THEME 6 User's Manual

By Magnus S. Magnusson

Work in progress.

Please send comments and suggestions to patternvision@patternvision.com

Table of Contents

History

In continuation of Magnusson's Medal Thesis on Ethological (biology of behavior) methodology at the University of Copenhagen in 1978, he developed THEME from the late 70's for "Temporal Configuration Analysis" (later called T-Pattern Detection and Analysis or TPA). His first THEME program was a three thousand lines Fortran IV program running on a PDP 8 computer at the Psychological Laboratory of the University of Copenhagen around 1980 and the first results were presented at an Artificial Intelligence workshop in 1981 at Uppsala University, Sweden. In 1995 an international inter-university research collaboration network was created around "Magnusson's analytical model" now with over 24 member universities and in such international context of his development and research application of the T-Pattern model and its extensions (the T-System), Magnusson has since 1983 continued the development of THEME™ at the National Museum of Natural History, then at the University of Paris (V, VIII and XIII) and finally at the Human Behavioral Laboratory that he founded and directs at the University of Iceland. THEME™ is now a 300 thousand line Delphi program for Windows (7 and up) that runs on modern PCs.

Useful References

With links.

Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. Behavior Research Methods, Instruments, & Computers (2000) 32: 93. doi:10.3758/BF03200792

Anolli, L., Duncan S. Jr., Magnusson MS and G. Riva. (2005) The Hidden Structure of Interaction: From Neurons to Culture Patterns. Approx. 300 pp., hardcover. ISBN: 1-58603-509-6

Each chapter describes a different research application of T-Pattern Analysis with THEME™

Casarrubea M, Jonsson GK, Faulisi F, Sorbera F, Di Giovanni G, Benigno A, Crescimanno G, Magnusson MS. (2015) T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. J Neurosci Methods.

Magnusson, M.S., **Burgoon**, J., **Casarrubea**, M. (Eds.) (2016) Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME™. Springer, 2016.

Each chapter describes a different research application of T-Pattern Analysis with THEME™

Magnusson MS. (2016) Time and Self-Similar Structure in Behavior and Interactions: From Sequences to Symmetry and Fractals. Chapter 1 in Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME™. Springer, 2016.

Magnusson MS (2017) Why Search for Hidden Repeated Temporal Behavior Patterns: T-Pattern Analysis with Theme. Int. J. Clinical Pharmacology & Pharmacotherapy 2017, 2: 128 (click to download free PDF)

See also a list of publications implicating T-Pattern Detection and Analysis with THEME™ at the Human Behavior Laboratory, University of Iceland: http://hbl.hi.is/publication

The Data

The following provides only essential descriptions, while further descriptions can be found in the sources listed above. Most similar: [Why Search for Hidden Repeated Temporal Behavior Patterns: T-Pattern Analysis with Theme.]

The kind of data, called T-Data that all definitions refer to and is treated by the T-Pattern detection algorithm are one or more point-series on a *discrete scale* within $n \geq 1$ intervals $[t_{i1}, t_{i2}]$; $i = 1..n$, $t_{i2} \geq t_{i1} \geq 0$, each series representing the positions of the beginning or ending of the same phenomenon, which may be, for example, a particular type of behavior in time or the positions of a particular kind atom or molecule along another such as DNA or proteins.

Single series within a single interval [1, T]

Fig 1

T-data Example

A T-data set of **N** *event-types* $\mathbf{E_{1..n}}$, sometimes also called *primitives*:

---

e_1 . . . . . .

e_2 . .

e_3 . . . .

.

.

e_i . . . .

.

.

e_n . . . .

---

1 T

Fig 2

The T-Pattern

A **T-pattern** is **m** ordered components, $\mathbf{X_1..X_i..X_m}$, any of which may be primitives or T-patterns, on a single dimension (time), such that, over the n occurrences of the pattern the distances $\mathbf{X_i} \rightarrow \mathbf{X_{i+1}} = t\mathbf{X_{i+1}} - t\mathbf{X_i}$ vary within a significantly small interval, called a **critical interval** (CI): $\mathbf{[d_{i1}, d_{i2}]}$,

$\mathbf{X_1\ [d_1,d_2]_1\ X_2\ [d_1,d_2]_2\ ..\ X_i\ [d_1,d_2]_i\ X_{i+1}\ ..\ X_{m-1}\ [d_1,d_2]_{(m-1)}\ X_m}$ (m = length)

The Detection Algorithm

This T-pattern Detection algorithm works **bottom-up**, **level-by-level** and uses **competition** and **evolution** to deal with redundant detections.

Any T-pattern $\mathbf{Q = X_1 X_2..X_m}$ can be split into at least one **pair** of shorter ones related by a **critical interval**:

$$\mathbf{X_{Left}\ [\ d_1\ ,\ d_2\ ]\ X_{Right}}$$

Recursively, $\mathbf{X_{Left}}$ and $\mathbf{X_{Right}}$ can thus each be split until the pattern $\mathbf{X_1..X_m}$ is expressed as the **terminals** of a **binary-tree** of primitives (event-types).

Detection works in the opposite direction. Starting with the comparison of all pairs of the series of T-Data, which may represent beginnings or endings of either event-types or (as detection progresses) T-patterns, searching for such critical **left** $[d_1, d_2]$ **right** relationships and when found connecting the two sides where they have that relationship and their occurrence series are **added to the data**.

Patterns are thus constructed bottom-up level-by-level as binary trees of detected critical intervals between primitives (event-types) and/or detected T-patterns.

Patterns are selected and evolve through a **completeness competition** where partial and equivalent patterns are removed (see below).

Detecting Critical Intervals

Relating Point Series A and B

Fig 3

Given point series A = ta_i (i = 1..na) and B = tb_j (j = 1..nb). The distances from each point ta_i in A to the first point tb >= ta_i in B are calculated searching for an interval [$ta_i + d_1$, $ta_i + d_2$] containing at least one tb and occurring after more points in A than expected by chance; assuming as a zero hypothesis that A and B are independent. A critical interval is called FAST if d1 is set to zero, that is, [d_1=0, d_2].

The search algorithm for FAST critical intervals is simpler than for the FREE critical interval, which is, however, a simple extension of the former. The FAST begins with an interval [d_1=0, d_2] where d_2 is initially set to the longest of the "A to first following or concurrent B" distance. If not significant, d_2 is set to the next longest until a significant interval is found or all distances have been tested.

The **FREE** algorithm does the same except that d_1 is initially set to the shortest "A to first following B" distance and then it proceeds as the **FAST** algorithm regarding d_2, but if significance is not found for a d_1 value it is set to the next higher and the FAST process is repeated for that value. This is repeated for each higher value of d_1 until a critical interval is found or all possible [d_1, d_2] intervals have failed to be significant.

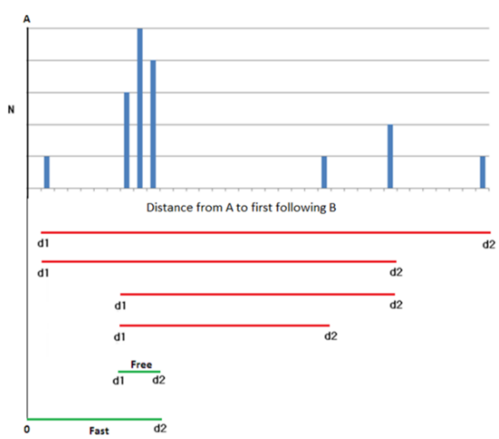


Fig 4

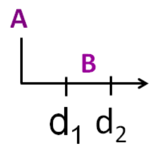


Fig 5

Evolution through Completeness Competition

All detected T-patterns are compared removing patterns that are

1. <u>Identical</u> to patterns already detected. Identical T-patterns have all the same primitives (event-types) occurrences, that is, all event-types are the same and occurring at the same occurrences times/positions.
2. <u>Equivalent</u> meaning identical except having a different binary tree structure. Then only the first one detected is kept.
3. <u>Partial</u> relative to another, meaning identical except for one or more primitives missing in the partial.

In this way only the most complete versions survive.

- ((A   B)   ((C   D)   (E   F)))
- ((A ((B   C) D)) (E   F))
- ((A   B)   (D   (E   F))
- ((A   C)   F)
- (B   E)
- (B   D)
- (A   F)

Fig 6

Validation

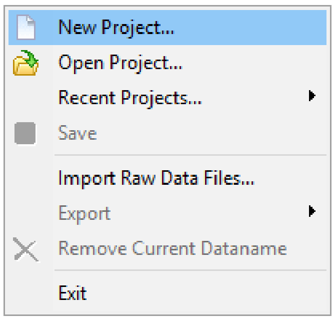
Validation is external and/or statistical; global and per pattern, using repeated simulation. See below.

The Project

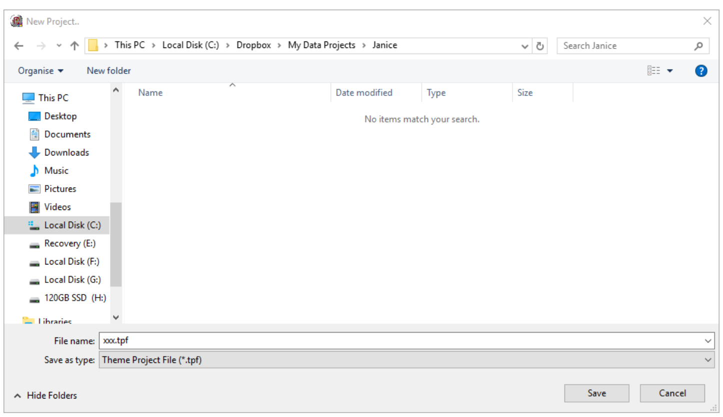
A Theme project refers to a dedicated **project directory (**which may have to be created) with one or more .txt raw data files (T-Data) and one or more **reference tables:** a **category table** and, optionally, an **independent variables table** and/or a **graph table.**

Before a Theme project can be started a **category table** with the reserved name **vvt.vvt** (vvt = variable-value table) must be stored in the dedicated project folder.

Create a new project



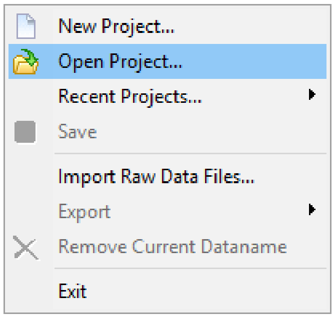
To create a new project in the dedicated project folder click: or **File**, **New Project** and



enter an name (here xxx.tpf) for your project using only standard English letters and numbers. The name <u>must begin with a letter</u>, then click **Save** in the project folder (here …Janice).

Opening an existing project

To continue work on an existing project, for example, to import, export or remove data files or to detect or analyze patterns click  or **File, Open Project…** or **Recent Projects…**



Data Entry and Preparation

Creating a category table (a reference table)

Compared to earlier Theme versions, creating a category table (reserved name vvt.vvt) has been simplified. The vvt.vvt file is a simple .txt file that can be easily created and edited using **Notepad**. It has two types of entries (only one per line): **Class** (or variable here shown in **bold** letters below) names starting at the first column and **Item** (or values or elements) starting at any other column in the line. For example, this simple table with the four classes (variables) Actors, b_e, locomotion and verbal acts each with a few values:

actors
  Sue
  Bill
  Jack
  Luis
b_e
  b
  e
locomotion
  run
  walk
  crawl
verbalacts
  question
  command
  verify

Empty lines can be added anywhere as they are ignored by the program. All letters are automatically changed to lower case. Be sure there is an empty line, a carriage return, after the last entry in the file.

The “B_E“ (short for *begin* and *end*) variable or class has this reserved name and, if used, must be the second variable in the vvt.vvt and must have only those two values „b“ and „e“ and in that order. When included, it is only used to specify whether some behavior is beginning or ending. For example, „sue,**b**,run“ meaning: sue **begins** running vs. „sue,**e**,run“ meaning: sue **ends** running. {For various reasons, it is recommended to always include this „B_E“ class even if only its „b“ value is used in the data.}

A new Theme project can only be created if a vvt.vvt file with this format is stored directly in the dedicated project folder (not in any of its subfolders).

Creating data files

The *tab delimited* (*Windows .txt*) text file format of Theme **raw data files** has only two columns: **time** and **event**,
that is, each line only has a **time stamp** (integer) and an **event code** (for example, 374 sue,b,run). Such files can be created using various simple text editors such as Windows Notepad.

Only one sample per .txt raw data file. In Theme 6, a data file can only contain one T-Data sample, that is, just one observation period.
The first line of the file contains the tab delimited column headings: **time** {tab} **event**.
Next, a line with the special time-stamped *start of observation* code "**:**" (for example, 24 {tab} :).
Then follow the time-stamped events one per line.
The last line of the file is the time-stamped *end of observation* code (**&**). See below.

It is often useful given data files (samples) **descriptive names**, for example, regarding their content such as subject name(s), when and where it was created and how (e.g. jun2015id3id7garden.txt).

Note also that Theme **automatically** orders all the files in a project **alphabetically**. The order can thus be easily decided by adding one or more letters at the beginning of the file names, for example, Ajan2015id3id7garden.txt, Bapr2015id3id7garden.txt, Cjun2015id3id7garden.txt, etc.

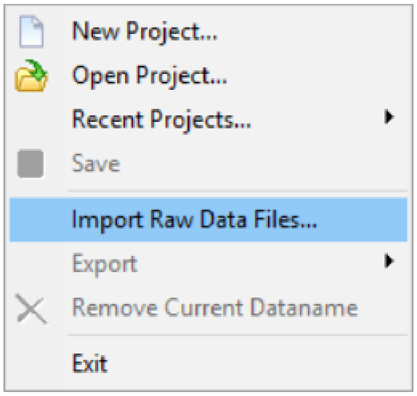
NB: **File names must not exceed X characters, only English letters, the digits 0 to 9 and must begin with a letter.**

A tiny tab-delimited Theme T-data file example:

Time	event
5	:
15	sue,b,run
17	jack,e,talk
201	sue,e,run
231	bill,b,smile
302	jack,b,talk
302	bill,e,smile
302	sue,b,smile
303	&

Observation begins at 5 and ends at 303 (the time unit is not specified, but is here likely to be, for example, a second). Note that time must be specified using only **integer**s representing the smallest discrete time value used and thus corresponding to the **time resolution**.
Any number of **different** event-types can occur at the same time, while the **same event-type** can occur only **once** at each time unit. If an event-type is coded more than once at the same time unit, only one occurrence is considered.

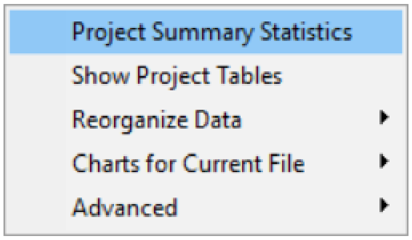
Importing data files



To import one or more files select **File**, **Import Data Files** and click **Open**

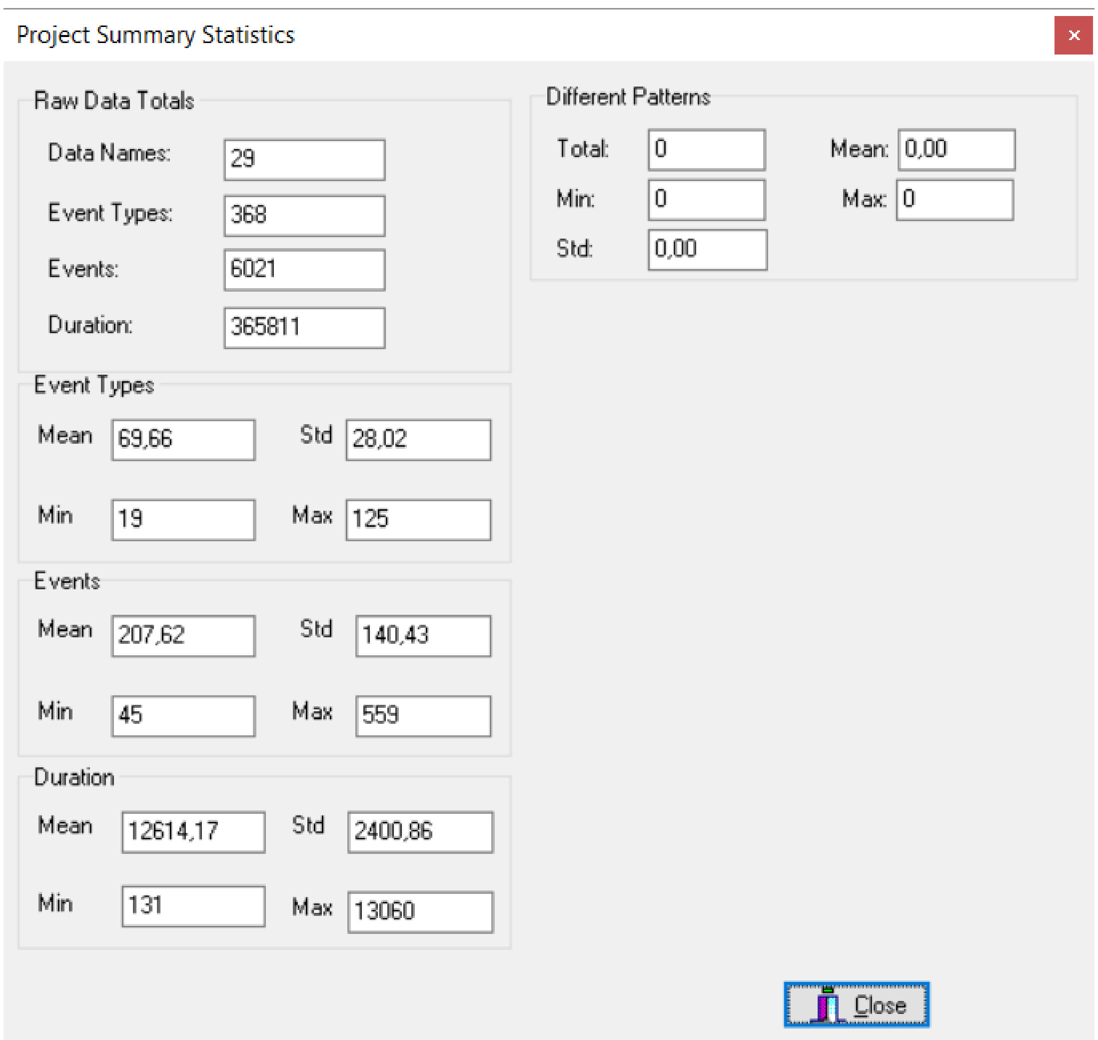
When complete, the sample file (data set) names will appear on the left hand side of the screen and to the right a table with summary information about each sample.

