears. In the left part of this window you see the names of both keys ('key1' and 'key2'), with a coloured bar in front of the keyword.

- ▶ Click at 'key1' (or at the coloured bar in front of it). The bar's height is increased (to show that this is the one you selected). Give the keyword a different colour, using the 'hue', 'saturation' and 'brightness' sliders.
- ▶ Click the **Store** button. The DEFINE KEYWORDS window is closed. In the TEXT KEYS window, all keys with keyword 'key1' in the key field, will obtain the new colour.

Each key in a sequence is uniquely characterized by its name (the keyword) and its position (from word x to word y).

- ▶ Move the mouse over 'key1' in the keys field. A tool tip is shown (you may have to wait a short moment), with the keyword and the position, like 'key1(30,70)'.
- ▶ Click at 'key1', and then at 'key2'.

The key in which you click becomes selected, the colour becomes lighter, to make it more easy to find the selected key. In the text field of the main sequence window, the corresponding text becomes flagged. Clicking at the white space in the keys field, deselects the key.

- ▶ Click at the ▶ triangle in the lower left corner of the TEXT KEYS window. A drawer appears with some information about the first and last keys in the sequence, a reminder of the colours of the different keywords, and a *notes* field.
- ▶ Click at 'key1' again.

The field below 'notes' becomes enabled. Moreover, 'notes' changes into 'notes for key1(30,70)' (or whatever is the range of key1). In the notes field you can enter any further description, for that key.

- ▶ Enter some text in the notes field, for example "This is my first key".
- ▶ Click at 'key2' in the keys field, and enter some descriptive text for this key. Close the drawer with the grey triangle and deselect the key (click at some white space in the keys field).

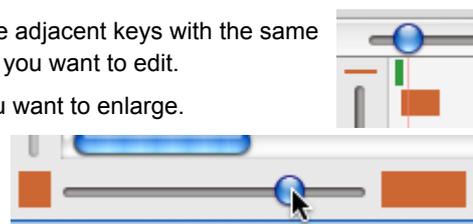
Maybe you observed that when you added the first key, a box with two popup menus was added to the main sequence window (in the lower right corner).

- ▶ Select 'key1' from the lower popup menu (the *keyword* popup menu). All texts having the keyword 'key1' are now coloured.
- ▶ Click at the  annotations button at the left of the main sequence window. A small yellow window, the *notes* window, appears.
- ▶ Move the mouse over the text field of the main sequence window. As soon as the mouse is over the text belonging to 'key1' or 'key2', the notes you added to the keyword, appear in the notes window. If both keys partly overlap, the notes for both keys are displayed, if the mouse moves over the overlapping text.

### Editing keys

You can resize keys, move keys to the left or the right, split keys, and combine adjacent keys with the same keyword. To be better able to edit keys, you may have to zoom in on the keys you want to edit.

- ▶ Move the key field slider at the top of the text keys window to the place you want to enlarge.
- ▶ Move the *horizontal zoom* slider below the keys field to zoom in on the keys. In addition you may have to scroll the keys field to the left.
- ▶ Click one of the keys to select it.



- ▶ Click at the move left  button to move the key one word to the left. The flagged text in the text field of the main sequence window shows you the effect immediately. The move right  button moves the key one word to the right.
- ▶ To move a key over a larger distance than one word, position the key field slider over the selected key. Check the 'Move' checkbox and move the key field slider with the shift key down.
- ▶ To extend a key with one word to the left, select it and click the extend left  button.



The remaining three buttons respectively shrink the key with one word at the left, shrink the key with one word at the right and extend the key with one word at the right.

- ▶ To resize a key with a larger number of words, position the key field slider exactly either at the left or the right border of the selected the key. The slider is correctly positioned when the 'Resize' checkbox becomes enabled. You may have to zoom in; because of the large number of words of the text, the slider cannot be positioned at each word in the text. You can also use the  or  button in the upper left corner of the TEXT KEYS window to move the slider one word to the right or to the left. Check the 'Resize' checkbox and move the key field slider with the shift key down.

➤ Now position the key field slider somewhere between the left and right border of the selected key. In the text field of the main sequence window, the word corresponding to the position of the key field slider is selected. The **Split** button (at the left of the TEXT KEYS window) becomes enabled. Click this button. The key will be split before the selected word in the text field.

Both keys will have the same keyword and the same note. You may also observe that the **Combine** button has become enabled.

➤ Select the right one of both keys and move it to the right with . The **Combine** button becomes disabled. Move the key to the left again to enable the **Combine** key.

➤ Click the **Combine** key. Both keys are combined. If the notes of the original keys are different, both notes are added to the notes of the new key.

#### Note

If you have split a key, both keys are exactly the same, except their position. When you close the key window, or go to a different sequence, both keys are automatically combined again, because it does not make sense to have two similar successive keys (without words between them). If you move one of both keys to insert a gap between both keys, or add a note to one of them, they are not combined of course (because they are not connected, or have different properties).

#### Automatically adding text keys

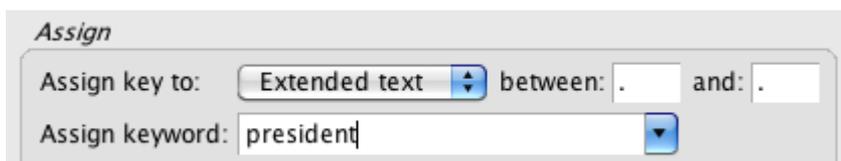
Adding keys manually is only one way to add keys. A different way is to assign a key to each instance of a particular piece of text.

➤ Select **Auto assign keys...** from the **Text** menu (leave the TEXT KEYS window open).

We now want to assign a key to all sentences with the word 'president'. We define a sentence as a piece of text bounded by a carriage return or a period *plus a space* (or, of course, the start or end of the sequence text).

➤ Type 'president' in the 'Find text' field of the AUTO ASSIGN KEYS setup window..

➤ In the 'Assign' box, select 'Extended text' with the pop-up menu after 'Assign key to'.



➤ Type a period in the edit box after 'between', and a period in the edit box after 'and'.

➤ Type a keyword for the new key, for example 'president'.

➤ Leave everything else as is, and click the **Assign all** button.

