

# Tropes<sup>®</sup>

Version 8.1

## Reference Manual

Software developed by Pierre Molette and Agnès Landré  
on the basis of the work of Rodolphe Ghiglione

The Romanian dictionaries were developed by Dan Caragea  
for UEFISCDI, Romania.

This version of Tropes software is distributed by:

Semantic-Knowledge  
[www.semantic-knowledge.com](http://www.semantic-knowledge.com)  
All rights reserved

and  
UEFISCDI  
[www.forhe.ro](http://www.forhe.ro)  
All rights reserved

## Table of contents

Introduction .....	6
CHAPTER 1 – Analyzing a text .....	7
Saving a document in “text” format .....	8
Starting a text analysis .....	8
Examining a text .....	9
Displaying the context.....	10
What is the text Style? .....	11
Most Characteristic Parts of text.....	12
Reference fields.....	13
References: what is the text about? .....	13
Actants and Acted .....	14
Which elements are frequently connected?.....	15
Note on Equivalent classes.....	15
Which word categories are frequently used? .....	16
Displaying all word categories.....	16
Lists of verbs and adjectives.....	17
Episodes and Bundles.....	18
Analysis options.....	19
Printing the results .....	21
Equivalent classes and Relations graphs.....	22
Bundles and Episodes graphs .....	26
General operation of graphs .....	27
Copying the graph .....	28
Copying the selected text.....	28
Graph properties .....	28
Changing the colors.....	28
Finding a word.....	30
Report writer .....	31
CHAPTER 2 – Semantic Scenarios - Ontologies .....	32
Introduction .....	33
Creating semantic Scenarios .....	34
Connecting an Equivalent class to a Semantic Group .....	36
Applying the Scenario to the text.....	38
The methodology of <i>Scenario</i> design .....	39
Scenario management .....	41
Terminology extraction .....	42
CHAPTER 3 - Borders .....	45
Corpus segmentation .....	46
Creating Borders .....	47
Borders files.....	48
CHAPTER 4 – Introduction to text analysis.....	49
How Tropes operates .....	50
Propositional hashing.....	51

## Tropes – Reference Manual

Ambiguity solving .....	51
Word categories .....	51
Statistical, probabilistic and cognitive analyses .....	54
Equivalent classes and Relations between equivalents .....	55
Reflections on the size of the texts .....	56
Analysis of heterogeneous utterances: open questions, dispatches, enumerations, etc. ....	57
Analysis of discourses and conversations .....	58
Literary studies .....	58
Comparing two texts .....	59
CHAPTER 5 - Appendices .....	60
Files conversion .....	61

### Important notice:

- This first Romanian version is a prototype.
- It only accepts Romanian texts, properly written, with diacritics.
- The conversion of PDF<sup>®</sup> documents may require the installation of an Adobe<sup>®</sup> IFilter component on your system. You can download the latest version of IFilter on Adobe's site.
- Romanian dictionaries and software grammar are in continuous evolution.
- Please download always the latest versions from our sites.
- For specific linguistic topics, specialized dictionaries, etc., please contact us.

### Typographical conventions:

The Windows commands, accessible by the menus or the dialog boxes, and the text referring to the buttons are written in *italics* and in square brackets *[]*. For example:

- *[File][Open]* refers to the *[Open]* command of the *[File]* menu.
- *[Cancel]* refers to the *[Cancel]* button of a Windows dialog box.

The specific vocabulary of the software is written in italics (for example: *Equivalent classes*).

### Notes:

Though it is possible to install this software in a different folder than the default folder, we will consider, in this manual, that the home folder of Tropes is named "tropes" and is located on the hard drive "c:", i.e. *[c:\Program files\Tropes\Tropes...]*

Most of the drawings used in this documentation have been captured in Windows XP. Although the aspect of windows and dialog boxes may slightly vary from that of other versions of Windows.

### Legal notice:

Please read **TROPES LICENSE AGREEMENT** before using this software.

Tropes software was developed by Pierre Molette and Agnès Landré on the basis of the work of Rodolphe Ghiglione. All rights reserved.

The dictionaries for Romanian version were developed by Dan Caragea for UEFISCDI, Romania. All rights reserved.

Redistribution and use in binary forms are permitted provided that the following conditions are met:

1. Tropes software must be distributed for free. Reselling this software or its components is strictly forbidden.
2. Redistributions of the software must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

## Tropes – Reference Manual

3. Redistributions of any form whatsoever and all advertising materials mentioning features or use of this software or its components must display the following acknowledgment: "Tropes software was developed by Pierre Molette and Agnès Landré on the basis of the work of Rodolphe Ghiglione. The dictionaries for Romanian version were developed by UEFISCDI. All rights reserved. The dictionaries for Romanian version were developed by Dan Caragea for UEFISCDI, Romania. All rights reserved. <http://www.semantic-knowledge.com>; <http://www.forhe.ro> "

4. The name "Tropes" and the name of the authors must not be used to endorse or promote products derived from this software without prior written permission.