Project (in main menu)

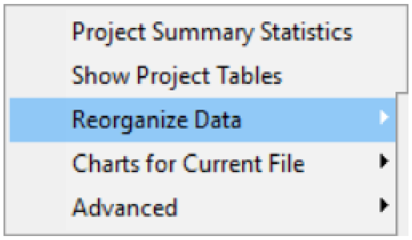


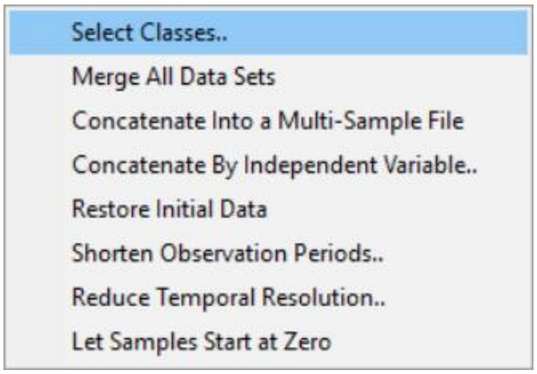
Project Summary Statistics

An overview of the raw data in a project before T-pattern detection is shown



Reorganize Data





Select Classes

Here the specification of events can be reduced by deselecting classes from the VVT table. For example, after unchecking of a class “speed” the event-type *bill,b,run,fast* becomes *bill,b,run*. Similarly: *jane,b,turn,left* after removal of a direction class becomes *jane,b,turn.*

Merge All Data Sets

All data sets in the project are merged assuming there is some overlap of their observation periods (intervals) and their section becomes the observation period of the resulting merged data set with all events over all the files occurring within that period.

Concatenate into a Multi-Sample File

Here **all** the single-samples are copied in **alphabetical order** (as in the data name list on the left) into a **single multi-sample file** that can be analyzed mostly as a single sample file, but after detection allows identification of patterns significantly more (or less) frequent within certain subsets of samples, usually corresponding to an experimental design.

Concatenate by Independent Variables

Here two or more *multi-sample files* are created according to a classification variable in the *Independent Variables Table*. See below.

Restore Initial Data

Returns the data into the initial state, that is, as just after importing the single-sample-files.

Shorten Observation Periods

The observation duration of each sample is reduced to that of the shortest.

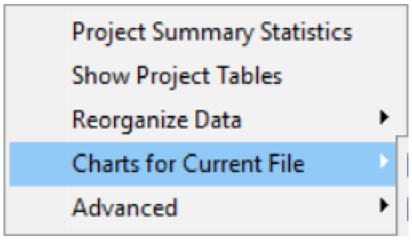
Let Samples Start at Zero

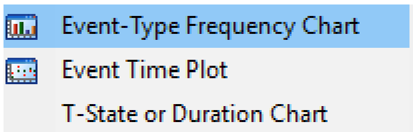
The start time of each sample is subtracted from all entries of that sample. All samples thus begin at zero.

Reduce Temporal (Positional) Resolution

Requests an integer value and all time (position) values in the data are divided by that value, for example, 10, 15, 100, 50000 or any other positive integer >0.

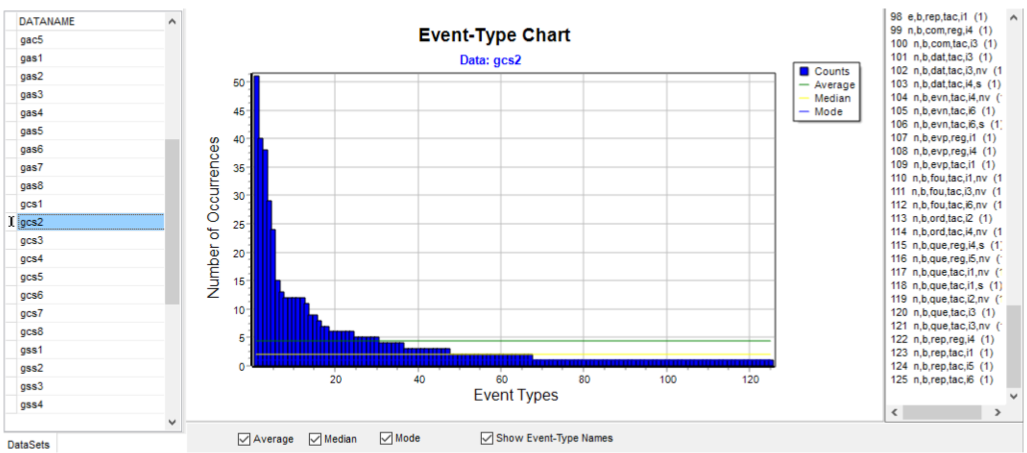
Charts for Current File





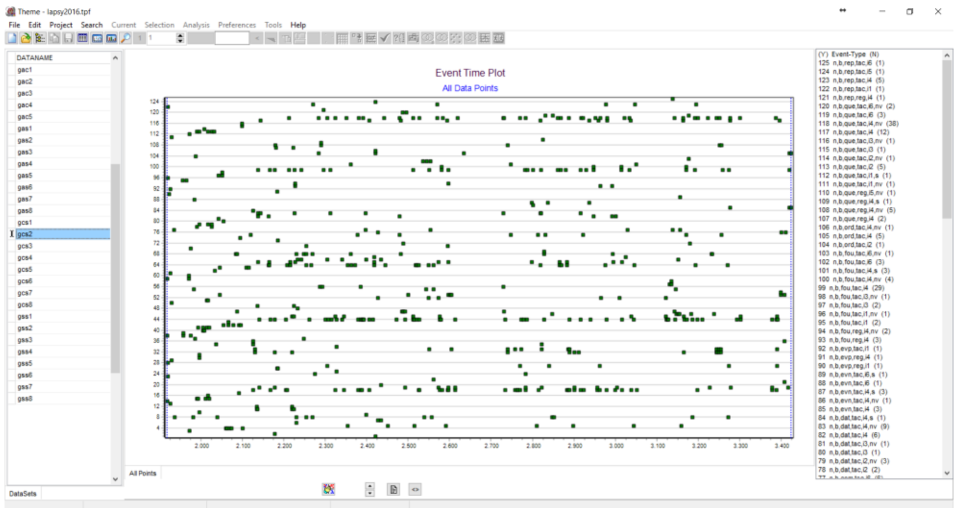
Event Frequency Chart

The Event Frequency Chart shows the number of occurrences of the event-types in the current data file. Here shown for file gcs2. Optionally shown are also average, median and mode as well as the lists of files in the project and the event-types (with counts) are here optionally shown.



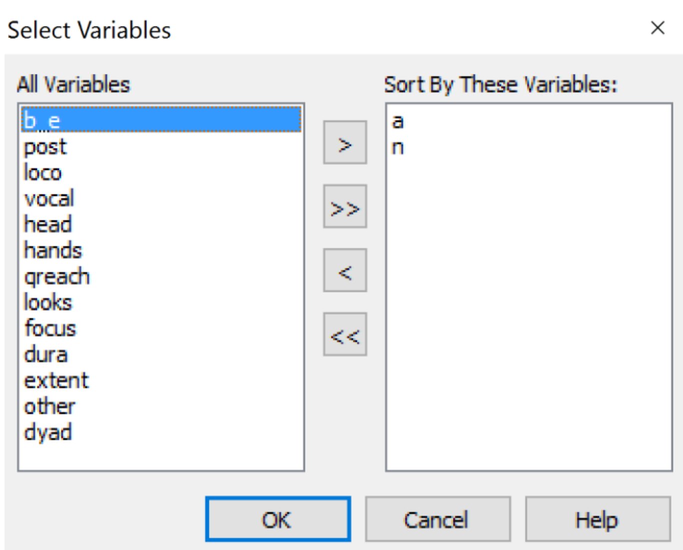
Event Type Plot

The Event Type Plot shows the (time, 1-D) locations of all events in the current data file.



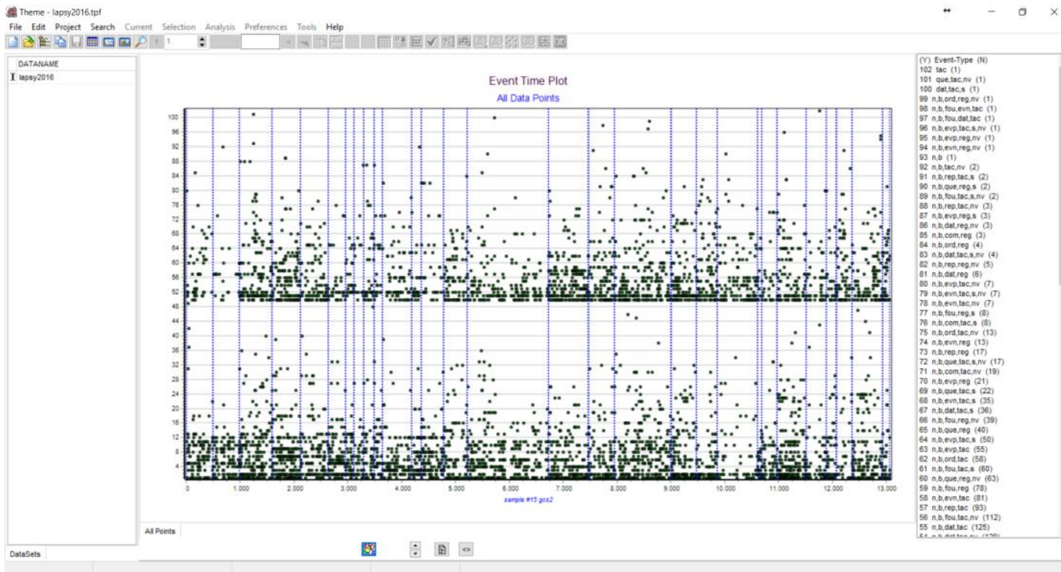
 Changes point size.  Hides/shows the event-type list on the right.  Hides/shows the list of samples on the left and the list of event-types on the right.

The **Sorting** button  at the bottom brings up a window as below for the **selection** of VVT classes (variables) to sort the event-type occurrence series, for example (here), the series will be sorted by actors and the frequency N.

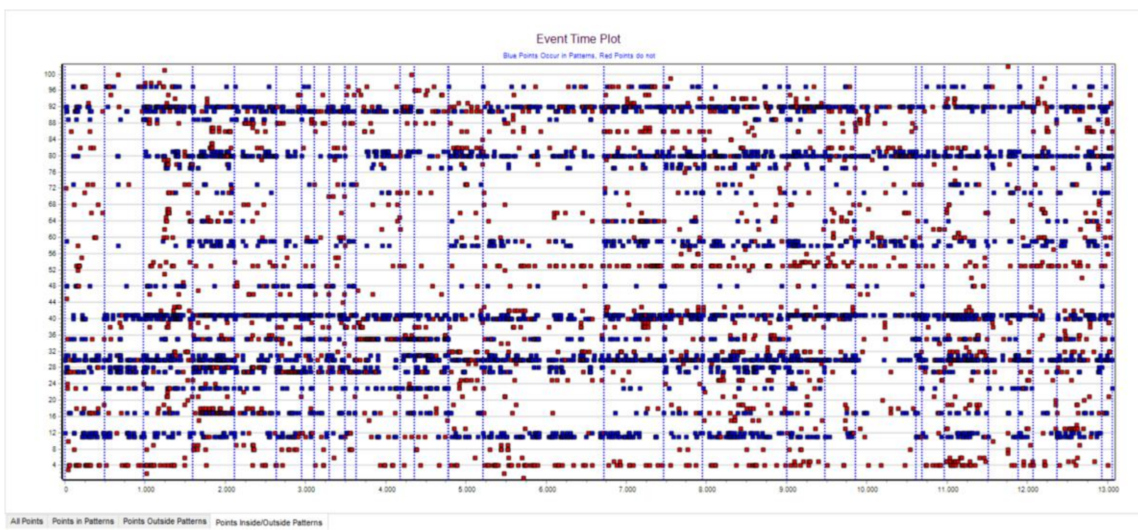


For multi-sample files, the samples (here 29) are shown separated by vertical (blue) dotted lines with the event-types (here 101) listed on the right.


Theme - lapsy2016.tpf
File Edit Project Search Current Selection Analysis Preferences Tools Help
DATANAME
lapsy2016
DataSets
Event Time Plot
All Data Points
sample #15 gcs2
All Points
(Y) Event-Type (N)
102 tac (1)
101 que,tac,nv (1)
100 n,b,tac,nv (2)
99 n,b,rep,tac,s (2)
98 n,b,rep,tac,nv (3)
97 n,b,rep,tac (93)
96 n,b,rep,reg,nv (5)
95 n,b,rep,reg (17)
94 n,b,que,tac,s,nv (17)
93 n,b,que,tac,s (22)
92 n,b,que,tac,nv (408)
91 n,b,que,tac (218)
90 n,b,que,reg,s (2)
89 n,b,que,reg,nv (63)
88 n,b,que,reg (40)
87 n,b,ord,tac,nv (13)
86 n,b,ord,tac (58)
85 n,b,ord,reg,nv (1)
84 n,b,ord,reg (4)
83 n,b,fou,tac,s,nv (2)
82 n,b,fou,tac,s (60)
81 n,b,fou,tac,nv (112)
80 n,b,fou,tac (459)
79 n,b,fou,reg,s (8)
78 n,b,fou,reg,nv (39)
77 n,b,fou,reg (78)
76 n,b,fou,evn,tac (1)
75 n,b,fou,dat,tac (1)
74 n,b,evp,tac,s,nv (1)
73 n,b,evp,tac,s (50)
72 n,b,evp,tac,nv (7)
71 n,b,evp,tac (55)
70 n,b,evp,reg,s (3)
69 n,b,evp,reg,nv (1)
68 n,b,evp,reg (21)
67 n,b,evn,tac,s,nv (7)
66 n,b,evn,tac,s (35)
65 n,b,evn,tac,nv (7)
64 n,b,evn,tac (81)
63 n,b,evn,reg,nv (1)
62 n,b,evn,reg (13)
61 n,b,dat,tac,s,nv (4)
60 n,b,dat,tac,s (36)
59 n,b,dat,tac,nv (129)
58 n,b,dat,tac (125)
57 n,b,dat,reg,nv (3)
56 n,b,dat,reg (6)
55 n,b,com,tac,s (8)



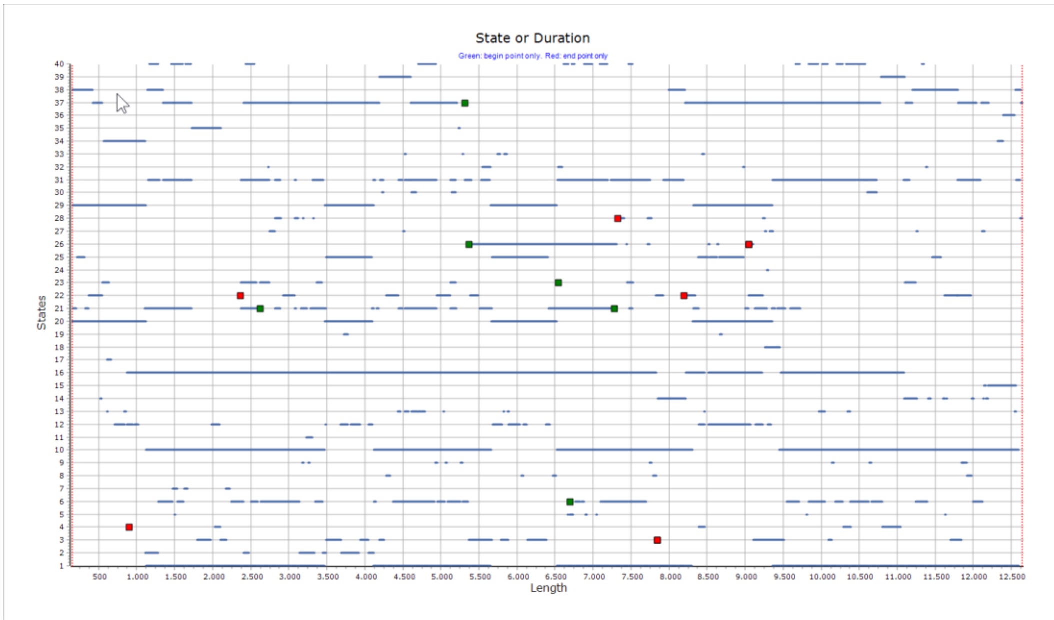
Here the event-type series have been sorted by actors (here, two actors shown in upper vs. lower half according to their order in the VVT) and frequency.



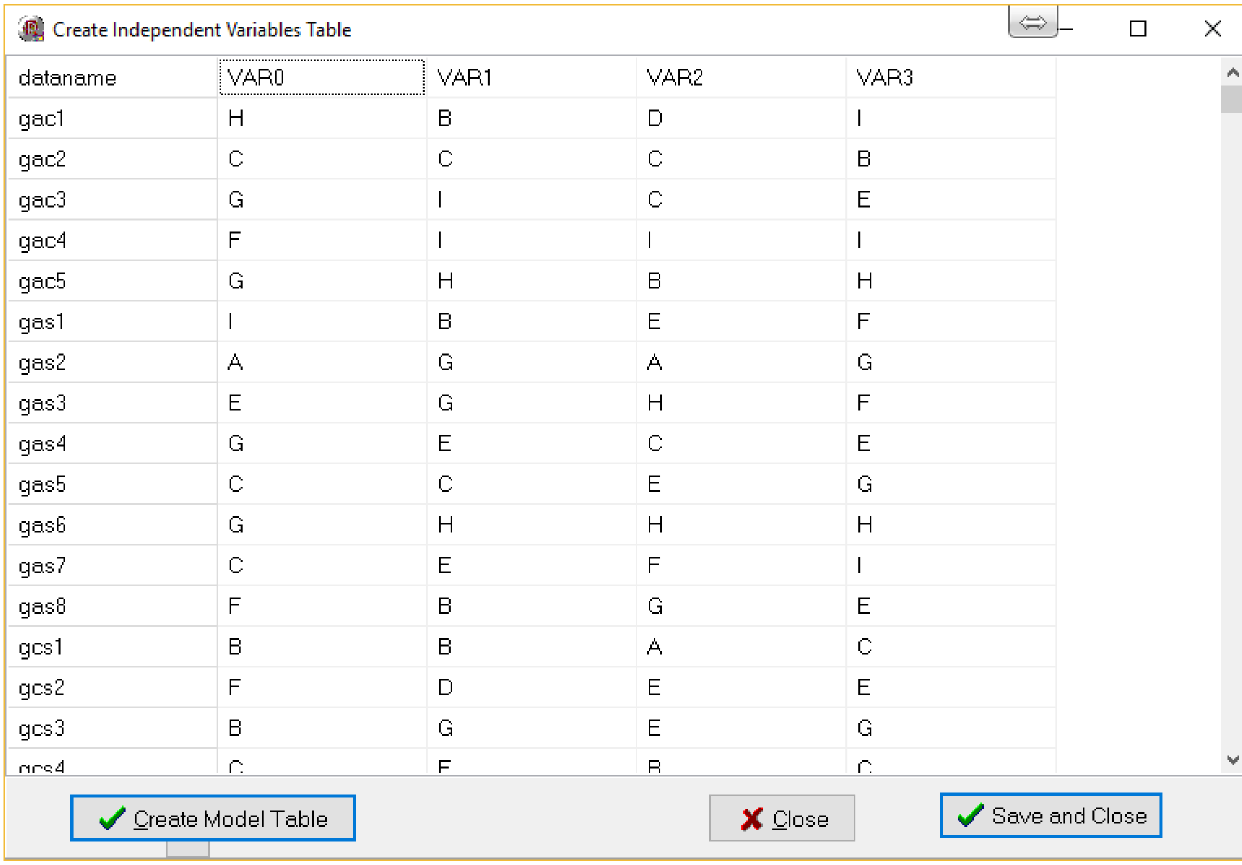
After patterns have been detected, the options at the bottom allow showing: All Points, *Points in Patterns (blue)*, *Points Outside Patterns (red)* or, as here, *Points Inside/Outside Patterns*.