As you can observe from the TEXT KEYS window, the keys are automatically drawn. As usual, the output window informs you about your actions. You can read here that one key could not be added because of overlap. The reason is that in one sentence the word 'president' occurs two times.

➤ Select 'president' with the keyword popup menu in the keys box of the main sequence window (in the lower right corner). All texts with the keyword 'president' will be coloured (probably this is already done automatically).

To be able to perform a simple analysis, we add another key to our file. Now we will assign pieces of texts with the word 'not' to a key.

➤ If you closed the AUTO ASSIGN KEYS setup window, open it again with **Auto assign keys...** from the **Text** menu.

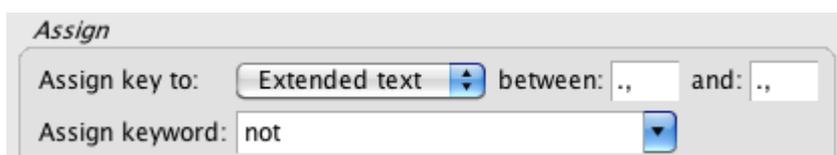
➤ Type 'not' in the 'Find' field.

➤ Check 'Whole words only', to prevent that a word like 'notion' is counted.

➤ Select 'extended text', and type both a period and a comma, both after 'between' and 'and'.

➤ Type a keyword for the new key, for example 'not'.

➤ Click the **Assign all** button.



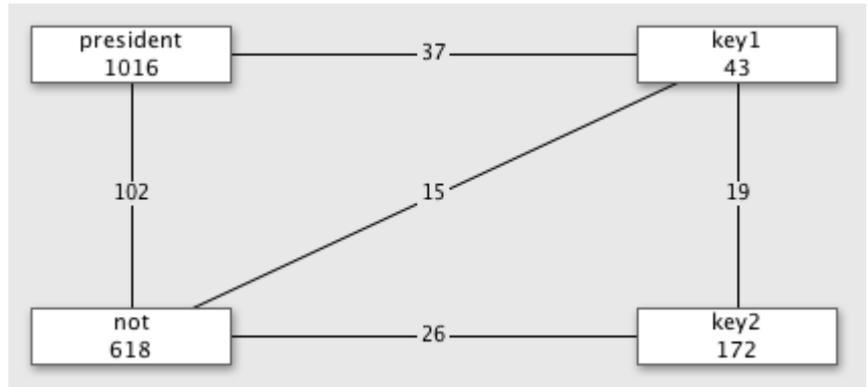
The result is as you probably expected: fragments are now bounded by comma's or periods (or the start and end of the event text). As a result, the keywords 'president' and 'not' partly overlap. Of course such overlap does not have much meaning here, but in case of more meaningful keys, it might be interesting to look after the degree of overlap. But first close the 'Auto assign keys' setup window, to prevent that the screen becomes cluttered with all kinds of windows.

➤ Select **Coincidence graph...** from the **Keys** menu.

➤ Click the **Calculate** button.

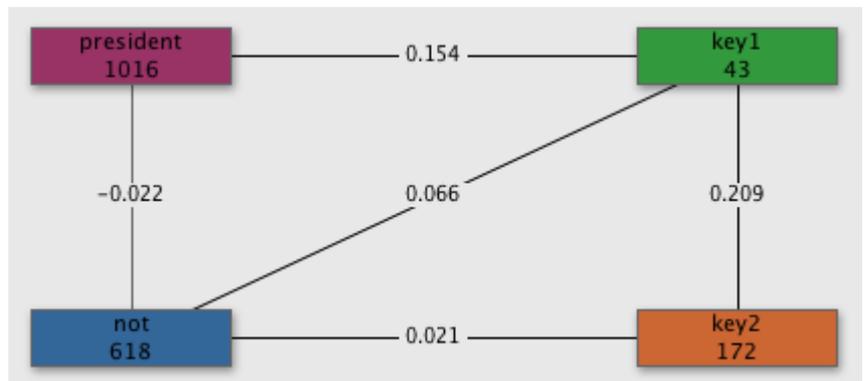
In each rectangle you see the keyword, and the total number of words belonging to that keyword. The lines between the fields show the total number of words that are common to both keywords.

- ▶ Hold the mouse down on a rectangle, to move the rectangle to a different position. If connecting lines happen to overlap, position the rectangles in such a way that the overlap is absent or minimal. The graph may look like the screen shot at the right (depending on the keys and their positions in the sequence text).



The 'president' and the 'not' keywords have the most words in common. This is not surprising of course, because these keywords are the most frequent ones and cover the most words.

- ▶ Click the **Options** button.
- ▶ Select 'phi' with the popup menu after 'show'.
- ▶ Check the 'Show colored fields' checkbox in the sheet.
- ▶ Click the **Apply** button.



The final graph may look more or less like the screen shot at the right. The fields with different keywords are now coloured with the same colours as used in the TEXT KEYS window. The phi value between 'president' and 'not' (-0.022) shows that the number of words both keys have in common seems to be accidental.

- ▶ Save the file as 'Bush.sv6'. We will use it later again.

## 8.2. Time keys

Time keys concern audio fragments and episodes of movies. Time keys are edited very much like text keys, but there are a few differences. We will demonstrate time keys with a short movie.

- ▶ Open the file 'SVdemo\_1.sv6' and go to sequence number 3.
- ▶ Click the 'link new AV file' button  in the 'linked AV files' section (below the horizontal scrollbar of the main sequence window).
- ▶ Link the file 'Sequence 3 movie.mp4' in the 'tutorial files' folder to sequence 3.

You may note that you now have two different files linked to this sequence, an audio file, and a video file. You can link up to four different files to a sequence, however these files should be perfectly synchronized (as is the case here). You cannot have both files open at the same time; if you select one of both files, the other one is closed first.

- ▶ Select **Time keys** from the **Keys** menu to open the TIME KEYS window.
- ▶ Move the key field slider, and observe that the frame of the movie corresponding to the position of the slider is shown. The number in the upper left corner of the TIME KEYS window, is the time point.