5. Products derived from this software may not be called "Tropes" nor may "Tropes" appear in their names without prior written permission of the authors.

THIS SOFTWARE IS PROVIDED " AS IS " AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

The information contained in this manual may undergo alteration without notice.

Tropes, Adobe, Microsoft, Windows, Excel, Word and Apple are registered trademarks.

To complement this manual, consult the Tropes user's guides on our website (<http://www.semantic-knowledge.com> and <http://www.forhe.ro> ).

1<sup>st</sup> edition, March 2012.

### Introduction

What is the content of a text? Or to be more precise: what are the core elements which must be identified in order to grasp the essential meaning of a text?

Whether it is a press article, a book, a speech, or any other sequence of language, every text contains a few key sentences conveying the ideas that make up its framework: its skeletal structure. The problem, then, is to locate this central core of the text that holds the essentials of its meaning. This is the crucial first step, before any attempt at interpretation can be made.

We can say that a text consists of various worlds in which different actors do, form or say various things in combination with other actors. And we can say that these worlds – which are invariably propositional in form – have different levels of importance in the structure of the text. And finally, we can say that some of these worlds – a very small proportion of them – constitute the foundations of the text, in that if they were removed, the "textual construction" would collapse, and the meaning would be lost.

Content analysis, then, applies a set of techniques to a given text to determine:

- the identity of the main actors,
- the relations in which they stand to each other,
- the hierarchy of these relations and how they evolve.

To sum up, content analysis consists in revealing the framework of a text, i.e. its meaning. This necessarily implies two things. First, there must be a theoretical conception of the text: this must describe both the textual organization of the things that are said and the structural organization of the thought-processes of the people who say them. Secondly, it implies the use of a tool derived from this theoretical conception and which rigorously excludes the subjectivity of the investigator – at least until the analysis is finished.

We now propose to introduce you to such a tool: **Tropes**.

# Tropes<sup>®</sup>

## CHAPTER 1

### Analyzing a text

### Saving a document in “text” format

You must save the documents in the text format: Microsoft Word files, UTF-8 (Windows), HTML (web pages), RTF, or other. These documents must have a text file extension (myfile.doc, WebPage.htm, etc.) Limitations may exist for some of these formats. Consult the software user's manual for further details.

We recommend that you save your texts in the [Text only, UTF-8] or web page format (HTML). To carry out a file conversion, use the [Text only] option of the [Save as] command of your word processor. You can save web pages on your hard drive by means of your web browser or of a "web-spider".

The text may include punctuation marks, capitals or special characters (such as parentheses, numbers, percentages, etc.) Only alphabet letters and punctuation characters will be used during the analysis. If you want a series of words to be considered as a single word, link these words together with the underline character "\_" (for instance: șeful\_statului).

To get better results, use punctuation and respect the typographical conventions in use in Romanian. The spellchecker of some word processors can enable you to ensure that these conventions are fully respected.

### Starting a text analysis

To analyze a text, start *Tropes*, use the [*File*][*Open*] command, select a file, and click on [*OK*].

You can open several files simultaneously by using the [Ctrl] and [Shift] keys of your keyboard when opening files.