T-State or Duration Chart

The State or Duration Chart is only relevant when the VVT contains the **begin-end** variable with the reserved name "b_e" and then only for behaviors where both beginning and/or end were coded. While providing a certain overview of a T-data set, this chart may also help locating coding errors as it shows with **blue** lines segments all intervals from the **beginning** (b,) to the **end** (e,) of each event-type. This can help identifying cases where only the beginning (**green**) or only the ending (**red**) of a state was coded by error or because they were not visible (being obstructed or occurring outside the observation period).



Create or Edit an Independent Variables Table (optional reference table)

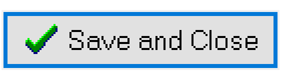


dataname	VAR0	VAR1	VAR2	VAR3
gac1	H	B	D	I
gac2	C	C	C	B
gac3	G	I	C	E
gac4	F	I	I	I
gac5	G	H	B	H
gas1	I	B	E	F
gas2	A	G	A	G
gas3	E	G	H	F
gas4	G	E	C	E
gas5	C	C	E	G
gas6	G	H	H	H
gas7	C	E	F	I
gas8	F	B	G	E
gcs1	B	B	A	C
gcs2	F	D	E	E
gcs3	B	G	E	G
gcs4	C	E	B	C

The optional **Independent Variables Table** allows the grouping of data files for T-pattern search according to an experimental design, by associating sample names (left) with the values of independent variables. An Independent Variable Table is automatically produced with all the samples ordered as in the project and associated with random temporary values for four independent classification variables, but only one can be used at the time when concatenating by independent variables.

Clicking 

allows editing of the table changing the names of the independent variables and their values in accordance with the experimental design.

Then click 

The Independent Variables Table is automatically merged with the *Overview Table* that may be exported as a tab delimited table for entry into, for example, standard statistical programs (SPSS, SAS, etc.) for visualization and analysis grouped by the independent variables to test their statistical effects.

Creating a *Grid or MAP File*

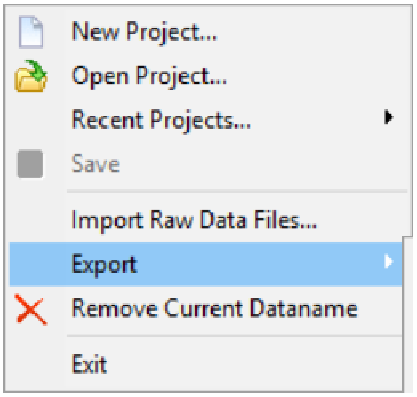
When Event-Type definitions refer to particular locations in physical or conceptual space, the x,y coordinates corresponding to such Event-Types can be indicated in a grid file, which is simple tab delimited text file with no column headings. It can be useful to first create such a file with random x, y coordinate to be consequently modified:

1. From the **Project** menu, choose **Create Random Grid Model File.** A gridfile.nxy will be created with in the first column the event types (zones) in your project and in the second and third column random values for the x,y coordinates.

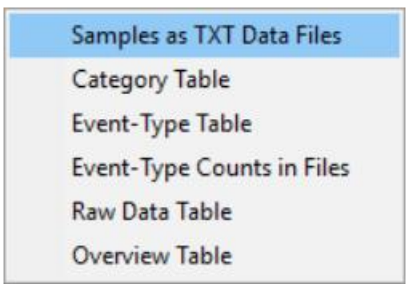
2. Open the file gridfile.nxy in Notepad and replace the random x,y coordinates with the values from your experiment.

3. Save the file as the text file "gridfile.nxy".

4. Close and reopen the project to activate the grid file. ???



Export



Samples as TXT Data Files

All data in project are stored as .txt files (may be useful if the data were entered into the project in non-standard or outdated ways).

Category Table

The vvt.dat (table) version of the vvt.vvt is exported as a tab delimited file.

Event-Type Table

The table eventtypetable.dat in subdirectory sysdir is exported as a tab delimited file and stored in the Project Directory.

Event-Type Counts in Files

The table FileETcounts.dat in subdirectory sysdir is exported as a tab delimited file to the Project Directory.

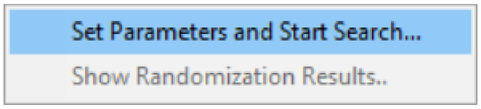
Raw Data Table

The table RawDataTable.dat in the sysdir subdirectory is exported as a tab delimited file and stored in the Project directory.

Overview Table

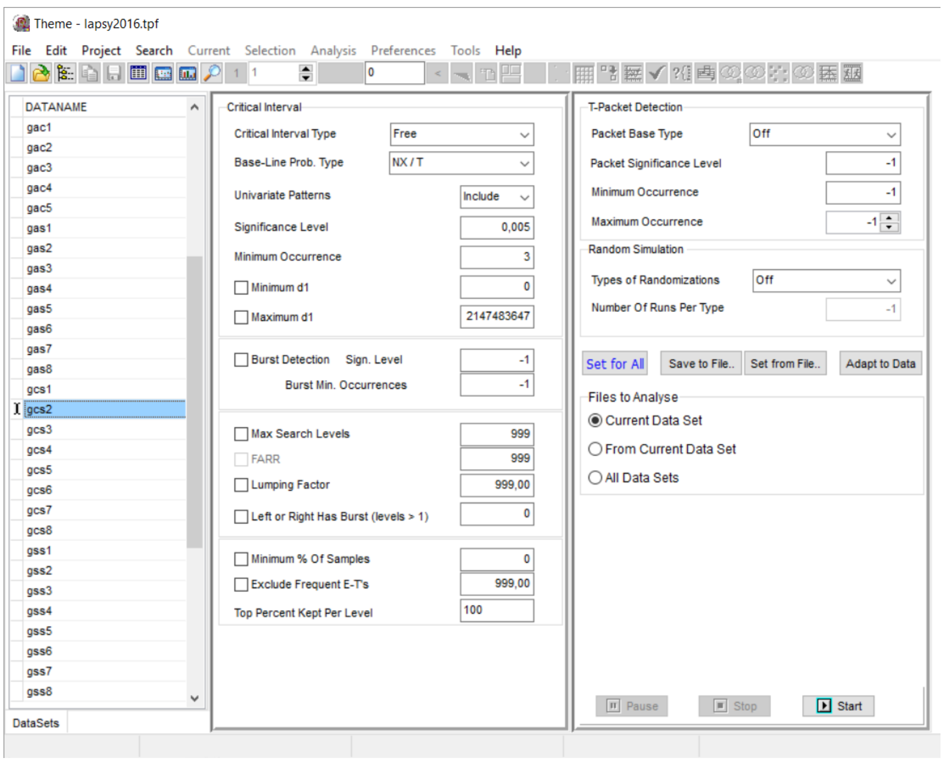
The table base.dat in the Project Directory is exported as the tab delimited file base.txt in the Project Directory.

Search



At this stage the data should be ready for T-pattern detection. Click the  or from the **Search** menu, choose **Set Search Parameters and Start Search...** a dialog box appears with the current settings of each parameter some of which are not checked (used). When a parameter is checked, it automatically gets a default value that may need to be changed depending on the data and goals of the search. The search parameters have crucial influence on which patterns can be detected. For example, if minimal pattern occurrences are specified higher than exist in the data, no patterns are, of course, detected, just as if a critical interval significance level is specified that does not exist in the data.

Set Parameters and Start Search

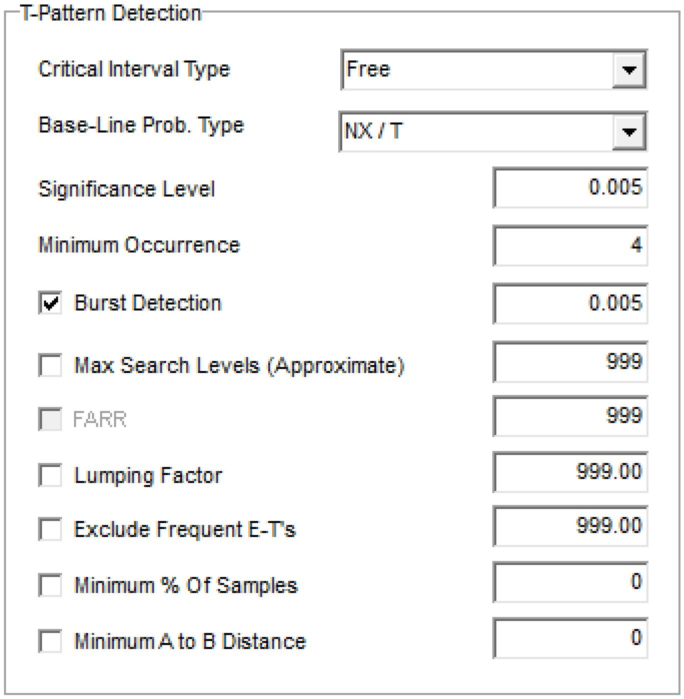


Using this dialog, parameters can be set separately for each selected dataname (here gcs2) and, optionally, in a project with more than one dataname (as here), the parameters for all datanames can be set the same as for the currently selected by clicking:

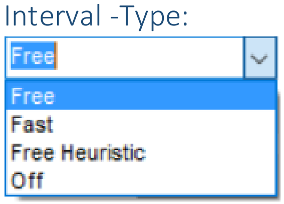
.

When the search parameters have been set the search can begin.

{To save the Search Parameters table: Click the **Save** icon in the toolbar or from the **File** menu, choose **Save** to save the search parameter settings in the file `SearchParam.txt`. The file is saved in the project folder on C:\Program Files\PatternVision\Theme 6.0. }



Critical Interval -Type:

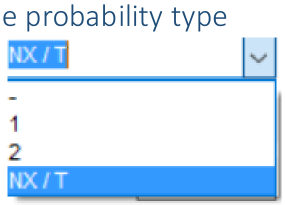


The default option, 'Free' uses an (exhaustive) algorithm guarantying that, if present, at least one critical interval (CI) is found (i.e., the closest free CI).

The 'Fast' option only searches for fast CI's (same as in Theme 5).

The 'Free Heuristic' algorithm is not exhaustive (similar to the default 'Free' option of Theme 5) and may thus miss some CI's.

Baseline probability type

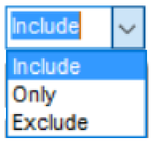


Decides how baseline probability is calculated (for all event-types and detected T-patterns).

NX/T is the default and means that the baseline probability for X is the number of occurrences of X (NX) divided by the duration of the observation interval (T).

The other options are not recommended for the time being.

Univariate Patterns



Decides whether patterns of the same X (event-type or pattern) are also searched for.

Minimum Occurrences

The minimum number, ≥ 2, of times a pattern must occur to be detected. Default is 3.

Significance Level

The maximum a priori probability of any critical interval relationship to occur by chance alone. Default = 0.005. Value less than 0.05 usually recommended.

Burst Detection

Allows the detection of T-bursts significant at the set level. Default = 0.005.

Max Search Levels

The maximum number of hierarchical *search levels* (≥1).
Default = not checked = 999 (meaning no real limit).
Lower values mean that only simpler patterns are detected, which often prevents overload in highly structured data.

Lumping Factor

For a critical interval **(A** [d1, d2] **B**) forming pattern (**AB**) occurring **NAB** times, if the **forward** conditional probability **NAB**/**NA** exceeds the **Lumping Factor** then **B** will be eliminated from the rest of the search process. Vice-versa, if the **backward** conditional probability (**NAB**/**NB**) exceeds the **Lumping Factor** then **A** will be eliminated.
In either case, the pattern (**AB**) is added to the list of detected patterns.
Default = 999 (meaning not used)

A lower value (≤ 1 or 100%) means that more lumping can take place so fewer patterns may be detected, which may help avoiding overload in highly structured data. Values between 0.7 and 1.0 often work well.

Exclude Frequent E-T's (Event Types)

Number of standard deviations above the mean number of occurrences of event-types. Allows excluding from the search process all event-types exceeding the set level. For example, a value of 2.5 excludes all event-types occurring more often than the mean + 2.5 standard deviations. The smaller the value (≥0, the more event types are excluded, and the fewer patterns may be found. Values between 1 and 2.5 often work well.

Note: Even though highly frequent event types may never occur in a pattern, some such event types do occur in all kinds of patterns, but add little useful information. They may thus clutter up all diagrams and even cause overloading of the search engine. Such event types can be more precisely eliminated (checked off) one-by-one (see below), but a quicker way to eliminate them (but possibly also some others) is through the use of this search parameter. A useful level may be found by inspecting the event-type frequency chart

Default = not checked (=999)

Minimum Samples

The minimum percentage x, $1 \leq x \leq 100$, of samples (datanames) in which a pattern must occur to be kept. Default = 1. For example, for a value of 50, a pattern must occur in half of all samples to be kept. This parameter is sometimes a powerful tool when too many patterns are detected possibly overloading the search engine.

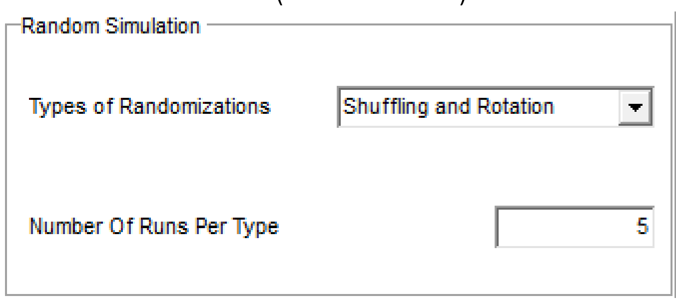
Running Simulation (Monte Carlo) Tests

For usual data sizes, the T-pattern detection algorithm calculates a large number of significance tests. So purely by chance some of them will be significant even if the data is random. It is therefore important to run the algorithm on randomized versions of the data to see if the set of detected patterns deviates significantly from such random expectation.

Theme provides three different randomization methods: 1) T-shuffling for a) point series and b) locations 2) T-rotation and T-position shuffling.

Setting Parameters

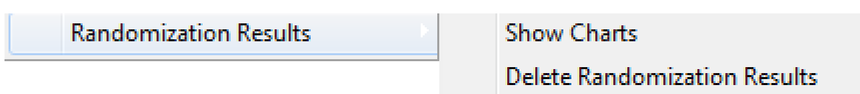
Random Simulation (Monte Carlo)



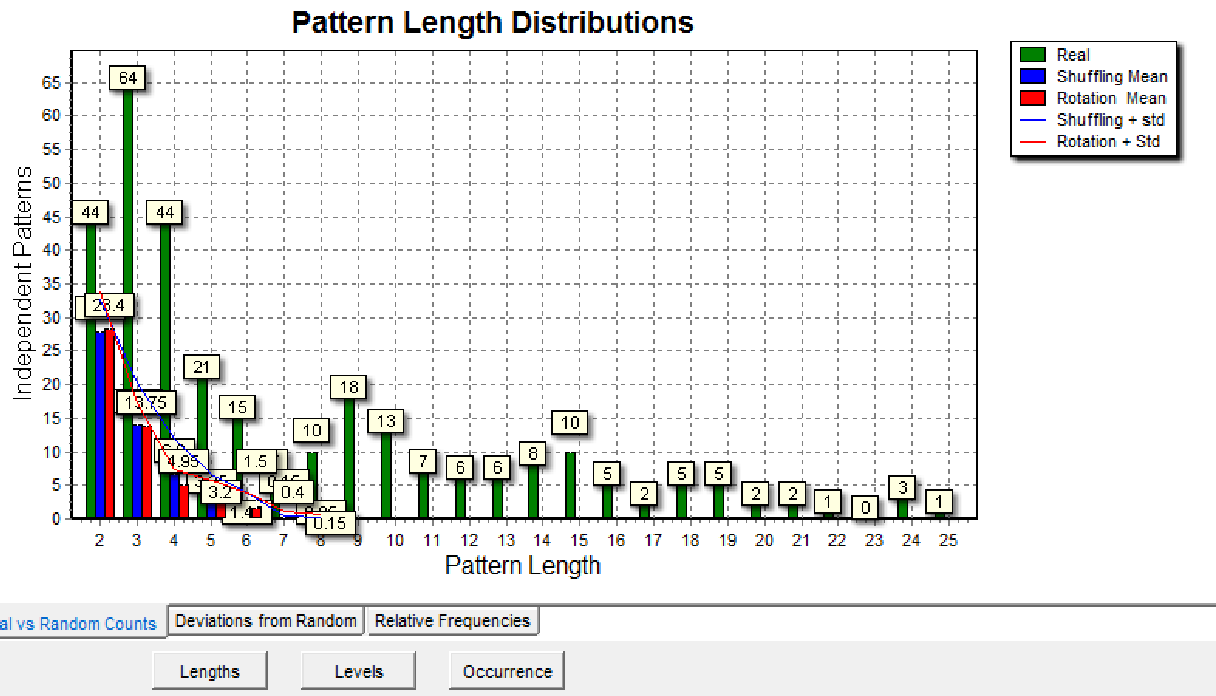
Types of Randomizations: The three options here are ‘Shuffling’, ‘Rotation’, and ‘Shuffling and Rotation’ (the latter not present in Theme 5).