In addition you may observe that (like the text key window), words in the sequence text field of the main sequence window are highlighted. The highlighted word only *roughly* corresponds to the time that word is spoken. Which word is highlighted is inferred from (1) the position of the time keys slider (that is, the time point), (2) the duration of the event, calculated from its onset and offset times (hence, if there are no onset and offset times, no words will be highlighted), and (3) the number of words of the event text.

### Adding a new time key

If you drag the key field slider, you may observe that at time point 21, the interviewer shortly turns his head in the direction of the respondent. Let's assign a time key to this behaviour. To add a new time key, you should *first* go to the time point at which the new key will start, before you click the **New** button in the TIME KEYS window.

- ▶ Go to the frame when the interviewer starts turning his head. You can use the key field slider of the Time keys window, but for precisely finding the correct frame, you can also use the  buttons of the PLAYER window. The time is about 21.
- ▶ In the TIME KEYS window, click the  button.
- ▶ Type a name for the key in the drawer, for example 'IntToResp' (for 'Interviewer looks at Respondent').
- ▶ Move the key field slider to move to the frame when the interviewer stops looking at the respondent (about 25). You cannot use the  buttons of the PLAYER window now, but you can zoom in on the fragment in the time keys field with the horizontal slider below the time keys field. Then click the  button.

The key appears in the time keys field.

#### *Other differences with text keys*

Except that adding a new time key is slightly different from adding a new text key (you have to move to the start of the new key before clicking the  button, which is not necessary for adding a new text key), another difference is that at the bottom of the TIME KEYS window, you can find a check box 'Play selected key'. If this checkbox is checked, and you select a time key, the episode corresponding to this key is automatically played. If you click the play  button in the TIME KEYS window, playing starts at the beginning of the selected key.

A more fundamental difference between text keys and time keys concerns the duration of keys. A text key from word 2 to 5, concerns 4 words; hence the length of the text key is 4. However, a time key from time point 2 till time point 5 takes 3 time units (like the time between 2 AM and 5 AM is three hours). As keys must have some length, a text key from word 3 to 3 is very well possible, but a time key from time 3 to 3 not, as its length would be 0.

## 9. Word links

Another data type in SEQUENCE VIEWER are word links. Word links are primarily intended for qualitative purposes. No quantitative analyses are available yet for word links, but this may change in future versions.

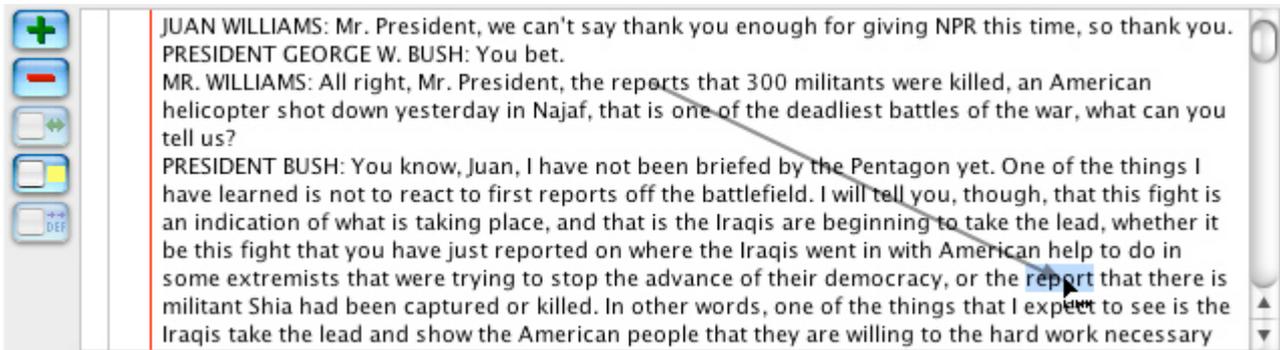
- Close 'SVdemo\_1.sv6' and open 'Bush.sv6' (see section 8.1). If you didn't saved the file, you need to repeat the first four steps from section 8.1 to recreate the file. It's not necessary to add text keys.

In the first question of Mr. Williams, he talks about 'the reports that 300 militants were killed'. In his answer, President Bush refers to this report: 'the report that there is militant Shia had been captured or killed'. You can add a word link to the text, to indicate that there is some relation between both text parts. Word links always concern single words, so you have to decide which words in both text parts you will use for the link. Let's take 'reports' respectively 'report'.

- Check the 'show word links' checkbox in the word links section at the bottom of the main sequence window. The  button becomes enabled.

- Click the  button

- Click at the word 'reports' (see screenshot below), hold the mouse down and drag the mouse to the word 'report' (see screenshot). When you move the mouse over a word, that word becomes highlighted, to assist you in selecting the correct word. The pointer changes to  when you move the mouse over the sequence text. An arrow connects both words. Leave the mouse.



- Move the mouse over the arrow. As soon as the arrow becomes surrounded by a black border, double click the arrow.
- At the right side of the main sequence window a drawer appears. Here you can enter a meaningful name for the word link, for example 'Reports'. You can add more information in the 'Info' field, for example 'reports about Shia killing'..
- Select '2' with the popup menu after 'Arrows'.
- Click at 'Arrow color' and select a colour for the arrow.
- Click .

The arrow is adjusted according to the changes you made.

- Click at the  annotations button at the left of the main sequence window to show the *notes* window.
- Move the mouse over the arrow. The notes window shows the information you just entered.

If you scroll the text, the link becomes invisible as soon as the start or end point becomes hidden. For this reason, it may be useful to resize the window when you are working with word links. If you add or delete text between begin and end of the word link, the link remains attached to the original words ('report' en 'reports'). If you delete one of both words, the word link is gone. You can however edit a word of the word link itself, as long as you don't delete the whole word. If you enter a space between two characters of the word of a word link (making two words of the word), the word link is attached to the first word. So in general, you can safely edit the sequence text, without disturbing the word links. Of course word links are saved together with your file.

## 10. Execution scripts

Most windows in SEQUENCE VIEWER concern *setup* windows. Such a window defines a setup for an analysis or a data manipulation. When you click the OK or similar button, the setup is executed (examples of windows that are not setup windows are the EDIT SEQUENCE window, the TEXT KEYS and TIME KEYS windows, and of course the main sequence window).