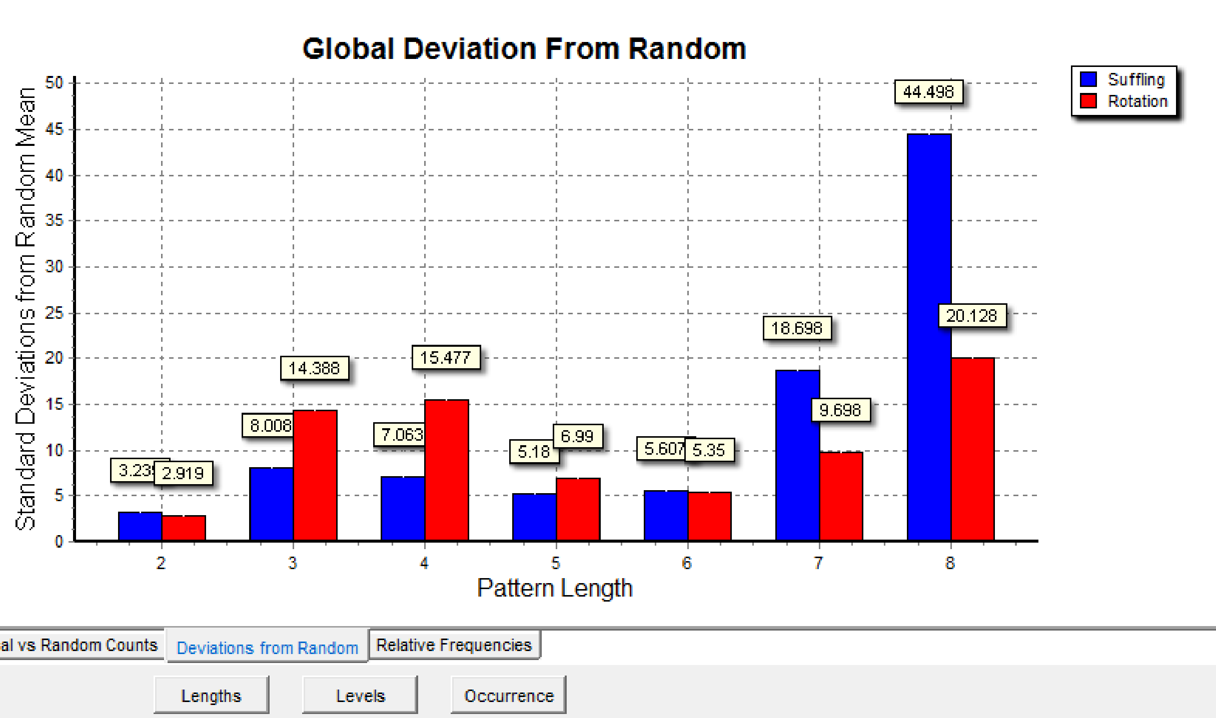
Number of Runs per Type: Specifies the number of randomizations-searches of each kind specified. Here a total of 10, that is, 5 for Shuffling and 5 for Rotation.

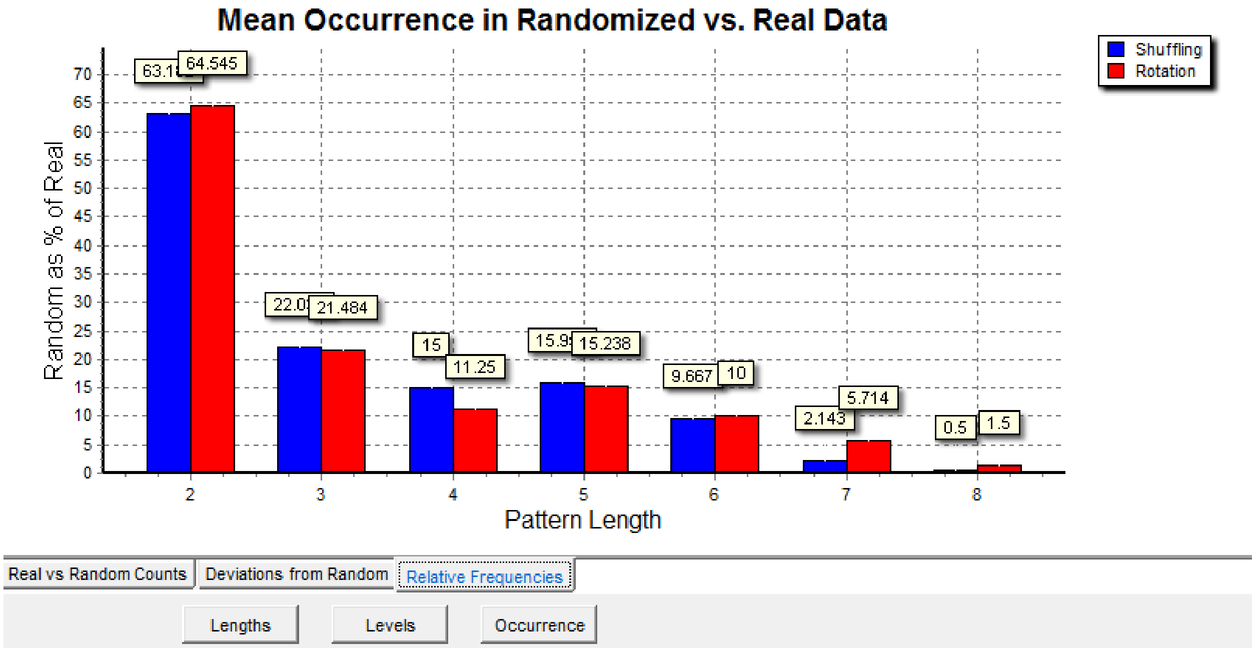
Randomization Results



Show Charts

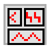

Pattern Length Distributions
Real
Shuffling Mean
Rotation Mean
Shuffling + std
Rotation + Std
Independent Patterns
Pattern Length
64
44
44
21
15
10
18
13
7
6
6
8
10
5
2
5
5
2
2
1
0
3
1
28.4
13.75
4.95
1.5
3.2
0.4
0.15
Real vs Random Counts
Deviations from Random
Relative Frequencies
Lengths
Levels
Occurrence

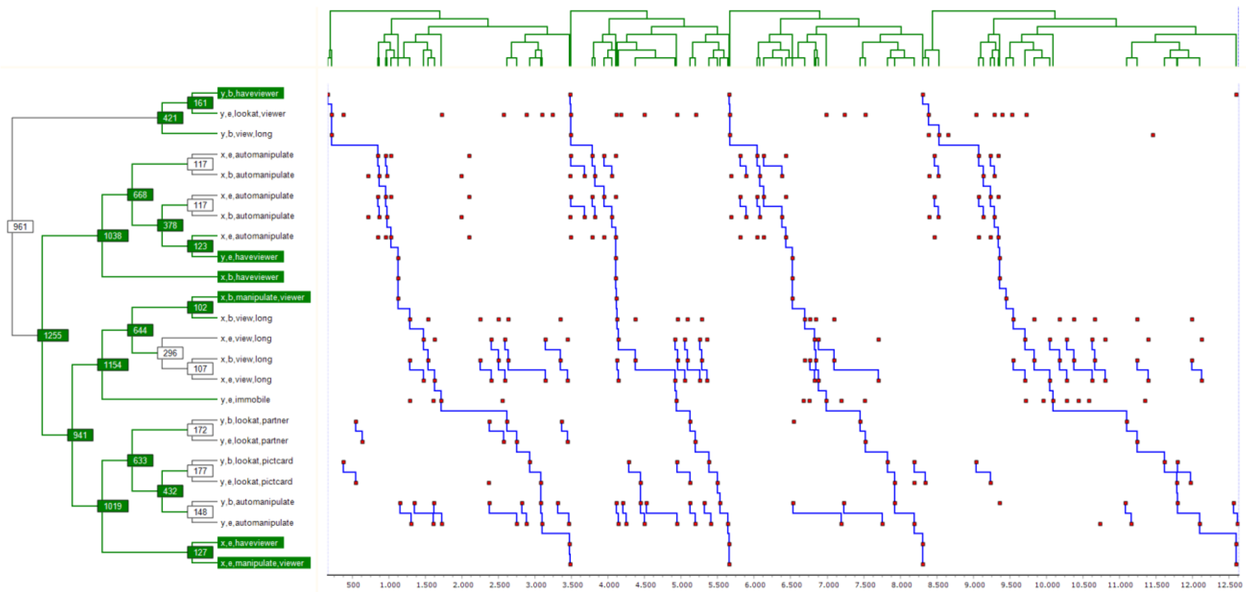

Global Deviation From Random
Suffling
Rotation
Standard Deviations from Random Mean
Pattern Length
44.498
20.128
18.698
9.698
14.388
8.008
15.477
7.063
6.99
5.18
5.607
5.35
2.919
Real vs Random Counts
Deviations from Random
Relative Frequencies
Lengths
Levels
Occurrence



Mean Occurrence in Randomized vs. Real Data
Shuffling
Rotation
Random as % of Real
Pattern Length
63.1
64.545
22.0
21.484
15
11.25
15.9
15.238
9.667
10
2.143
5.714
0.5
1.5
Real vs Random Counts
Deviations from Random
Relative Frequencies
Lengths
Levels
Occurrence

Selecting and Analyzing Detected Patterns

To visualize the currently selected pattern(s) select **Current, Graphic Displays, Diagrams** or simply click the  button at the top, which shows by default a diagram of three closely related parts to be described below. Here showing a pattern detected in the above data.

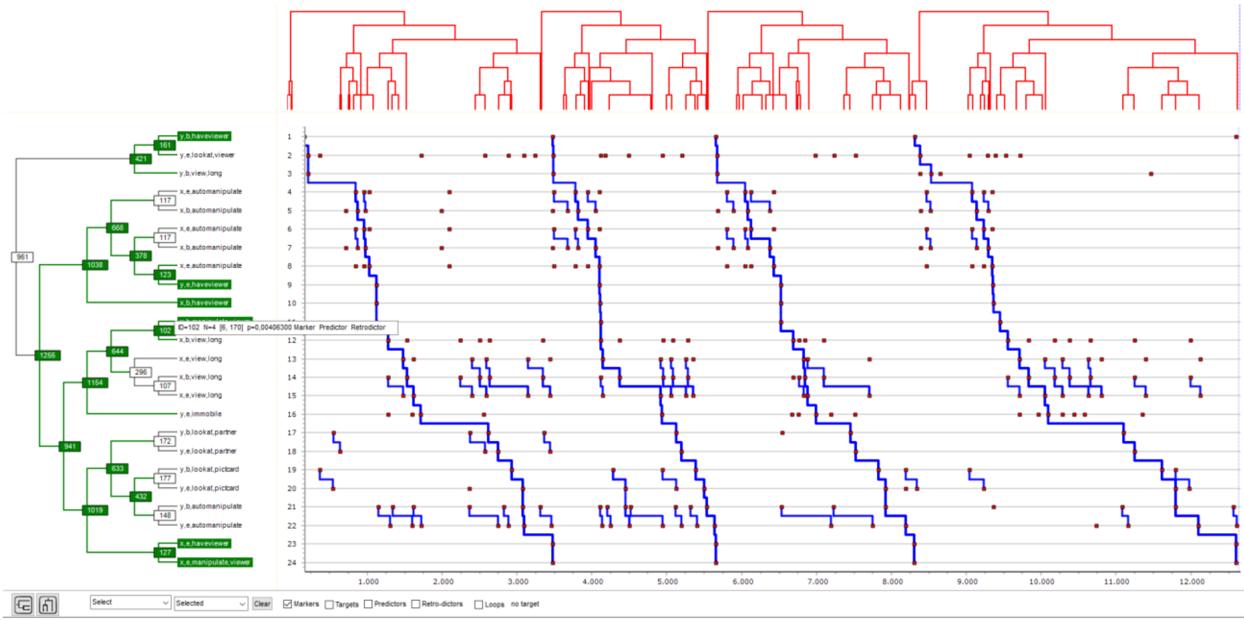


The three parts show

a) The Detection Tree at the left shows the bottom-up, level-by-level detection of the currently selected pattern

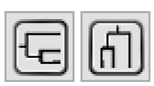
b) The dynamic Occurrence Trees show (here the four) occurrences of the pattern. Each tree is a kind of copy of the Detection Tree that has been rotated and its terminals (without the labels) are located in accordance with their occurrence times (locations) and aligned with the Connection Chart below.

c) The Connection Diagram of dots and lines, shows the raw data series of each of the event-types in the pattern and how some of their points are connected forming the recurrent pattern. Note that the point series of an event-type repeated within a pattern is shown each time.

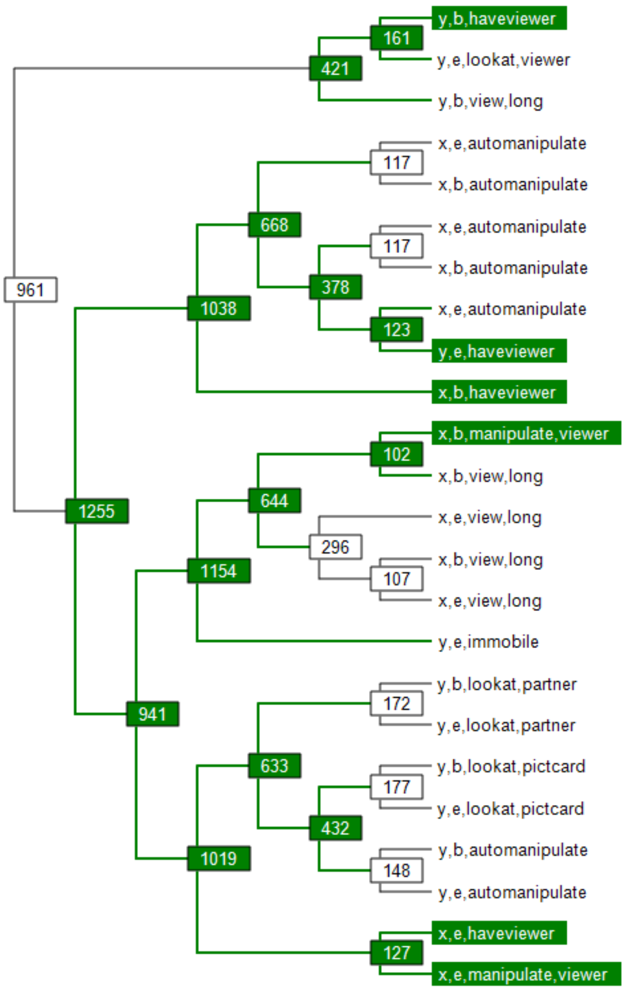


The appearance of each part can be changed by right-clicking each of them and selecting the various options, for example, concerning the colors, widths and sizes of text, lines and points.

By left-clicking and holding the boundaries of the parts and dragging the mouse the relative sizes of the three parts can also be changed.

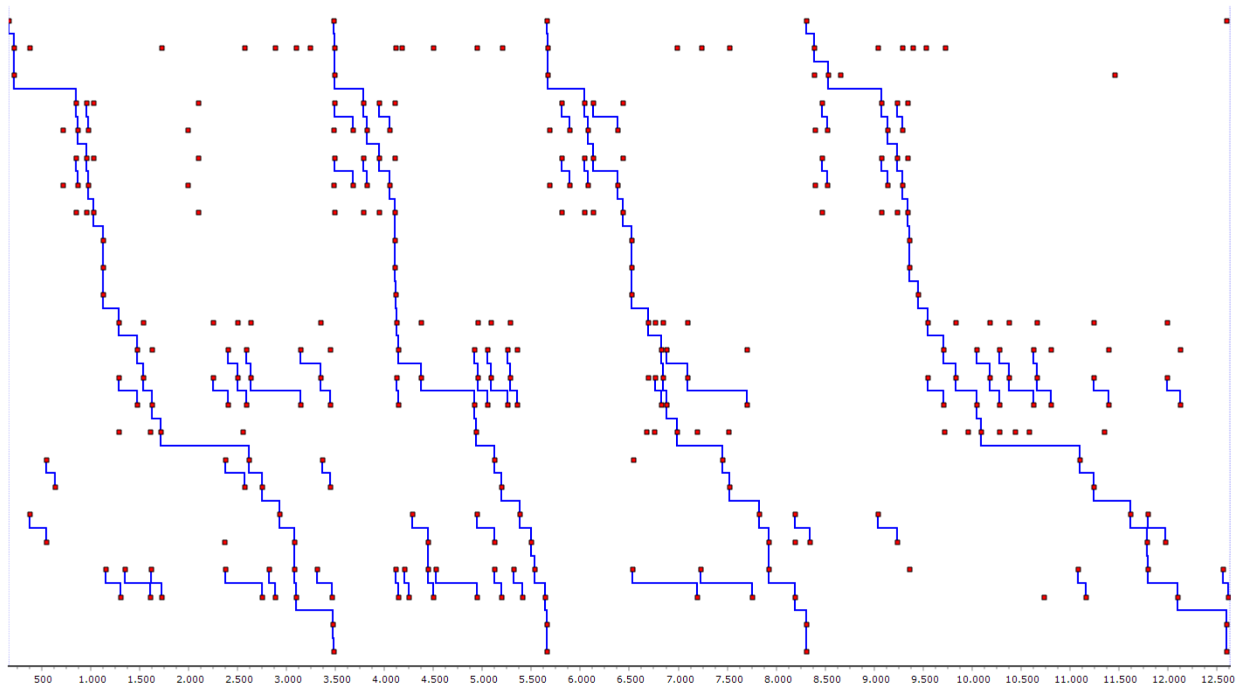
Using these buttons  (lower left) the left and the top parts can be included or excluded.

Placing the mouse over a node in the left detection tree brings up some of the parameters of that node as here can be seen for node (ID=) 102. Which shows that it occurs 4 times and the critical interval at the node is [6, 170] with p = 0.004063 and that it is a T-marker, a T-predictor and a “T-retrodictor” of this pattern (see below).

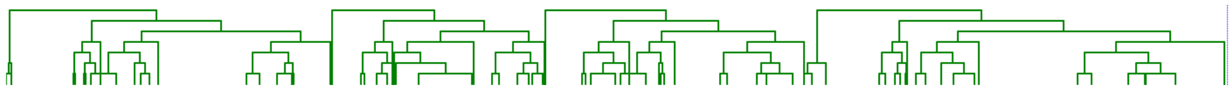


This part of the three-part diagram shows how the recurrent pattern was detected bottom-up and level-by-level (i.e., from right to left).

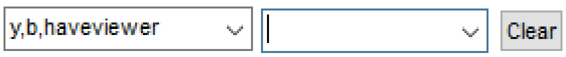
When Markers (see below) is checked these will be shown in **green.** As can be seen, markers can be either event-types or T-patterns.



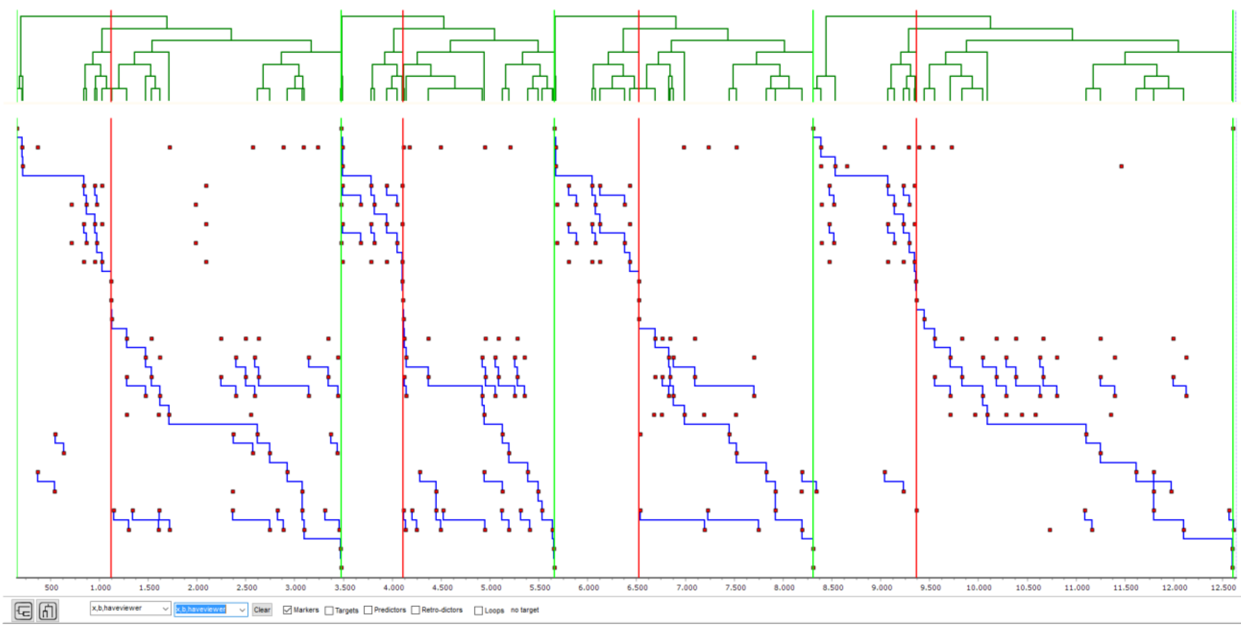
This Connection Chart shows exclusively the occurrence point series of the event-types in the pattern. If an event-type or pattern are repeated at different positions within the same pattern their series are also repeated.



This top part in the pattern diagram shows the Detection Tree as it recurs, here four times. It is as if the Detection Tree had been rotated and put on top so that its terminals correspond to the corresponding points in the Connection Chart. Thus in the following figure the event-type y,b,haveviewer (light green) is the first terminal in the pattern and x,b,haveviewer (number 10 in the pattern) have been selected and assigned colors by selecting them from the lists at the lower left of the screen.

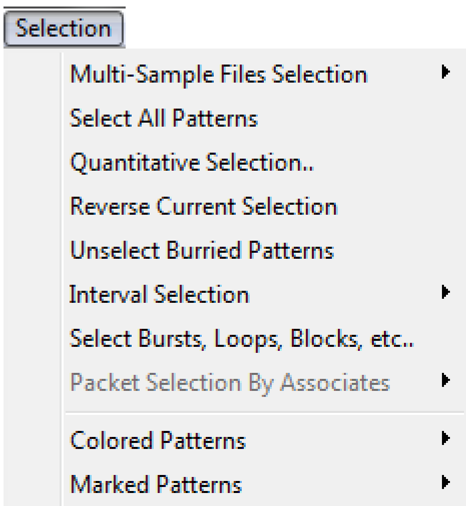


The left list contains all the event-type names and as they are selected and colors assigned to each, they will appear in the right list that can be cleared as well as all thir lines using the **Clear** button.

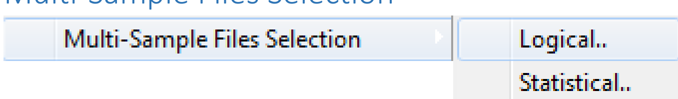


Pattern Selection

Selection and analysis options are somewhat loosely classified as quantitative, qualitative and structural. Some are only available and meaningful for multi-sample files where the focus is sometimes on patterns detected in different samples or sets of samples corresponding to an experimental or observational design. The “Logical and Statistical Sample Selection” options are thus only enabled for multi-sample files.

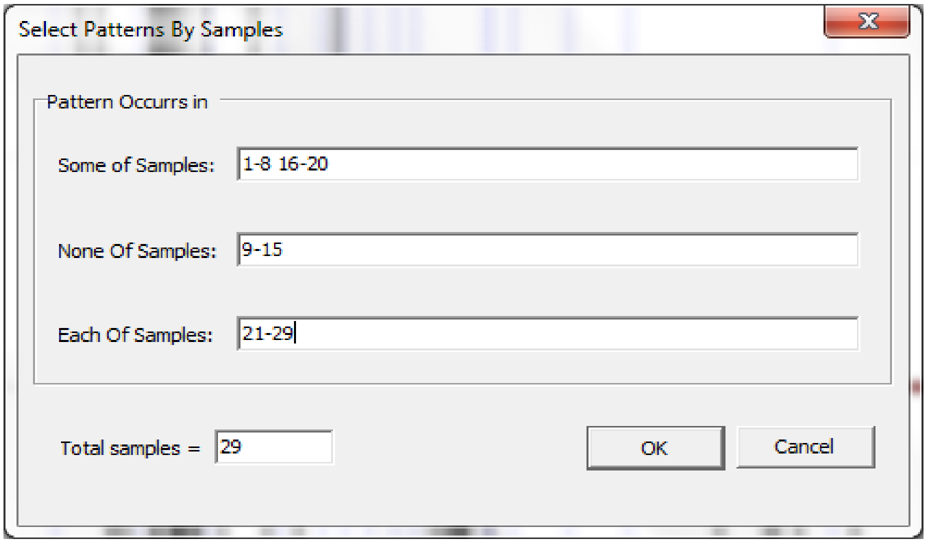


Multi-Sample Files Selection



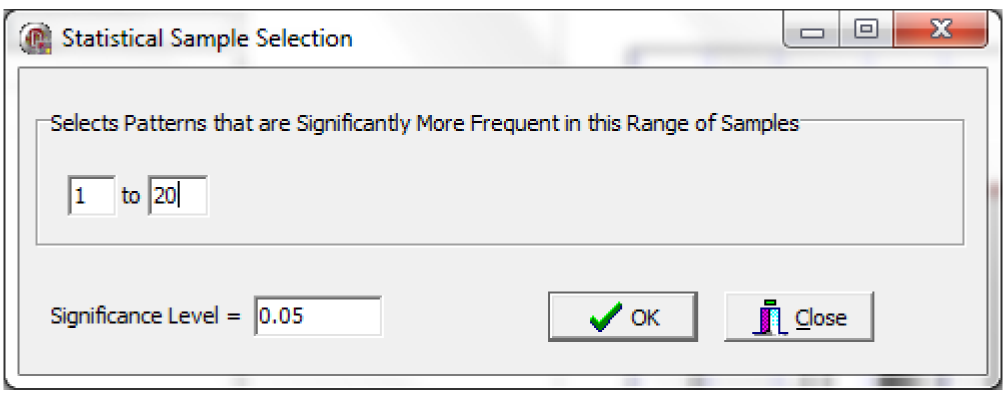
Multi-sample files allow selecting patterns that occur significantly more in some samples than in others. Samples are automatically ordered alphabetically by their names, so adding a letter at the beginning of a sample name can be used to indicate experimental group. To comparing experimental groups in this way the user must know the ranges of samples corresponding to the different experimental sub-sets. For example, if there are four groups the sample names may begin with a, b, c or d, respectively.

Logical

4. 

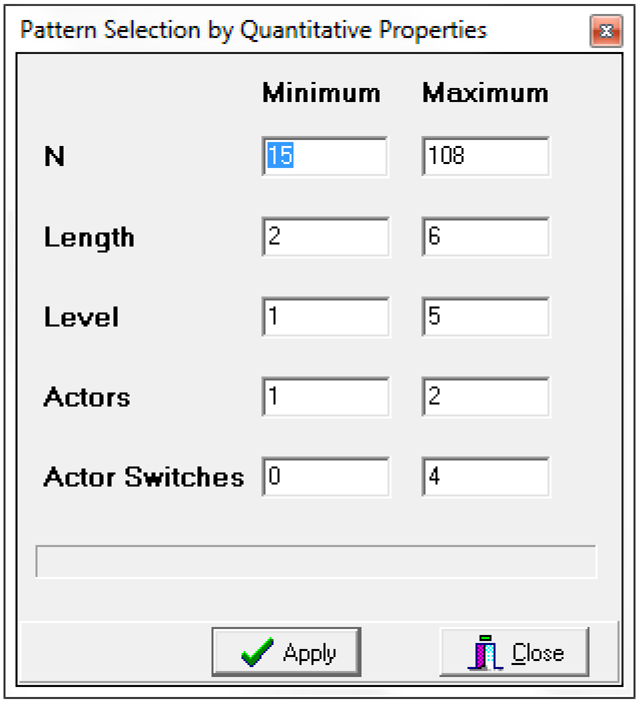
This query will select only those patterns that occur in <u>some of samples</u> 1-8 and 16-20, in <u>none of samples</u> 9-15, and in <u>each of the samples</u> 21-29. (Any one or more of the three fields can be used.)

Statistical



Here patterns are selected that appear significantly more often in samples 1 to 20 (combined) than in the multi-sample file as a whole. The default level of significance is 0.05, but this value can be changed by the user.

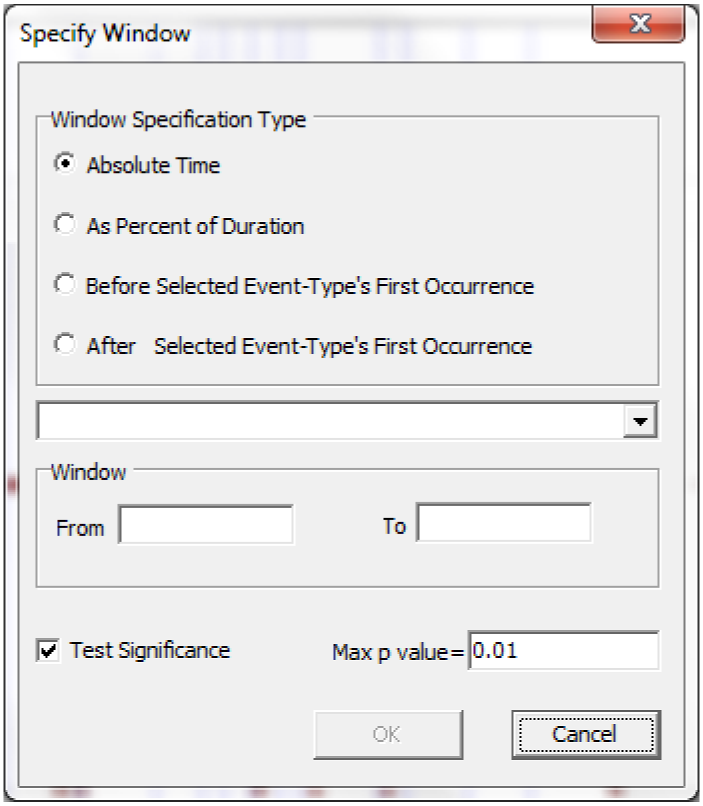
Quantitative Selection



This table shows the value ranges of five parameters of the currently selected patterns in the project. Shrinking these ranges by editing their values and then clicking the Apply button, the corresponding subset of patterns is selected.

Interval Selection

Select By Time Window:



To decide whether some patterns tend to occur significantly more during particular time-windows, here such windows can be specified in various ways as well as the significance significance level. The limits of the window is specified in the fields “From” and “To” and the numbers are interpreted in accordance with the above selection.

- Absolute Time {Numbers are considered absolute}
- As Percent of Duration {For example, [50, 100] means from the middle to the end}
- Before Selected Event-Type’s First Occurrence {an event-type is then slected froma list}
- After Selected Event-Types’s First Occurrence {an event-type is then slected froma list}

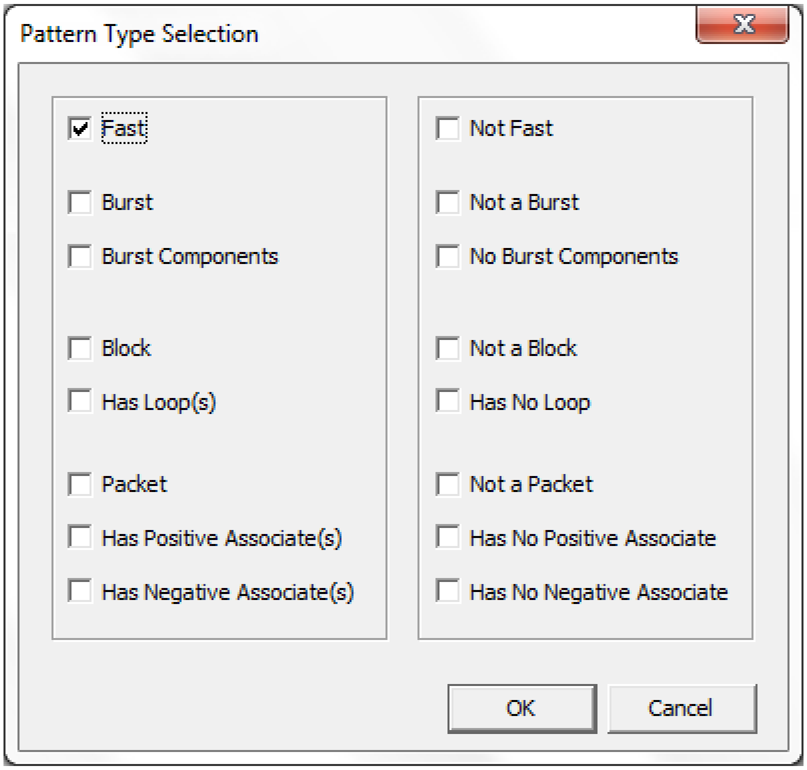
{Note that by adding event-types possibly occuring only once each (and thus never occurring in patterns) the last two options can be used to find patterns that increase or decrease infrequency before or efter such event-types . Other event-types can also be used, of course.}

Revese Current Selection

Currently selected patterns are unselected, while those unselected are selected.

Structural Selection

Select Packets, Bursts, Loops, Blocks and more.



Here patterns can be selected according to the presence/absence of particular aspects.

Burst

Pattern occurs as a T-burst

Burst Component

Pattern has at least one sub-pattern that occurs as a T-burst.

Block

Pattern is a T-block

Has Loop(s)

At least one event-type recurs within the terminal string (list) of the pattern.

Packet

The pattern occurs as the base of a T-Packet (i.e., has at least one T-associate).

Has Positive Associate(s)

The pattern (packet) has at least one positive T-associate.

Has Negative Associate(s)

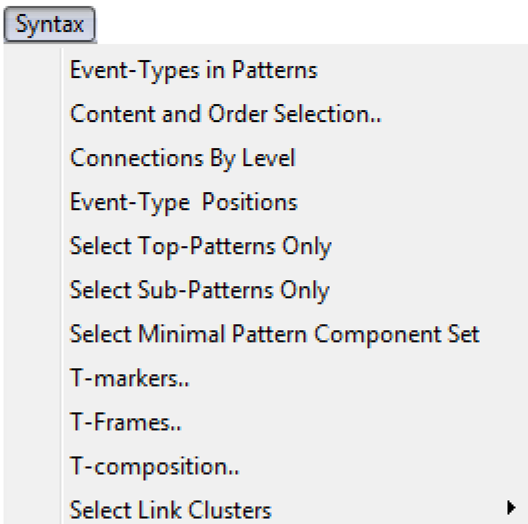
The pattern (packet) has at least one negative T-associate.

Redundancy Reduction

Unselect Buried Patterns

Patterns that have event-type (terminal) occurrences that all occur within some other pattern are unselected.

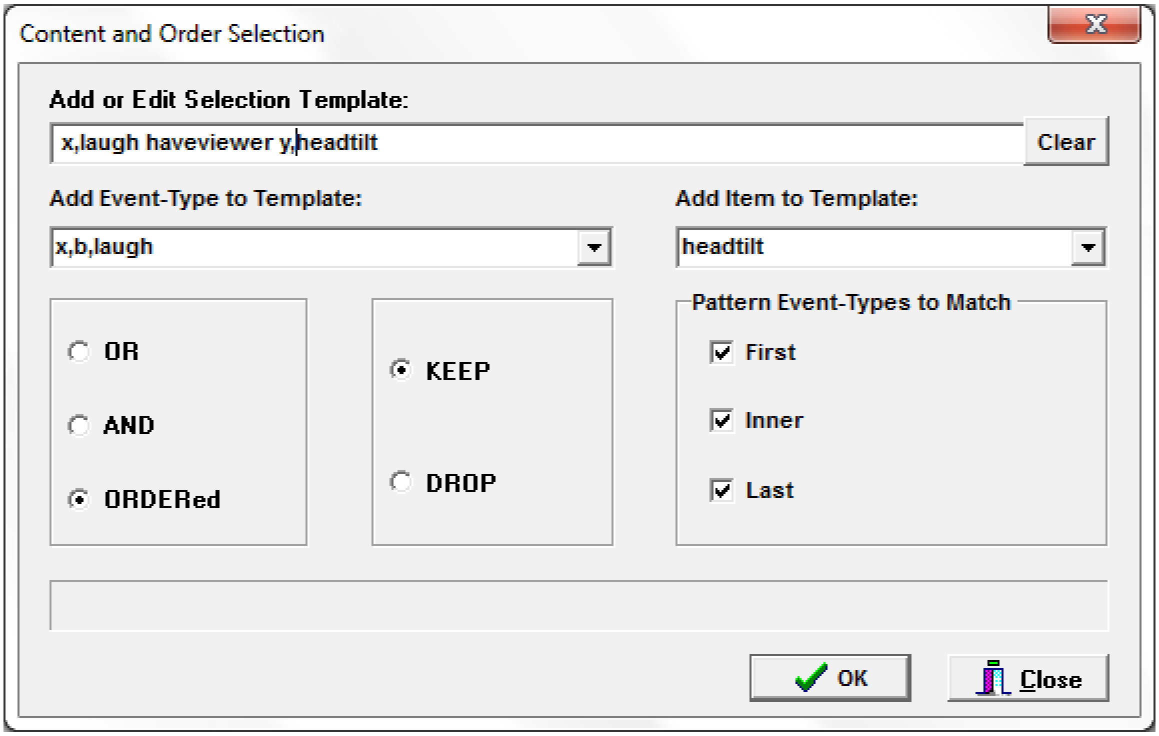
Syntax



Many and even most tools in Theme 6 could be said to be aimed at the study of syntax, but some of them have been tentatively assembled under the label Syntax in the main menu.

Qualitative Selection

Content and Order Selection



Filtering is based on (fully and) partially specified event-type names (templates) that may therefore match more than one fully specified event-types. For example, sue,fast will match sue,b,run,fast,left and sue,e,fast and sue,fast,left etc.. Similarly, X,laugh matches, for example, both x,b,laugh and x,e,laugh and haveviewer matches any event-type that contains the item haveviewer.

One or more such fully or partially specified templates can be added to the filter (see top line) either by entering (typing) them directly or by selecting them from the event-type list or the items lists, where they can also be edited.