A setup usually consists of a number of different parts. For example, if you perform a tree analysis, you define the start code, the tree depth, the minimum frequency, and whether or not only one branch per sequence is taken into account. In addition, you may apply a sequence definition, or select a particular subset of sequences on which the analysis should be performed. Finally the setup is executed by clicking the OK button.

Such a setup can also be described in a number of text lines, or (execution) *script*. Such a script can be saved to disk, and reloaded and executed in a next session. A script may concern the setups of a number of different setup windows.

### 10.1. Creating a script with the Command Editor

In our example, we will create a new file; hence if you have a SEQUENCE VIEWER file open, close it first.

- ▶ Select **Command editor...** from the **Edit** menu to open the COMMAND EDITOR window.
- ▶ In the COMMAND EDITOR window, select **File** from the popup menu after 'Select'. This popup menu lists the menus of the menu bar (a menu item 'Sequence window' is added to the popup menu to account for commands available via the main sequence window).

The second popup menu ('Select menu item') will now show the menu items, available from the **File** menu (not all items are available, as these are not covered yet by the Command editor). Menu items usually correspond to setup windows in SEQUENCE VIEWER.

- ▶ In the COMMAND EDITOR window, select **New** from this second popup menu.
- ▶ Click the **Enter** button (to the right of the second popup menu). In the *script field*, the line (in this case a *menu command* line) **File : New** appears, preceded by a tab. You may also observe that the second (and empty) line of the script field becomes highlighted.
- ▶ In addition, the NEW FILE window appears, asking you about the number of code variables and whether you want to add event times. Choose whatever you want (e.g. 3 code variables and adding event times). *Do not press the 'OK' button in the NEW FILE window.*
- ▶ Click the **Record** button in the upper right corner of the COMMAND EDITOR window.

A number of script lines (*setup lines*) are added to the script, each line describing the value of a particular control. Please observe that the script lines reflect the setup from the NEW FILE window. The last setup line always concerns the push button that should be pressed to execute the setup of a window.

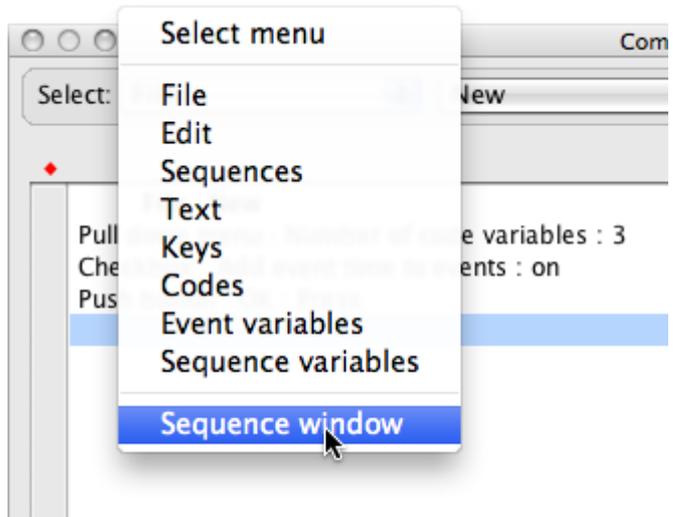
- ▶ Click the **Cancel** button to close the NEW FILE window.
- ▶ Be sure that the last (and empty) line of the script field is highlighted and select **Sequence window** from the first popup menu (after 'Select'). The second popup menu now shows a number of possible actions, available from the main sequence window.
- ▶ Select **Mark sequence** from this second popup menu.
- ▶ Click **Enter** to add this command line to the script lines.

The command line **Sequence window : Mark sequence** has no associated window, and hence no associated setup lines; it is an executable command by itself.

- ▶ Now select **Sequences** from the first popup menu and **Add sequences** from the second popup menu. Click **Enter**.

Although the command line **Sequences : Add sequences** is added to the script, you also obtain an error message, telling you that the menu item ADD SEQUENCES is disabled (you may check that from the menu bar). The reason is of course that no Sequence viewer file is open to which sequences can be added. So we first have to execute the previous part of the script, in order to create a Sequence viewer file.

- ▶ Uncheck the 'Test' checkbox (in the lower part of the window, left to the **Step** button).



- ▶ Click at the very first line in the setup field (**File : New**) to select this line. Execution will start with the highlighted line.
- ▶ Click at the **Step** button a number of times in succession, and observe that all actions according to the setup lines are executed (the NEW FILE window may be hidden after the COMMAND EDITOR window; move the COMMAND EDITOR window to be able to observe the changes in the NEW FILE window).

After 'Push button : OK : Press' is executed, the NEW FILE window is closed, and a new 'untitled' Sequence Viewer file is created.

- ▶ Click the **Step** button once again to mark the one sequence from the new file.

Observe that the one sequence of this file becomes marked after execution of **Sequence window : Mark sequence**.

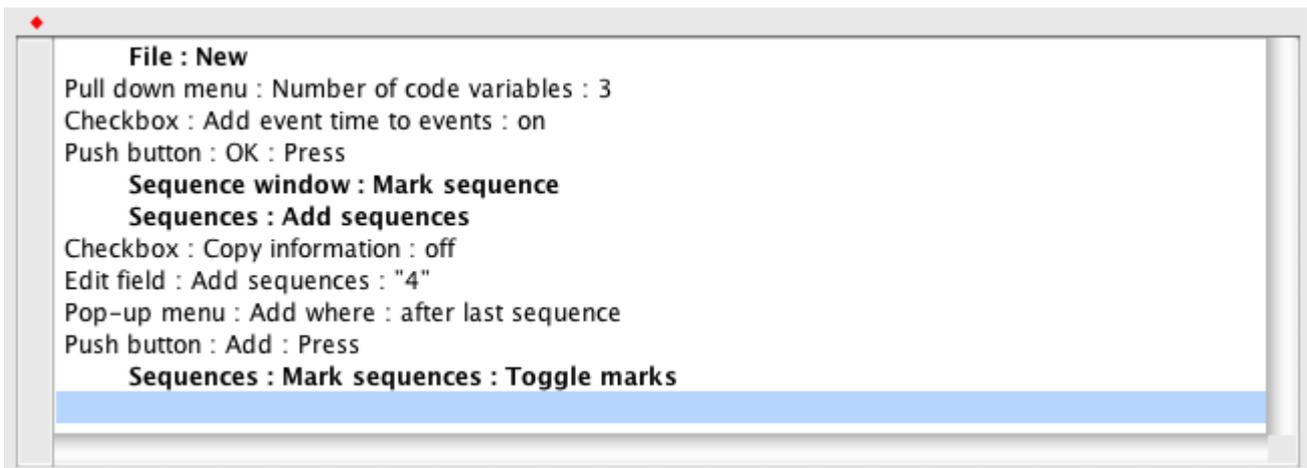
- ▶ Click the **Step** button to execute **Sequences : Add sequences**.