The options OR, AND, ORDERed mean

OR: at least one of the templates must match

AND: all templates must match

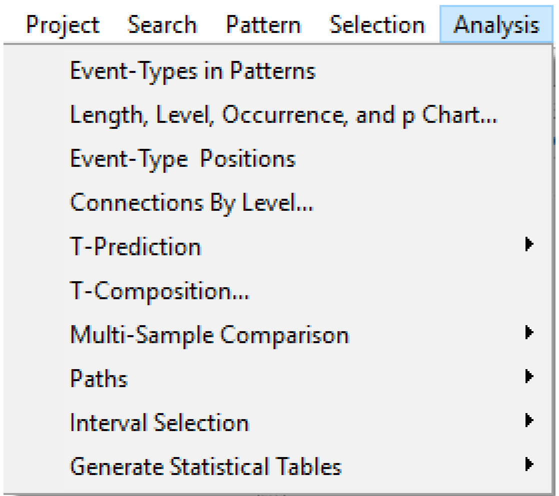
ORDERed: all must match and in the order of the filter

The KEEP vs. DROP options specify whether the matching patterns should be selected (kept) or unselected (dropped).

The **First, Inner, Last** options. Sometimes, only the first (or the last) behavior in patterns is of interest, that is, how they begin (or end). When only First is checked, the filtering only tries to match with the first event-type in each pattern. When only Inner is checked, neither the first nor the last event types of a pattern are considered. Any combination of the three can be checked or unchecked.

In Theme 6 it is now possible to match any one or more of the first, last or intermediate terminal event-types when filtering (selecting) patterns. (The default with all checked corresponds to Theme 5.)

Analysis



Event-Types in Patterns

EventType	N	AsNODE	AsFirst	AsInner	AsLast	R_NODE	R_First	R_Inner	R_Last	First_p	Inner_p	Last_p
x,b,automanipulate	16	16	16	16	13	1,00	1,00	1,00	0,81	0,0377	0,0056	99,0000
x,b,glanceat,partner	11	11	11	9	8	1,00	1,00	0,82	0,73	0,1422	0,3558	99,0000
x,b,glanceat,pictcard	5	5	5	5	5	1,00	1,00	1,00	1,00	0,5697	0,3797	99,0000
x,b,glanceat,viewer	4	4	4	0	0	1,00	1,00	0,00	0,00	0,6867	98,0000	98,0000
x,b,haveviewer	4	4	4	4	4	1,00	1,00	1,00	1,00	0,6867	0,5171	99,0000
x,b,laugh	14	14	14	7	7	1,00	1,00	0,50	0,50	0,0648	99,0000	99,0000
x,b,lookat,partner	4	4	0	4	4	1,00	0,00	1,00	1,00	98,0000	0,5171	99,0000
x,b,lookat,pictcard	11	11	11	9	9	1,00	1,00	0,82	0,82	0,1422	0,3558	99,0000
x,b,lookat,viewer	6	6	6	0	0	1,00	1,00	0,00	0,00	0,4643	98,0000	98,0000
x,b,manipulate,viewer	4	4	4	4	4	1,00	1,00	1,00	1,00	0,6867	0,5171	99,0000
x,b,order,viewer	7	6	4	5	6	0,86	0,57	0,71	0,86	99,0000	99,0000	99,0000
x,b,view,long	22	22	22	22	22	1,00	1,00	1,00	1,00	0,0069	0,0004	0,0594
x,b,view,short	9	9	9	4	6	1,00	1,00	0,44	0,67	0,2339	99,0000	99,0000
x,b,walk	10	10	10	6	7	1,00	1,00	0,60	0,70	0,1830	99,0000	99,0000
x,e,automanipulate	16	16	16	16	16	1,00	1,00	1,00	1,00	0,0377	0,0056	0,1626
x,e,glanceat,partner	11	11	7	10	11	1,00	0,64	0,91	1,00	99,0000	0,1542	0,3518
x,e,glanceat,pictcard	5	5	0	0	5	1,00	0,00	0,00	1,00	98,0000	98,0000	99,0000
x,e,glanceat,viewer	4	4	0	0	4	1,00	0,00	0,00	1,00	98,0000	98,0000	99,0000
x,e,haveviewer	4	4	0	4	4	1,00	0,00	1,00	1,00	98,0000	0,5171	99,0000
x,e,laugh	14	14	5	5	14	1,00	0,36	0,36	1,00	99,0000	99,0000	0,2235
x,e,lookat,partner	5	4	0	0	4	0,80	0,00	0,00	0,80	98,0000	98,0000	99,0000
x,e,lookat,pictcard	12	12	9	11	12	1,00	0,75	0,92	1,00	99,0000	0,1122	0,3035
x,e,lookat,viewer	6	6	4	4	6	1,00	0,67	0,67	1,00	99,0000	99,0000	0,6793
x,e,manipulate,viewer	4	4	4	4	4	1,00	1,00	1,00	1,00	0,6867	0,5171	99,0000
x,e,view,long	21	21	16	21	21	1,00	0,76	1,00	1,00	99,0000	0,0007	0,0706
x,e,view,short	9	9	6	7	9	1,00	0,67	0,78	1,00	99,0000	0,5433	0,4662
x,e,walk	9	9	6	7	9	1,00	0,67	0,78	1,00	99,0000	0,5433	0,4662
y,b,automanipulate	21	21	21	21	19	1,00	1,00	1,00	0,90	0,0092	0,0007	0,4137
y,b,glanceat,partner	10	10	10	0	6	1,00	1,00	0,00	0,60	0,1830	98,0000	99,0000
y,b,glanceat,pictcard	6	6	6	0	0	1,00	1,00	0,00	0,00	0,4643	98,0000	98,0000
y,b,glanceat,viewer	8	5	0	5	0	0,63	0,00	0,63	0,00	98,0000	99,0000	98,0000
y,b,haveviewer	5	5	5	0	4	1,00	1,00	0,00	0,80	0,5697	98,0000	99,0000
y,b,headtilt	5	5	5	5	5	1,00	1,00	1,00	1,00	0,5697	0,3797	99,0000
y,b,immobile	17	17	17	15	12	1,00	1,00	0,88	0,71	0,0286	0,0691	99,0000
y,b,kneel	5	5	5	0	0	1,00	1,00	0,00	0,00	0,5697	98,0000	98,0000
y,b,laugh	5	5	5	4	4	1,00	1,00	0,80	0,80	0,5697	99,0000	99,0000
y,b,lookat,partner	8	8	8	8	5	1,00	1,00	1,00	0,63	0,2968	0,1338	99,0000
y,b,lookat,pictcard	10	10	10	7	4	1,00	1,00	0,70	0,40	0,1830	99,0000	99,0000
y,b,lookat,viewer	25	25	24	24	23	1,00	0,96	0,96	0,92	0,0152	0,0009	0,2724
y,b,manipulate,nose	4	4	4	0	0	1,00	1,00	0,00	0,00	0,6867	98,0000	98,0000
y,b,manipulate,viewer	5	5	4	0	5	1,00	0,80	0,00	1,00	99,0000	98,0000	99,0000
y,b,sit	10	10	4	8	9	1,00	0,40	0,80	0,90	99,0000	0,4441	99,0000

This table shows to what extent event-types take each of three positions within patterns; as the **first**, the **last** or in some position **in between**, called **inner** positions.

The leftmost column lists the event-types found in patterns and the number of times they occur in the data (N), as a terminal (AsNode), as the first terminal (AsFirst), as an inner terminal (AsInner) and as a last terminal in some pattern(s). Note that an event-type occurrence can simultaneously be a part of a number of different patterns and at different positions in each. Example, standing up can at the same time be the end of a lunch pattern and a beginning of “taking the telephone in another room” pattern and thus possibly somewhere in the middle of a working day pattern.

The following columns:

R_Node, R_First, RInner and R_Last are Asnode, AsFirst, AsInner and AsLast divided by N.

First_p, Inner_p and Last_p are rough estimates of the tendency of the event-type to occur in each of the three position relative to other event-types. ………………………. missing…….see T5 manual….

Length, Level, Occurrence and p Chart

Event-Type Positions

Connections by Level

There are two different tables in this view that can be selected using the tab at the bottom: “Connections by Level” and “Connections per Event-Type”.

Both tables show relations between event-types within the currently selected patterns and up to the level selected in the top left corner (see “Up to Level”) .

Table “Connections by Level”:

For the event-types listed in the first two columns this table shows the event-types that occur **later** in one or more of the same patterns up to binary pattern tree level selected, while the table entries indicate the lowest level at which this happens. For example, *x,order,viewer* (id=14) is followed by event-type *x,e,manipulate,viewer (id=33)* within at least one pattern, but this occurs first at level 6 so below that level no such relation exists in the selected patterns.

Up To Level= 7   Connections= 642   Terminals= 59   % Event-Types= 72

ID	Event Type	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23	e24	e25	e26	e27	e28	e29	e30	e31	e32	e33	e34	e35	e36	e37	e38	e39	e40	e41	e42	e43	e44	e45	e46
1	x,b,automanipulate		4			3						1		3	2				3			1	4			6						1		6				3			3					6	5
2	x,b,glanceat,partner										3								1				1								3							1									4
3	x,b,glanceat,pictcard																							1																							
4	x,b,glanceat,viewer																								1																						
5	x,b,haveviewer		4											1					2				4			5								5				3			1					6	5
6	x,b,immobile																																														
7	x,b,kneel																																														
8	x,b,laugh																												1																		
9	x,b,lie																																														
10	x,b,lookat,partner																														1																
11	x,b,lookat,pictcard	2	4	2															2			3	4	3								1						2			2						
12	x,b,lookat,viewer																		2														1					2									
13	x,b,manipulate,viewer		3																1				3			5								5				2			5					6	4
14	x,b,order,viewer		4			2								3					3			2	4			6								6				3			2					6	5
15	x,b,pull																																														
16	x,b,sit																																														
17	x,b,stand																																														
18	x,b,view,long		2																	3			2			5						4		5				1	3		3					5	4
19	x,b,view,short																		2																			3	1								
20	x,b,walk																																							1							
21	x,e,automanipulate	1	4			2						1		3	1				3				4			6						1		6				3			2					6	5
22	x,e,glanceat,partner										3								1												3							1									4
23	x,e,glanceat,pictcard																																														
24	x,e,glanceat,viewer																																														
25	x,e,haveviewer																																	1												2	
26	x,e,immobile																																														
27	x,e,kneel																																														
28	x,e,laugh								1																																						
29	x,e,lie																																														
30	x,e,lookat,partner																																														
31	x,e,lookat,pictcard	1	4	1								2							2			1	4	2														2									
32	x,e,lookat,viewer																		1																			2			3						
33	x,e,manipulate,viewer																																													1	
34	x,e,pull																																														
35	x,e,sit																																														
36	x,e,stand																																														
37	x,e,view,long		1								3								1				2			4					3	4		4							3					4	4
38	x,e,view,short																		1																			2									

Connections By Level   Connections Per Event-Type

Table “Connections per Event-Type”:

As in the previous table event-types are listed with their ID numbers in the first two columns, while the next two columns show its number of occurrences (Freq) in the data and the number of different event-types (NLinked) that follow it in at least one pattern up to the level selected, while their IDs are listed in the columns to the right. Here, for example, with maximum level 7 selected, *x,e,haveviewer* (id=25) that occurs four times in the data is followed with within at least one pattern by three different event-types with the ids 33, 45 and 55.

Up To Level= 7 Connections= 642 Terminals= 59 % Event-Types= 72

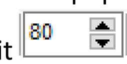
ID	Event Type	Freq	Nlinked	
1	x,b,automanipulate	16	29	2 5 11 13 14 18 21 22 25 31 33 37 40 45 46 47 51 52 53 55 59 62 67 68 69 73 74 75 77
2	x,b,glanceat,partner	11	10	10 18 22 30 37 46 59 62 68 75
3	x,b,glanceat,pictcard	5	1	23
4	x,b,glanceat,viewer	4	1	24
5	x,b,haveviewer	4	22	2 13 18 22 25 33 37 40 45 46 47 51 52 55 59 62 68 69 73 74 75 80
6	x,b,immobile	2	0	
7	x,b,kneel	1	0	
8	x,b,laugh	14	1	28
9	x,b,lie	1	0	
10	x,b,lookat,partner	4	1	30
11	x,b,lookat,pictcard	11	16	1 2 3 18 21 22 23 31 37 40 47 53 59 69 74 75
12	x,b,lookat,viewer	6	5	18 32 37 62 75
13	x,b,manipulate,viewer	4	21	2 18 22 25 33 37 40 45 46 47 51 52 55 59 62 68 69 73 74 75 80
14	x,b,order,viewer	7	25	2 5 13 18 21 22 25 33 37 40 45 46 47 51 52 55 59 62 67 68 69 73 74 75 77
15	x,b,pull	1	0	
16	x,b,sit	4	0	
17	x,b,stand	2	0	
18	x,b,view,long	22	26	2 19 22 25 31 33 37 38 40 45 46 47 51 52 53 55 57 59 62 68 69 73 74 75 78 80
19	x,b,view,short	9	6	18 37 38 47 69 75
20	x,b,walk	10	2	39 78
21	x,e,automanipulate	16	30	1 2 5 11 13 14 18 22 25 31 33 37 40 45 46 47 51 52 53 55 59 62 67 68 69 73 74 75 77 80
22	x,e,glanceat,partner	11	8	10 18 30 37 46 59 62 68
23	x,e,glanceat,pictcard	5	0	
24	x,e,glanceat,viewer	4	0	
25	x,e,haveviewer	4	3	33 45 55
26	x,e,immobile	2	0	
27	x,e,kneel	1	0	
28	x,e,laugh	14	1	8
29	x,e,lie	1	0	
30	x,e,lookat,partner	5	0	
31	x,e,lookat,pictcard	12	12	1 2 3 11 18 21 22 23 37 47 59 69
32	x,e,lookat,viewer	6	6	18 37 40 53 62 75
33	x,e,manipulate,viewer	4	1	45
34	x,e,pull	1	0	
35	x,e,sit	4	0	
36	x,e,stand	2	0	
37	x,e,view,long	21	23	2 10 18 22 25 30 31 33 40 45 46 51 52 55 57 59 62 68 69 73 74 75 80
38	x,e,view,short	9	5	18 37 47 69 75

Connections By Level | Connections Per Event-Type

Prediction

The following features are concerned more directly with prediction of event-types and detected T-patterns given occurrences of other event-types and detected T-patterns.

T-Markers

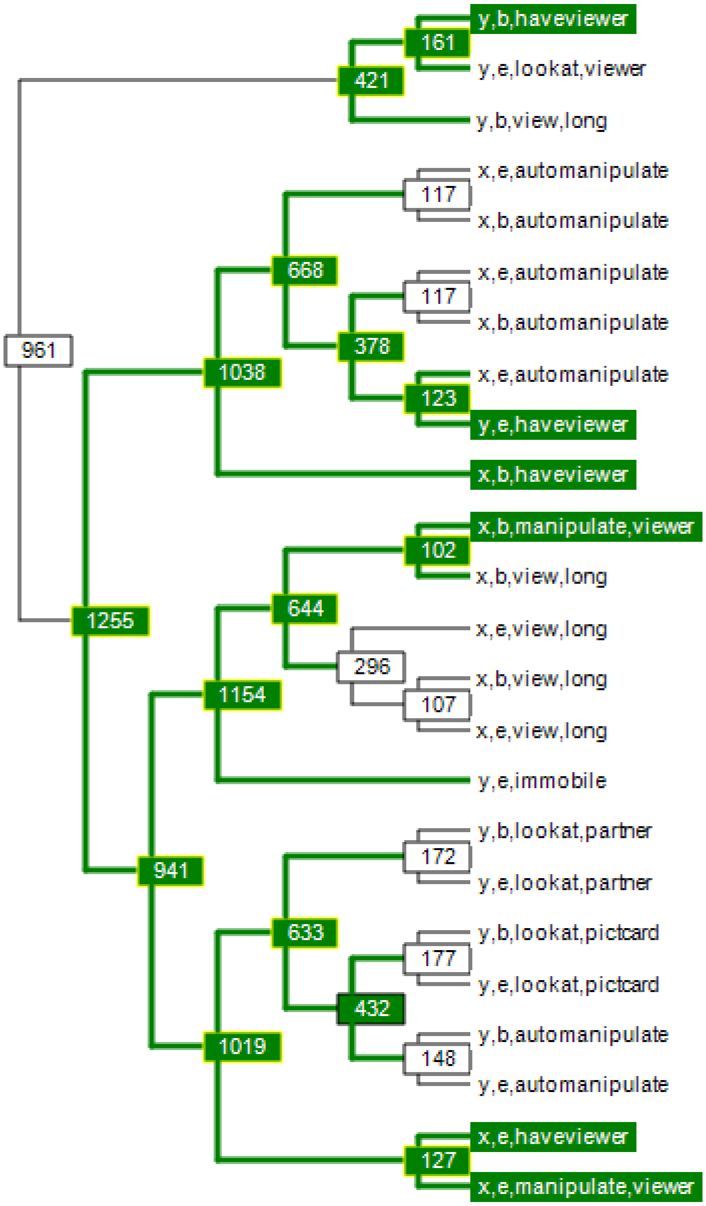
- For the current data, a k% T-marker component (Event-Type or T-pattern) indicates that a T-pattern is ongoing (in progress) with k% probability given an occurrence of that component. That is, if k% of the times component X occurs, it occurs as a component of T-pattern Q then X is a k% T-marker of Q.
- A T-marker that occurs before the end of a pattern thus predicts the following parts of the pattern with k% probability, while a T-marker occurring sometime after the first component in the pattern "retro-dicts" the earlier part(s) of the pattern. Such markers are called, respectively, T-Predictors and T-Retrodictors.
- Theme automatically finds all T-markers, T-Predictors and T-Retrodictors with k values equql to or higher than a cut off value specified by the user with the mouse and the spin edit 

to the right of the row of buttons just under the main menu. The default value is k = 80.