The 'Add sequence' window is opened. Now we can continue with the creation of our execution file.

- ▶ Create a setup for ADD SEQUENCES: leave the 'Copy information' checkbox unchecked; add 4 sequences; and add these sequences after the last sequence. Just create the setup, don't Click **Add**!
- ▶ Click the **Record** button in the COMMAND EDITOR window (if the **Record** button happens to be disabled, just click at the empty line below **Sequences : Add sequences**, even if this line is already highlighted). Again some script lines are added to the script field.
- ▶ Leave **Sequences** from the first popup menu in the COMMAND EDITOR window selected, and select **Mark sequences** from the second popup menu.

A third popup menu appears, to account for the sub items of the 'Mark sequences' item.

- ▶ Select **Toggle marks** as the sub item from this third popup menu and press **Enter** to add this command line to the script lines. Close the ADD SEQUENCES window with **Cancel**. The script lines should look like the screen shot below.



- ▶ Leave **Sequences** from the first popup menu selected, and select **Delete sequences** from the second popup menu. Click **Enter**.

Again you obtain an error message, telling you that 'Delete sequences' from the menu bar is disabled. The reason is not that you have other windows open that may conflict with the 'Delete sequences' window (as the error message suggests). Now the reason is that the Sequence Viewer file you just created only has one sequence. Hence deleting sequences does not make sense (a Sequence Viewer file always has at least one sequence), and that's why Sequence Viewer disables the menu item. Of course, after execution of the setup thus far, there will be more sequences, but Sequence Viewer just checks the present situation. Hence, again we have to execute part of the execution file, to be able to proceed with recording window setups.

- ▶ Click at the line **Sequences : Add sequences** in the script field to select this line.

To execute the setup from this line, you may use the **Step** button as described above. Alternatively, you can use the **Trace** button, which is similar to pressing the **Step** button repeatedly. This will usually give you a lot of screen activity (but also informs you about which part of the script is running, in case of long scripts).

- ▶ Execute the part of the script using **Trace**.

After execution of the script, the 'Delete sequences' window appears. Now we want to delete the marked sequences (you may have figured out that the preceding setup creates a file with five sequences, the last four marked). So select the 'All marked sequences' radio button in the 'Delete sequences' setup window, and click the **Record** button in the COMMAND EDITOR window (if the **Record** button is disabled, just click once at the 'Sequences : Delete sequences' setup line).

▶ Click the  button to close the 'Delete sequences' window.

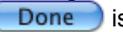
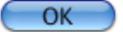
You're done with the script! To execute the whole script, close the 'untitled' file you just created (without saving). Select the very first line of the script (**File : New**), and click . The 'Run' button executes the script, starting at the highlighted line, without feedback (although progress windows may appear in case of tedious analyses).

You may save the script to disk, with the 'Save' button, and load it again with 'Load'. You should first create a folder called something like 'SV scripts', e.g. in the same folder where the SEQUENCE VIEWER application resides, to store execution scripts. In the Sequence Viewer preferences ('editor' tab), you should set this folder as the default folder for scripts. This has the effect that if you click 'Load', this folder will be shown in the file selector window. You can now close the 'untitled' file, without saving.

## 10.2. Auto recording scripts

A second way to create scripts, is let SV automatically record your actions. The next example shows you how.

- ▶ First clear the script field. Click at any of the bold lines in the script, hold the option key down, and click the  button.
- ▶ Close the COMMAND EDITOR window, if still open (the command editor may ask you to save any unsaved scripts).
- ▶ Close the newly created SV file if still open and open the file 'SVdemo\_1.sv6'.
- ▶ Click the Record button  in the toolbar (if this part of the toolbar is hidden, select Edit > Toolbar > Show auto recording). The recording light (below the record button) is turned on (becomes red), whereas the stop button  becomes enabled.
- ▶ Select **Frequencies** from the **Codes** menu. You may notice that right to the  button, the red recording light tells you that your setup is automatically recorded.
- ▶ Option-click the Sequence definition button  to open the SEQUENCE DEFINITION window. Select Sequence definition 1, check 'Mask events', and check the third checkbox (thus masking the 'adequacy' code variable). Click .
- ▶ Click the  button of the FREQUENCIES OF EVENT CODES window. The analysis is performed and the results are shown in the output window.
- ▶ Click the Stop button  to stop auto recording. The COMMAND EDITOR window opens (if not, select **Command editor...** from the **Edit** menu to open the COMMAND EDITOR window).

The script field of the COMMAND EDITOR window shows you that all your actions are recorded. You may note that although you first opened the FREQUENCIES OF EVENT CODES window, the script concerning the sequence definition precedes the script concerning the frequencies analysis, which is as it should be: to apply the sequence definition to the frequencies analysis  is pressed in the SEQUENCE DEFINITION window *before* you pressed  in the FREQUENCIES OF EVENT CODES window.

Don't close the COMMAND EDITOR window if you want to do the exercise described in section 10.3.

### note

During a session, you can always click the Stop button  before a particular analysis, and click the Record button  after the analysis, to prevent that this analysis is auto recorded. If you click the Stop button , the COMMAND EDITOR window is opened. You don't need to close the COMMAND EDITOR window before continuing auto recording, but can click the Record button  with the COMMAND EDITOR window open. In that case the COMMAND EDITOR window becomes locked, although you can observe all recording actions.

## 10.3. Auto running saved scripts

You can run the recorded script from the COMMAND EDITOR window with the  button, but there is a different way to run scripts, which is especially useful if you often run the same script.