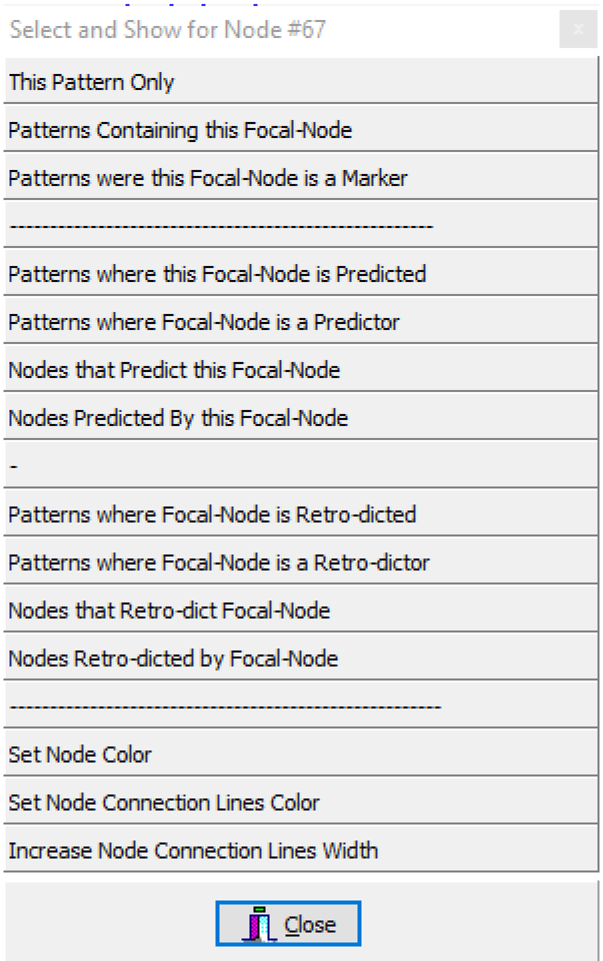
T-Prediction and T-Retrodiction (backward) of T-Targets

- Each k% T-Predictor may occur in a number of different T-Patterns where it may predict a number of following components called its **T-Targets**. Each T-Predictor may predict a number of T-Targets within the same or different T-patterns. Each T-Target may have more than one T-Predictor.
- Exactly the same is true for k% T-Retrodictors only backwards, that is, each k% T-Retrodictor may occur in a number of different T-Patterns where it may "retrodict" a number of preceding components called its T-Targets. Each T-Target may have more than one T-Retrodictor and each T-Retrodictor may retrodict a number of T-Targets within the same or different T-patterns.

Theme 6 has new interactive features to allow effective use of T-Markers, T-Predictors, T-Retrodictors and T-Targets. In the pattern tree diagram on the left in the *Pattern Diagrams* view, markers for the currently specified k value are optionally shown in green as below:



Clicking on any of the nodes, including the terminal event-type nodes, brings up the following menu of T-marker, T-predictor, T-retrodictor and T-target relations regarding the selected node, called Focal-Node, (here id = 67),

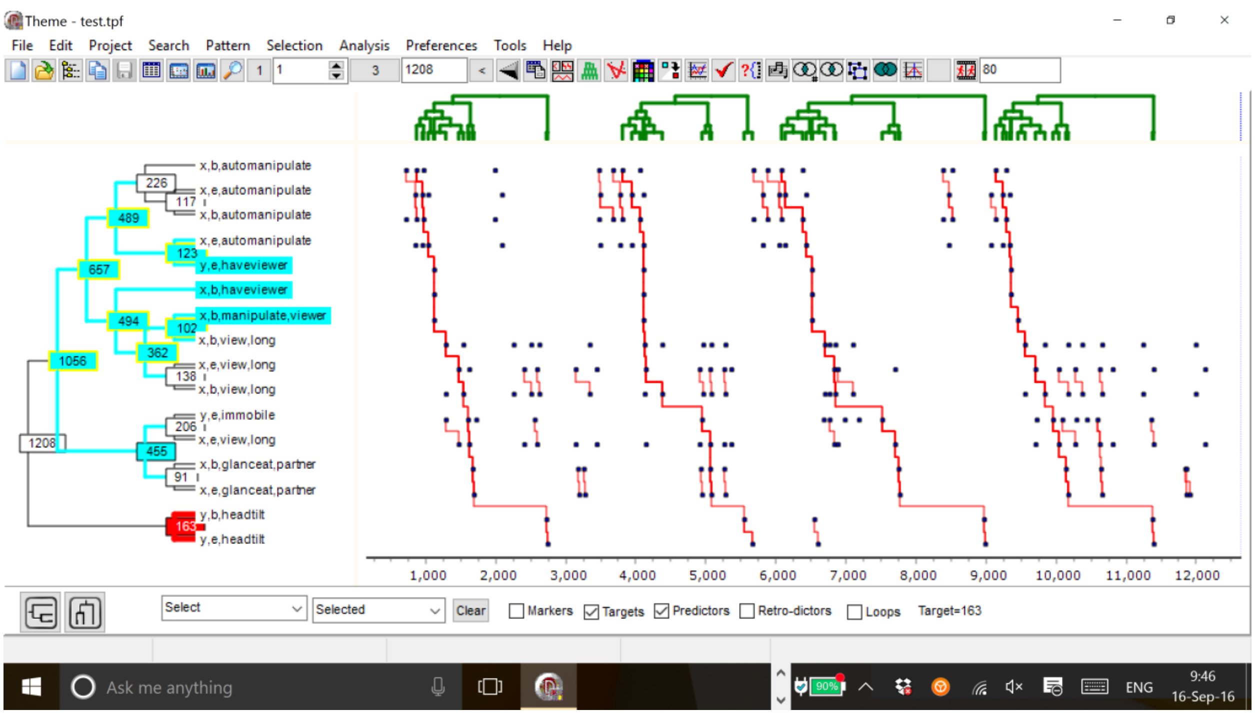


Selections made with this menu concern all the patterns in the current (possibly multi-sample) data file. That is, all patterns are first selected, including dependent patterns. This is different from selections made using the main menus where each selection is a sub-setting of the current selection and, by default, excludes dependent patterns. (Each selections with this special menu thus works as if including dependent patterns and then selecting all patterns.)

This menu also allows changing the looks of various aspects of the node and its connection lines.

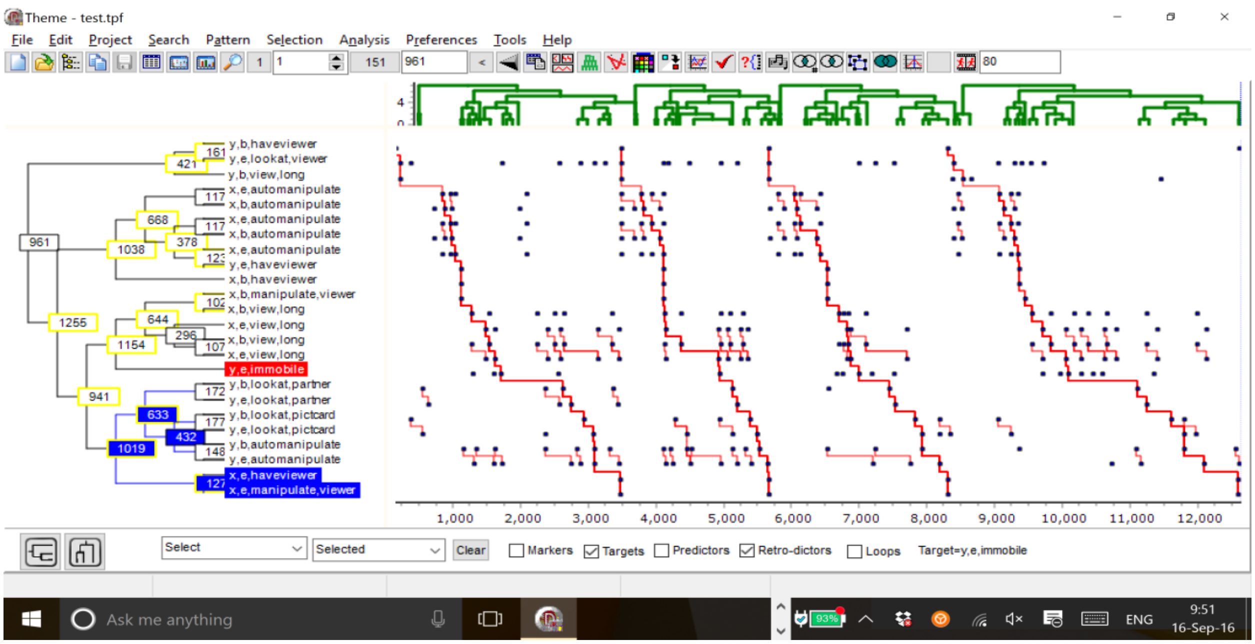
T-Prediction

Clicking on a potential T-Target (here 163) and using the above menu to ask for patterns where it is predicted will bring (if any are found) up a diagram of the kind below, where the target is marked red while its T-predictors in this pattern are (by default) marked by the aqua color.



T-Retrodiction

Clicking on a potential T-Target (here 163) and using the above menu to ask for patterns where it is retrodicted will bring (if any are found) up a diagram of the kind below, where the target is marked **red** while its T-retrodictors in this pattern are (by default) marked by the **blue** color.



Tables, Charts

Event-Types in Patterns

EventType	PercentFiles
e,b,que,tac,z4,nv	74
e,b,fou,tac,z4	61
n,b,fou,tac,z4	52
n,b,que,tac,z4	52
e,b,dat,tac,z4	48
e,b,fou,tac,z4,nv	48
n,b,que,tac,z4,nv	48
e,b,ord,tac,z4	39
n,b,dat,tac,z4	39
n,b,dat,tac,z4,nv	39
e,b,dat,tac,z4,nv	35
e,b,evn,tac,z4	35
e,b,que,tac,z4	35
n,b,fou,tac,z4,nv	35
e,b,evp,tac,z4	30
e,b,fou,tac,z6	30

Connections by Level

Up To Level= 7 Connections= 263 Terminals= 38 % Event-Types= 30

ID	Event Type	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14
1	dat,tac,z4,s														
2	e,b,com,reg,z5														
3	e,b,com,tac,z1														
4	e,b,com,tac,z3														
5	e,b,com,tac,z4								5						
6	e,b,com,tac,z4,s														
7	e,b,com,tac,z5														
8	e,b,com,tac,z6														

Up To Level= 7 Connections= 263 Terminals= 38 % Event-Types= 30

ID	Event Type	Freq	Nlinked											
1	dat,tac,z4,s	1	0											
2	e,b,com,reg,z5	1	0											
3	e,b,com,tac,z1	1	0											
4	e,b,com,tac,z3	5	0											
5	e,b,com,tac,z4	8	18	8	19	20	32	33	44	45	47	52	64	65
6	e,b,com,tac,z4,s	1	0											
7	e,b,com,tac,z5	2	0											
8	e,b,com,tac,z6	12	4	55	68	99	118							
9	e,b,com,tac,z6,s	1	0											
10	e,b,dat,reg,z2	1	0											

Event-Type Positions in Patterns

EventType	N	AsNODE	AsFirst	AsInner	AsLast	R_NODE	R_First	R_Inner	R_Last	First_p	Inner_p	Last_p
e,b,com,tac,z3	5	3	0	0	3	0.60	0.00	0.00	0.60	98.0000	98.0000	99.0000
e,b,com,tac,z4	8	7	6	7	6	0.88	0.75	0.88	0.75	0.5784	0.1066	0.2086
e,b,com,tac,z6	12	10	6	3	5	0.83	0.50	0.25	0.42	99.0000	99.0000	99.0000
e,b,dat,tac,z2	4	4	4	3	3	1.00	1.00	0.75	0.75	0.4636	0.6381	0.5615
e,b,dat,tac,z4	40	37	33	33	17	0.93	0.82	0.82	0.42	0.0048	1.74715941	99.0000
e,b,dat,tac,z4,nv	15	15	8	13	14	1.00	0.53	0.87	0.93	99.0000	0.0093	0.0005
e,b,evp,tac,z4,s	6	5	5	4	0	0.83	0.83	0.67	0.00	0.5303	0.5946	98.0000
e,b,fou,reg,z4	12	10	10	8	0	0.83	0.83	0.67	0.00	0.2113	0.3041	98.0000
e,b,fou,reg,z4,nv	13	13	8	13	13	1.00	0.62	1.00	1.00	99.0000	0.0008	0.0002
e,b,fou,tac,z2,nv	4	4	3	3	0	1.00	0.75	0.75	0.00	99.0000	0.6381	98.0000

...missing

Select Top-Patterns Only

Selects exclusively patterns that do not occur as sub-patterns of any other patterns.

Select Sub-Patterns Only

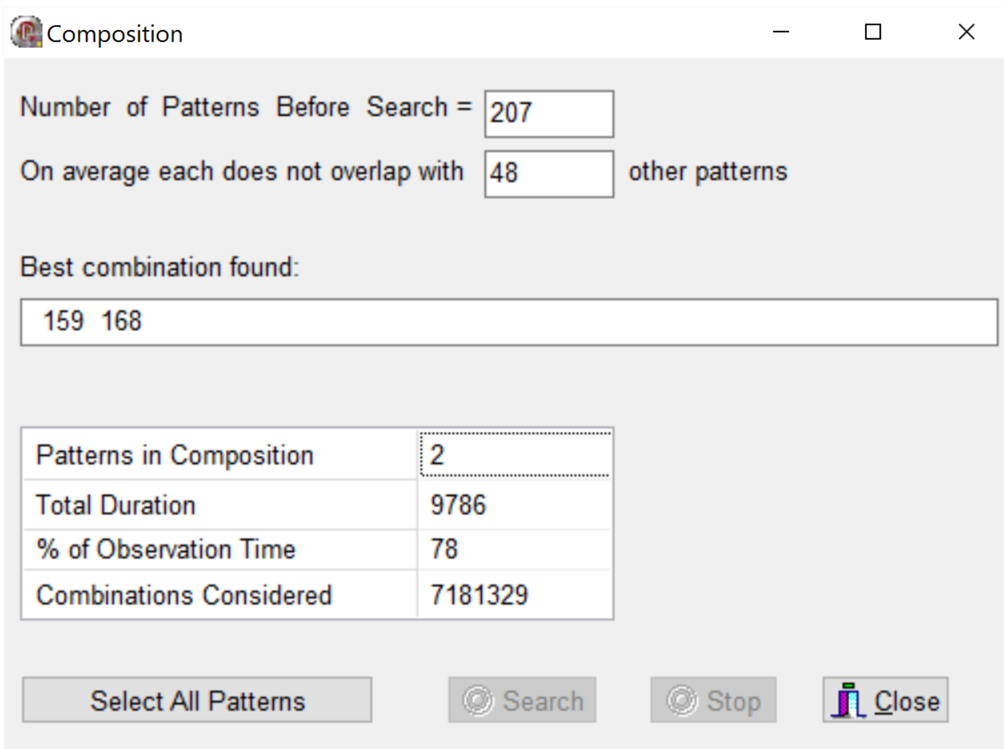
Selects exclusively patterns that only occur as sub-patterns.

Select Minimal Pattern Component Set

Selects exclusively patterns that together with the initial event-types are sufficient to compose all other patterns.

T-Composition

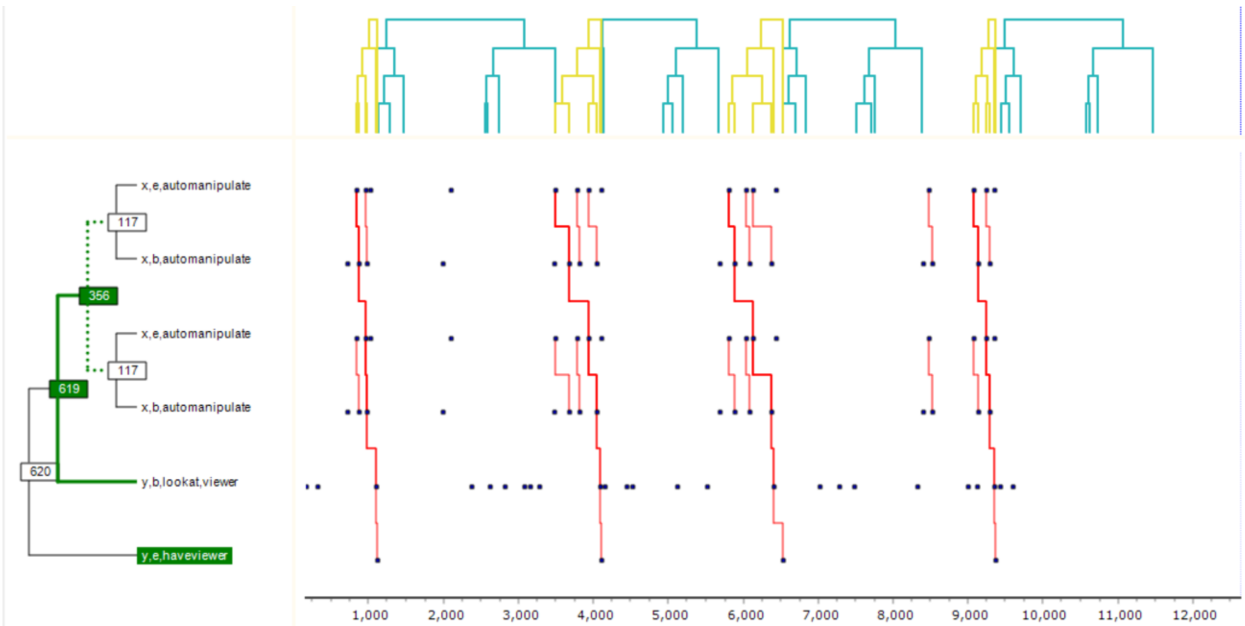
Searches the currently selected patterns for a subset such that no occurrences of its patterns overlap, that is, all the occurrences of all the patterns are mutually exclusive in time while the combined duration of the patterns subset is the of any other such combination. This search is exhaustive so if the patterns in the selection are many the processing may take a very long time and possibly fail. A fairly strict preselection of pattern is therefore recommended. Actually, if there an important such risk a warning is issued.



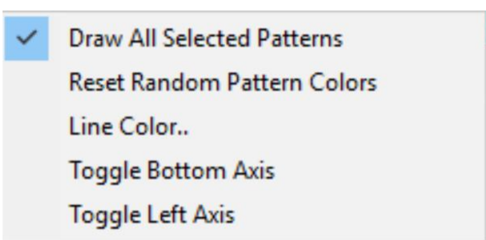
This image above appeared after a successful T-Composition search telling that the search started with 207 patterns selected and that on the average each of these patterns did not overlap with 48 other patterns. It also tells that there are only two patterns in the T-composition subset, Id=159 and id=168.

Their total duration is 9786 time units which means that their occurrences cover 78% of the observation time. Combinations considered in this exhaustive search were here 7181329.

By clicking the **Close** button, the two patterns in the T-composition set are selected and all their occurrences are shown at the top of the Pattern Diagram each in a different randomly set color as shown below:



Right-clicking the occurrence (top) diagram brings up this menu:



Each time "Reset Random Pattern Colors" is clicked a new random combination of colors is set,. Checking/unchecking the "Draw All Selected Patterns" menu item switches between single pattern and all patterns display mode.

T-Clusters

Here the currently selected patterns are clustered on the basis of at least one shared event-type. Each member in a **S**ingle **E**vent-Type **O**ccurrence (SEO) T-cluster occurs the same number of times and shares all the same occurrences of at least one event-type with at least one other in the cluster. This results in mutually exclusive clusters. A pattern thus belongs to only one cluster. This kind of clustering can serve to reduce the number of similar T-patterns arising from complex zones within the behavior (data stream).