If you haven't done this already, you should first create a default folder for your script files:

- ▶ Open the 'Preferences' window and click the 'editor' tab.
- ▶ Click the 'Browse' button, select the folder containing your scripts (or create a folder) and click .

Setting a default folder where your scripts are saved is obliged, if you want to make use of the options discussed below!

- ▶ Go back to the COMMAND EDITOR window and click the  button. Be sure that the script file is saved in the default script folder (this folder will be automatically shown)! Call the script file 'Example.txt' (or whatever you like).

- ▶ Close the 'Edit commands' window.
- ▶ Click the Stop button , if auto recording is still on.
- ▶ Go to the toolbar and click the  button. If you cannot find this button, maybe this part of the toolbar is hidden. Select Edit > Toolbar > Show edit script menu to make it visible.
- ▶ A drawer appears. Click the  button in this drawer.
- ▶ Select the 'example.txt' file you just saved.
- ▶ Type a name in the field after 'Script name', for example 'code frequencies'.
- ▶ In addition you can type a larger descriptive text in the field after 'Description', for example 'adequacy masked'.
- ▶ Click .
- ▶ In the toolbar, select 'Code frequencies' with the 'Select script' popup menu.
- ▶ Click the  button in the toolbar.

The scripts now runs outside of the command editor.

*note*

The  button is disabled if auto recording is on, or the COMMAND EDITOR window is open.

# 11. Waveforms

You can find a general introduction to waveforms in the help file of SEQUENCE VIEWER, or the SEQUENCE VIEWER manual. Please read this introduction first before you continue.

## 11.1. The waveform window

In this section we will first have a quick look at the different parts of the WAVEFORM window.

- ▶ Open 'SVdemo\_2.sv6'. This file is similar to SVdemo\_1.sv6, except that the fourth sequence you added earlier to the file SVdemo\_1.sv6, is already transcribed and coded, and includes onset and offset times.
- ▶ Go to sequence 4.
- ▶ Select AV file 'Sequence 4 sound.aif'.
- ▶ Select 'Waveform...' from the 'Edit' menu.

The WAVEFORM window is opened, and after a few moments, the waveform of the sound is drawn.

- ▶ Click the 'Assign event times' tab (if not selected yet).

In the upper part of the window some information can be found about the sound file; for example that the sample rate is 22050, and the sample format is 16 bits. Below this information you find a number of controls. At the right you find the channels popup menu. If you have two channels (stereo), you can mix both channels, or display either one channel. Because the present file has only one channel, this channel is displayed and the setting of the popup menu does not have effect.

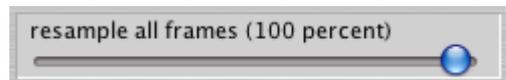
At the left of this popup menu you find the 'Normalize' check box.

- ▶ Check the 'Normalize' check box. The amplitude of the waveform is now set to its maximum, which makes easier to see the differences between soft and loud parts of the sound. *Analyses are affected by the setting of this checkbox, in particular loudness and silence analyses.*



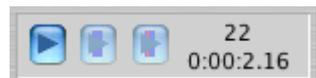
At the left of the 'Normalize' checkbox is a slider. With this slider you can change the resample rate. The original sample rate of the sound is 22050; the resample rate is set to 10 percent, that is 2205. Because of his low resample rate, the waveform is displayed much faster.

- ▶ Drag the resample slider to 100 percent.



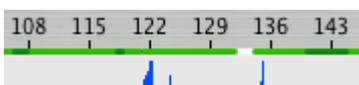
Now it takes much more time to draw the graph, whereas there are only small differences! *However, the setting of the resample rate will affect the results of analyses, most notably the pitch analyses.* For reliable pitch analyses, you should not use a resample rate lower than 7000 Hz, if human speech is involved. You can find the resample rate in the right part of the information window.

- ▶ Drag the resample slider to 33 percent. The resample rate is now 7350. Leave the sample rate at this value, until you are told to change it.
- ▶ Now click the play button  at the left. The sound starts playing, and the button changes into . The waveform slider with the red vertical line, moves together with the sound. We call the waveform slider plus vertical red line, the *time slider*. Click  to stop the sound.



- ▶ Go to the main sequence window, and click at one of the event codes, for example the first on, IQA.

The corresponding part of the waveform becomes selected. The selection is based on the onset and offset times of the event. If no event times are available, clicking an event code has no effect.



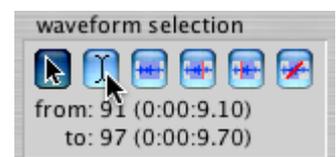
The part between onset and offset time of an event has a green coloured bar at the top of the waveform field. A gap means that there is a pause between the offset time of an event and the onset time of the next event. There is one such gap at time 133.

Dark green bars means that there is overlap: the offset time of an event exceeds the onset time of the next event. There may also be red or black bars; these indicate that the onset time of an event exceeds the offset time of the same event (that is, the duration is negative, which should be corrected of course). If the bar is black, there is also overlap with another event.

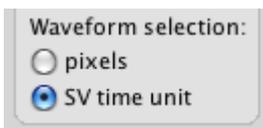
- ▶ Click the play selection  button. The selected sound is played. The sound stops when the end of the selection is reached.



- ▶ Find the 'waveform selection' box at the lower right corner (the box with the six icons). Click the  button and go to the waveform. You can now select any part of the waveform you wish.



If you select part of the waveform, you may observe that the selection is adjusted a bit. The reason is that the radio button 'SV time unit' is selected. You can find this radio button

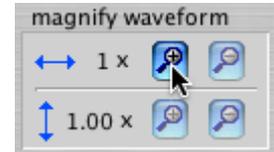


in the lower right part of the 'tab' section. This radio button is useful if you want to assign event times using the waveform, because it doesn't make sense to make more precise selections than the time unit, when assigning onset and offset times. You can also choose 'pixels' as the unit of selection. In the three analyses modes (loudness, silence and pitch), always pixels are used to select part of the waveform.

The button selects the whole waveform; the button selects the waveform from the beginning of the sequence until the time slider, whereas the button selects the waveform from the time slider until the end of the sequence.

Click the deselection button to get rid of any selection.

Find the 'Magnify waveform' box at the lower right corner. Click the button after the horizontal double blue arrow a few times to zoom in on the waveform, at the position of the time slider.