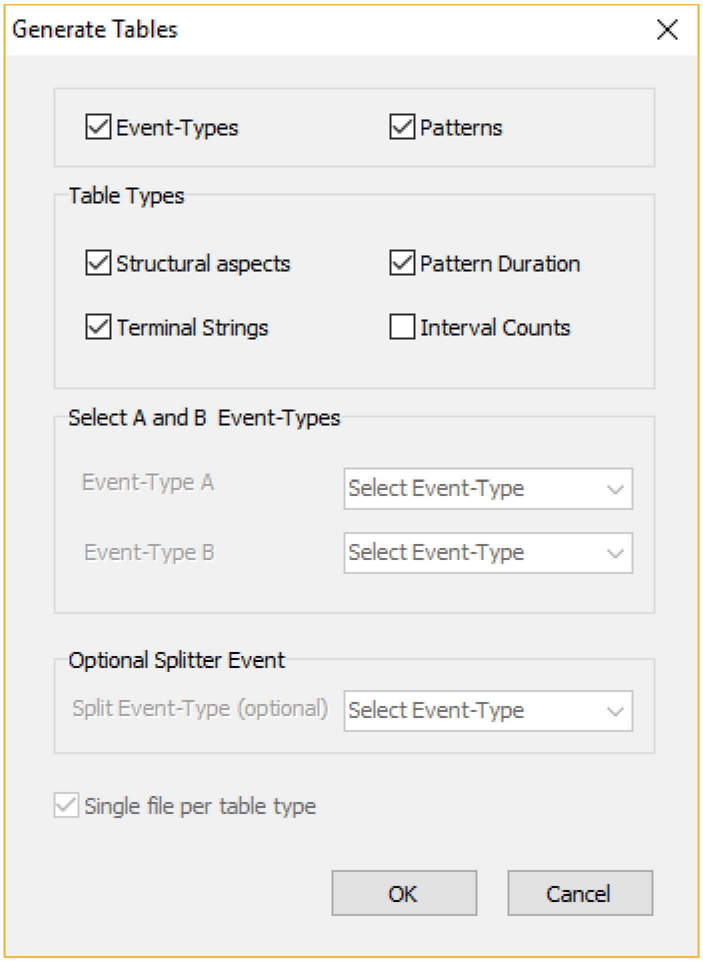
Largest Cluster

The patterns of the largest cluster are selected exclusively.

Longest From Each Cluster

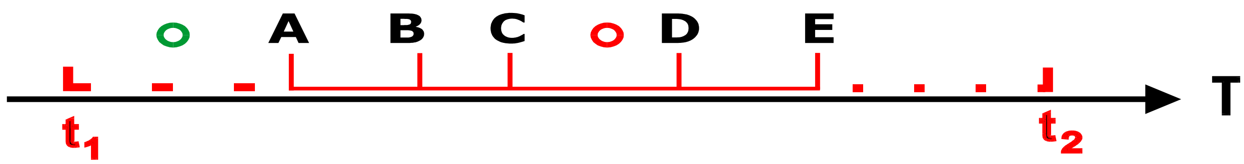
Here the longest pattern in each cluster is selected. The number of patterns selected is therefore equal to the number of clusters. (In Menu *Pattern* item *Select with Linked Patterns* the current pattern is exclusively selected together with all the other members of its SEO SI cluster.)

Generate Tables….



T-Packets

The T-packet combines T-Patterns (or Event-Types) and **T-Associates** and have positive and/or negative **Gravity Zones**



The above figure shows an instance of a T-packet with two T-associate instances. The positive gravity zone, [t_1, t_2], of a T-pattern extends from its earliest to its latest occurring positive associate.

The negative gravity or repulsion zone (not shown) is similarly the interval within which negative associates tend not to occur.

T-packets are thus simultaneously structured and non-structured repeated real-time patterns.

T-Associates

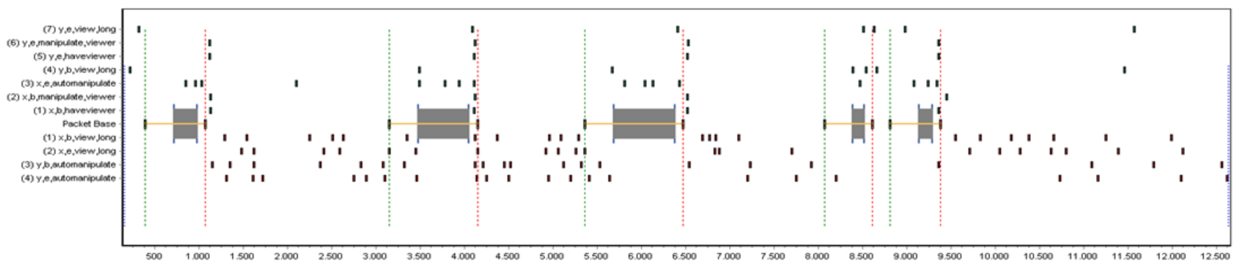
With X as an Event-type or a T-Pattern; a positive or negative (+/-) associate of X is some behavior that is not a part of X, but occurs during or around it significantly more (positive) or less (negative) of its occurrences than expected by chance.

Such associates may occur **only**, **always, sometimes** or **never** within or near their corresponding T-pattern.

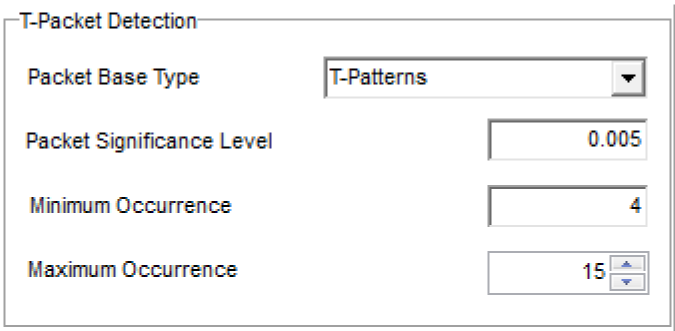
The **"only and always"** case is called a **T-Satellite.**

The ≈**never** case is called a **T-Taboo.**

a



The above figure shows a T-Packet, a T-Pattern with positive (above) and negative (below) T-Associates.



T-Packet detection was not available in previous versions and these are the T-packet search parameters:

Packet Base Type

T-packets may be discovered around detected T-patterns and/or Event-types as selected by the user, that is,

- ‘Event-Types’
- ‘T-Patterns’
- ‘Event-Types and T-Patterns’

Packet Significance Level

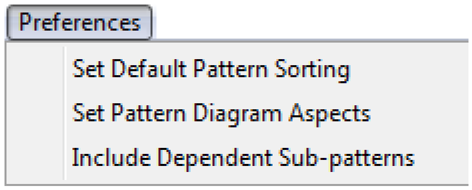
This is the significance level (here 0.005) for the detection of each T-associate and is be default set to the value used when the T-patterns were detected.

Minimum Occurrence

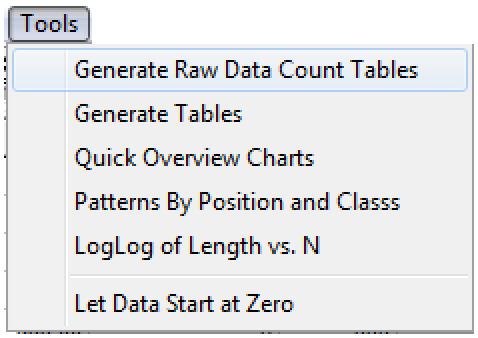
This sets the minimum number of occurrences for a T-packet (here set to 4).

Maximum Occurrence: As the algorithm for T-associate (and thus T-packet) detection becomes more computationally intensive with higher values, setting this parameter >15 may become overly time consuming and thus effectively impossible.

Preferences

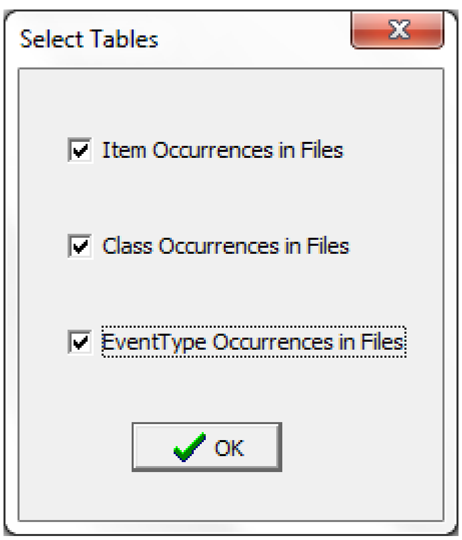


Tools

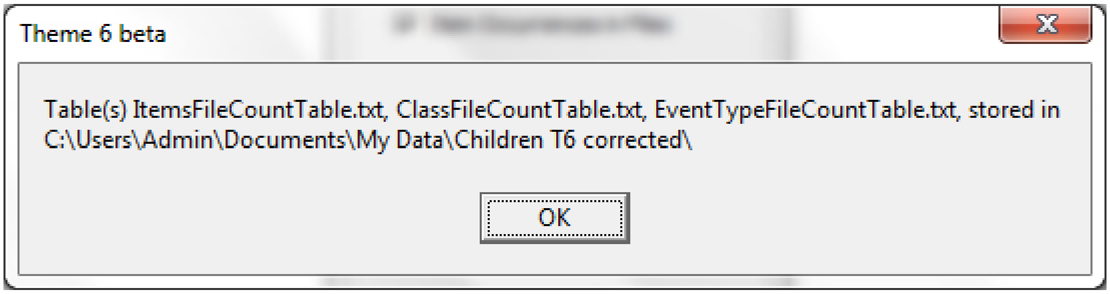


Most Theme tables that appear on the screen can be easily saved for further analysis by clicking the Save button and/or through the **File – Export** feature, but other tables can also be created using the following features which includes a number of options.

Generate Raw Data Count Tables



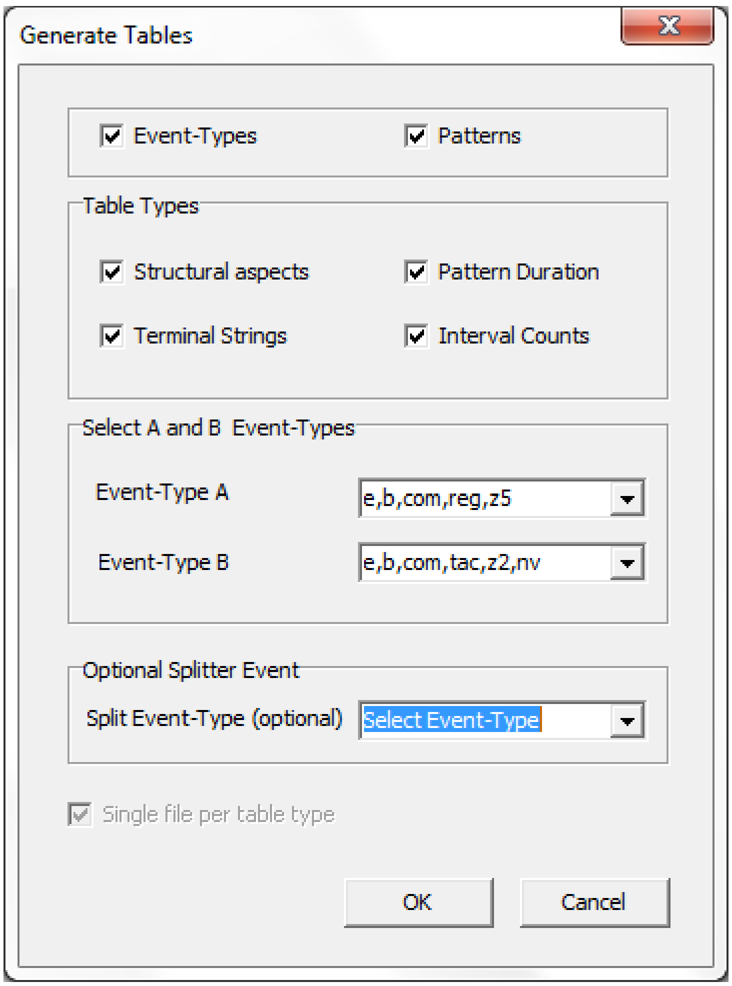
Here, one or more of three optional tables can be created. These tables contain one line per file containing the occurrence counts for either items, classes, or event-types depending on the table type. When the processing is complete a message specifies the names of the tab limited files generated and where they have been stored (always in the project directory):



Generate Tables

This feature allows the generation of tables concerning the occurrence of detected patterns (and the initial event-types) for statistical analysis. (Tab limited tables for easy entry into Excel, statistical packages, etc.).

This feature can only be used if patterns have already been detected.



The two first check-boxes decide whether Event-Types and/or Patterns should be considered. The Table Types options allow the generation of up to four different kinds of tables. The following variables are common to two or more tables:

Dataname the name of the data file, which may be a single- or multi-sample file.

Sample the serial number of the sample (always =1 for single-sample files)

Id the id number of the pattern or event-type

Length the number of event-types in a pattern

Level the highest depth (level) of the pattern

Structural Aspects {various aspects}

Dataname

Id

Length

Block {is a block}

Burst {is a burst}

Hasbursts {pattern contains a burst}

Hasloop {pattern contains a loop; one or more event-types recur as pattern terminals}

Packet {the pattern or event-type is a packet base}

Associates {number of associates}

Pattern Duration {the duration of pattern occurrences}

Dataname

Id

Sample

Starttime {start-time of pattern occurrence}

Endtime {end-time of pattern occurrence}

Duration {duration of pattern occurrence}

Terminal Strings {the terminal pattern strings of each patterns}

Dataname

Id

N

Length

Level

N_actors {number of actors in pattern}

N_switches {number of switches between actors within pattern}

Patstring	{the pattern as a string with parentheses}

Interval Counts {counts for event-types and/or patterns within intervals}

Dataname	
Id	
Sample	
N	{number of occurrences in sample}
Length	
Level	
Intervals	{number of occurrences of specified intervals [t1,t2] }
Between	{number of corresponding intervals [t2, t1] }
t1	{beginning-time of occurrence}
t2	{end-time of occurrence}
i_inside	{occurrence in [t1 , t2] }
i_between	{occurrence in [t2 , t1] }
Splittime	{the optional split time}
Beforeafter	{occurs before or after the split time}

Checking and modifying data and VVT

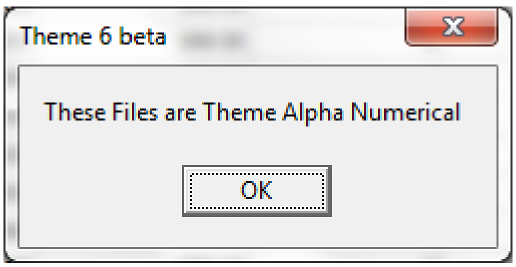
Theme provides tools for checking and correcting the vvt.vvt file and raw data (.txt) files: Click **Tools, Checking and Modifying Data and the VVT**

Syntax Check for Raw Data

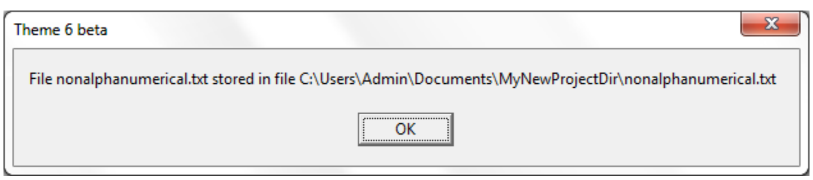
Legal characters in .txt raw data files are: ' ', tab, '0' to '9', 'a' to 'z', 'A' to 'Z', '!', '?', ';', '{', '}', '-', '.', ',', '=', '>', '<', '_', '&', ':', '*', '/', '"'

Here select the .txt raw data files you want to check – normally all the .txt files of the project – and click **Open**

If none of the selected files contains an illegal character this message will appear:



Otherwise a file (nonalphanumerical.txt) with error messages is created and its location is indicated:



Check the VVT.vvt file

Checks if a vvt.vvt file is present and whether it contains illegal characters.

Create a VVT.VVT from VVT.DAT

If a vvt.dat version of the category file exists (created with an earlier Theme version) in the project directory a vvt.vvt version will be automatically created. *The vvt.dat version should then be removed.*

Adapt .TXT Data Files and VVT.VVT

Corrections are here attempted of certain errors in the formatting of .txt raw data files vis-à-vis the vvt.vvt file.

To be revised

Working with multiple data sets

The search in multiple data sets (samples) can be done in two different ways:

1. Each data set is searched separately and the results for each are shown in the *Pattern Statistics* table. Note that in this case it is sometimes important that all the data files have the same duration and it this is not the case their duration can be set equal to the shortest one.
2. By concatenating data sets (joining together the samples one after the other) in a new **multi-sample data file**. The order of the samples may be important as it can represent the experimental design. As the samples are concatenated in alphabetical order this can be achieved by adding a letter at the beginning of data names. Concatenating data sets (samples) allows detecting of patterns that do not recur in any single sample. It is also particularly useful for exploring and the testing experimental effects as may be done using Theme features some of which are now to Theme 6.

Setting file duration equal to the shortest one

In order to obtain reliable statistics when using multiple data sets to search for effects of independent variables on pattern parameters, it is often important that all the data sets (files) have the same length. In practice this means that you have to set the file duration equal to the shortest one:

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.8).

Select **Set File Duration Equal To the Shortest** and click **OK.**

Click **OK** in the **Class Selection** dialog box.

To restore the initial file length

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

Select **Use Initial States** and click **OK.**

Click **OK** in the **Class Selection** dialog box.

Concatenating samples:

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

Select **Join All Data Sets into a Single File** and click **OK.**

Click **OK** in the **Class Selection** dialog box.

To restore the initial state

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

Select **Use Initial States** and click **OK.**

Click **OK** in the **Class Selection** dialog box.

Joining data sets by independent variable

Joining data sets by independent variable is useful when you want to test the effects of your independent variables.

To join data sets by independent variable

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

Select **Join Data Sets by Independent Variable**, select the independent variable of your choice and click **OK.**

Click **OK** in the **Class Selection** dialog box.

To restore the initial state

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

Select **Use Initial States** and click **OK.**

Click **OK** in the **Class Selection** dialog box.

Changing the time scale

The types of patterns that **Theme** detects are independent of scale. The program is capable of analyzing data whatever time units are used, micro-seconds, hours, days, months, years or even longer. In fact, the data need not be time-based: they just need to be collected along some form of axis. If you want to see the effect of reducing the amount of information about timing, you can change the time scale to ordinal:

1.
From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

2.

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

3.

Select **Ordinal Time Scale** and click **OK.**

4.

Click **OK** in the **Class Selection** dialog box. A message appears asking you whether you want to restore the original time scale. Click **No**.

To restore the initial time scale

1.

From the **Project** menu, choose **Select Classes**. The **Class Selection** dialog box opens (see Figure 7.5).

2.

Select **Advanced**. The **Advanced Options** dialog box opens (see Figure 7.10).

3.

Select **Initial Time Scale** and click **OK.**

4.

Click **OK** in the **Class Selection** dialog box.

Working with graphs and charts

Parts of most graphs can be expanded by selecting a part of a graph with the mouse: holding the left-button and moving the mouse.

Many graphs can be copied to the clipboard using Copy in the Edit menu and then pasted into documents such as, for example, Word and PowerPoint documents.