The vertical magnification buttons are disabled, if the 'Normalize' checkbox is checked. The amount of vertical magnification does not affect analyses however, unlike checking the 'Normalize' checkbox.

## 11.2. Sound analyses

### Loudness analysis

Open 'SVdemo\_2.sv6' and go to sequence 4.

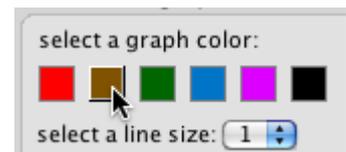
Set the resample rate to 33 percent and check the 'Normalize' button.

Select 'Sequence 4 sound.aif' as the linked AV file.

Click the 'Loudness analysis' tab.

Click to select the whole graph.

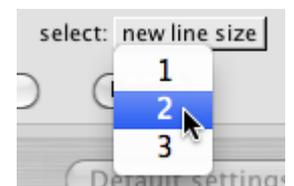
Click at one of the colours to select a graph colour. The selected colour has a small white border around it. Select a line size of 1 with the popup menu.



Click the **Get loudness** button.

After a few moments a loudness graph is drawn. As you can see, the graph nicely corresponds with the amplitude of the waveform. The colour and line size of the graph depend on the selected colour and line size. Especially the colour is important, because it enables you to select a particular graph as we will see.

Be sure the same colour is selected as the colour of the graph. Click at the 'new line size' pull down menu and select a new line size of 2. This has the effect that the line size of all loudness graphs with the selected colour is changed to 2. Don't confuse the popup menu after 'select a line size' (the line size for new graphs), with the 'new line size' pull down menu (to change the line size of existing graphs).



Select a different colour with the 'new colour' pull down menu, preferably black (unless this colour was already selected of course). Be sure the selected line size is still 1. You may notice that the 'new colour' and 'new line size' pull down menus become disabled: there are no graphs with a black color.

Change the resample slider to 100 percent.

Click the **Get loudness** button. The analysis clearly takes more time. The new graph is drawn over the old one. It is apparent that despite the higher resample rate, there is not much difference.

Press the button a few times, to be better able to observe any differences. Use the horizontal scrollbar at the bottom of graph field, to inspect the whole graph. As you can see, the differences are very small

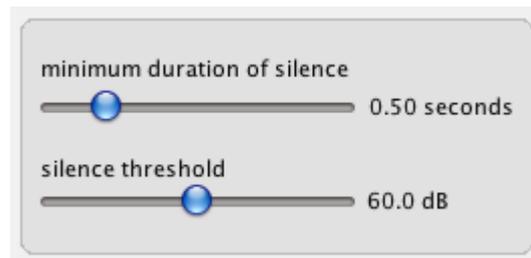
Change the resample slider to 10 percent, select a different colour again, keep an eye on the statistics and click the **Get loudness** button. Although the differences are a bit larger, they are still quite small, as is also shown by the differences in statistics. These statistics are expressed in dB (decibels). If the 'Normalize' checkbox is unchecked, these values would have been a bit smaller.

Select one of the graph colours and click the **Delete** button. The graph with this colour is deleted. Delete all loudness graphs you have created, except one. We need this graph for our next analysis.

### Silence analysis

Be sure the 'Normalize' button is checked, the resample rate is 33 percent, and there is one loudness graph. A loudness graph is not necessary for a silence analysis, but better illustrates the effect of particular settings, as we will see. And click at the button until the magnification is '1 x'.

- ▶ Click the 'Silence analysis' tab. The lower part of the loudness graph disappears (is hidden).
- ▶ In the lower left corner of the WAVEFORM window, is a checkbox, called 'Show settings'. Check this checkbox (if not already checked) to obtain the settings drawer.
- ▶ Move the 'silence threshold' slider to the left and to the right, and watch the effect on which part of the loudness graph becomes hidden.



As you already guessed, the setting of this slider defines what should be viewed as a silence. The other setting is used to define the minimum length of a silence.

- ▶ Set the minimum duration to 0.50 seconds, and the **silence threshold** to 60 dB. In addition, select a colour and set the line size to 3. Select the whole waveform. Click the **Get silence** button.

The silence 'graph' consists of bars at the bottom of the graph field. There is only one instance of a silence detected (between time 82 and 88).

- ▶ Delete the **silence graph** you just created, set the minimum duration to 0.20 seconds. Leave the silence threshold at 60 dB. Click the **Get silence** button.

You should now have six silences. You can click the  button to select pieces of the waveform, where silences appear and play the corresponding sound. Or select other 'gaps' in the loudness graph, not viewed as silences, to determine whether or not these are 'real' silences. Adjust the sliders of the silence settings until the bars correspond to what should be viewed as silences. If you are going to analyze a large number of sequences (to assign sound characteristics to event variables), that use sound files with the same quality and sound level (for example if the same sound file is used for a number of sequences), you can safely use these settings for these other sequences.

- ▶ Delete all silence graphs, uncheck the 'Normalize' checkbox, leave the minimum duration at 0.20 seconds and the silence threshold to 60 dB. Click the **Get silence** button.

No silence bars at all! Whether or not 'Normalize' is checked, clearly affects what is viewed as a silence. Because the not normalized sound is less loud, the silence threshold has a different effect. To obtain silence bars again, you may increase the threshold to for example 62.0 or 62.5 dB. If you analyze different sequences with different sound files, that may have different sound levels, you can check 'Normalize' to make the sound level of different sound files more comparable. Generally you should not normalize the sound if the different sequences use the same sound file.

### Pitch analysis

Again, be sure the 'Normalize' button is checked, the resample rate is 33 percent, and the magnification is 1x, but delete all silence and loudness graphs. To delete a loudness graph, you need to select the 'Loudness' tab first.

- ▶ Click the 'Pitch analysis' tab.
- ▶ Check 'Show settings'.
- ▶ Select red for the line colour and 1 for the line size.
- ▶ Click the **Default settings** button (to be sure that results are similar to those we will discuss here).
- ▶ Select the whole waveform with , and click the **Get pitch** button.

The analysis takes a few moments. When the analysis is finished, the first thing you can observe is that there are lots of gaps in the pitch graph. Silences, and very short pieces of speech are not analyzed. Particular sounds like 's', or fragments with too much noise relative to the speech, are analyzed neither. We will call the part of a waveform that is assigned a pitch, a fragment.

- ▶ Play the whole audio file. With one exception (at time 119-120), it is clear that female voices are associated with a higher pitch than male voices.
- ▶ Move the time slider to point 120 and magnify the waveform 4 times with . Click  and select the waveform between about time point 115 and 125.
- ▶ Play the selected part of the waveform with . It is clear that the word 'very', spoken by the (female) interviewer, is assigned a fundamental frequency that is more like the fundamental frequency of the voice of the (male) respondent.
- ▶ Set the resample rate to 100 percent, select black for the graph colour and 1 for the line size.
- ▶ Click  to select the whole waveform, and click the **Get pitch** button again.

As you can see, the 'black' graph partly overlaps the 'red' one, but there are also some differences.

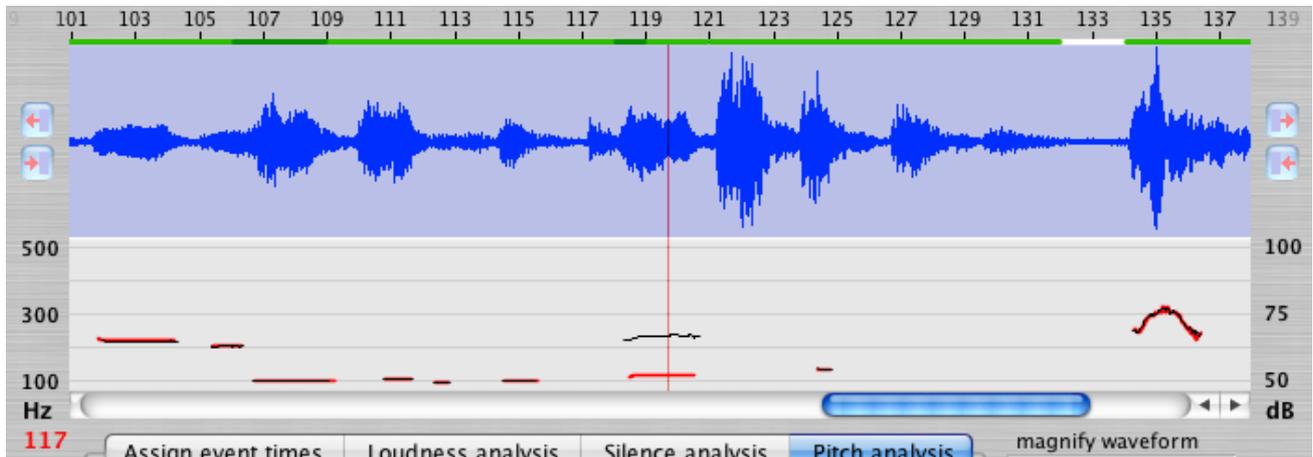
- ▶ Select the red colour and change the line size of the 'red' graph with the 'new line size' pull down menu to 2.

Now it is much easier to compare the differences between both analyses! This is very important in order to decide on the best settings.

▶ Set the magnification to 4x, and scroll through the waveform.

The differences between both graphs are not very large, with a few exceptions. Some fragments show a 'red' graph, but no 'black' graph. More important however is that the fragment at time 119-120 has both a 'red' and a 'black' graph. Because the fragment concerns a female voice, it is clear that the 'black' graph, does a better job in this respect.

▶ Click the red colour to select the red graph. Move the time slider over the red graph between time point 118 and 121. Below the Hz scale at the left of the graph field the fundamental frequency of the waveform at the position of the time slider is shown (in red) according to the red graph.



▶ Click the black colour to select the red graph. Move the time slider over the black graph, between time point 118 and 121. Now the fundamental frequency of the waveform according to the black graph is shown.

As you can see, the fundamental frequency of the black graph is about twice as high as the fundamental frequency of the red graph. This is a common problem in finding the fundamental frequency in waveform analysis (see also the section on 'Pitch correction' in the help file or the Sequence Viewer manual).

▶ Set the resample rate to 33 percent, select blue for the graph colour and 1 for the line size. Change the 'low pitch correction' in the settings drawer to 25. Click the **Get pitch** button.

▶ Click  to select the whole waveform, and click the **Get pitch** button once again.

The blue graph now shows the correct fundamental frequency for the fragment between 118 and 121. Moreover, the fragments between 7 and 9 are also assigned a pitch. With this simple correction we obtain a clearly better graph.

Deciding on which settings should be used to obtain the 'best' pitch graph can often be tedious, especially because the setting on one type of correction may influence the effect of the setting on another type. The high pitch correction should be used if a fragment obtains a too high pitch (like this example). Noise correction affects whether less clear fragments should be assigned a pitch value or not. If this correction is (near) 0, you probably obtain no graph at all. If the noise correction is too high, you may obtain the wrong pitch for particular fragments. In general, follow these steps:

1. Use the default settings.
2. Find fragments with a too low pitch.
3. Adjust the low pitch correction to the *lowest* value that corrects for the error.
4. If correction is not possible, set the noise correction to a lower value.
5. Do the same for the high pitch correction.
6. If you didn't adjust the noise correction, you may try using a higher value.

In some cases it is not possible to correct each and every fragment. Our last resort is using the other three settings. Problems in detecting the correct fundamental frequency may be caused by overlapping speech. If you check 'Exclude overlap and pauses', the part between the offset time of an event and the onset time of the next event, are excluded from the graph. If the offset time is less than the onset time of the next event, this is a pause. Usually these parts are already excluded (as there will be no speech). If the offset time exceeds the onset time of the next event, there is apparently overlapping speech and will be excluded too.

Another problem is that especially for very short fragments it may be difficult to calculate the correct fundamental frequency. If short fragments have wrong pitches, you may increase the minimum duration of a fragment.

And a final problem is that also relatively soft fragments may cause problems (because the noise will be relatively high). You can exclude fragments with a low average loudness with the 'minimum loudness' slider. This correction is not the same as the silence threshold we used for the silence analyses: the minimum loudness concerns the *average* loudness of a whole fragment; thus particular parts of the fragment are very likely less than this minimum.