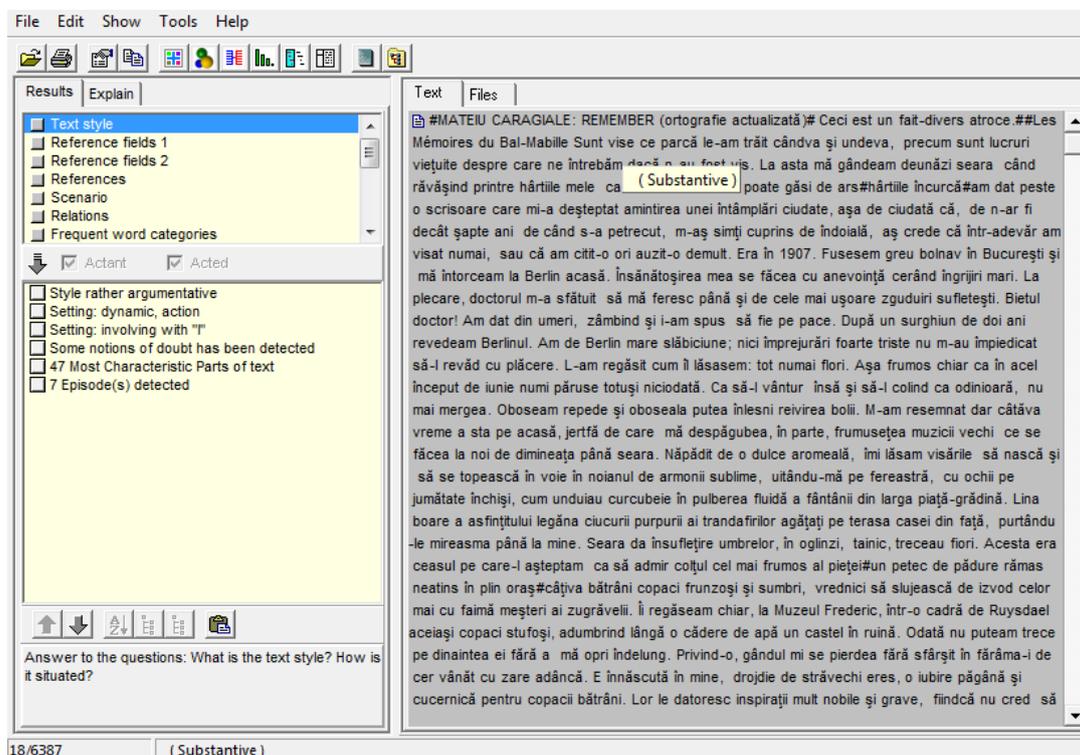
It is also possible to use the *drag & drop* function of Windows Explorer: grab a text with the mouse and drop it on the main window of *Tropes*.

After starting to run the analysis, you can also use the [File] tab to add or delete documents, display them, open them, sort them, etc.

### Examining a text

As an introduction to the software, we suggest the analysis of the example file: *Remember.txt*.

Once a text is opened, and as soon as the analysis is over, its content is displayed:



The main analysis results are displayed inside the frames on the left of the screen, while the text and the graphs are shown on the right.

This window “responds” according to the text:

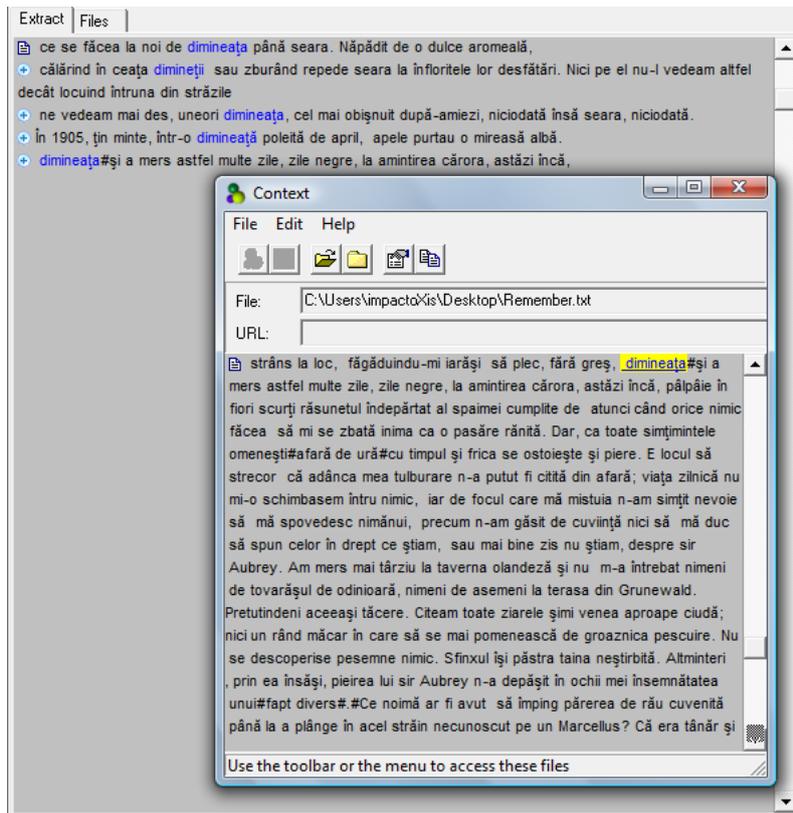
-When the **mouse cursor** is placed on a word of the text, the message bar at the bottom of the screen shows the category of this word (in the above example, this word is a “substantive”). In this case, the *[nn/ww]* indicator on the left of the message bar indicates the occurrence **number** of the word pointed at by the mouse.

-When the display applies to a subset of the text, only the categories concerned appear in color. To display the whole text, use the *[Show][All the text]* command or the *[All]* button of the tool bar:

### Displaying the context

To display the context of a proposition, click on the icons on the left margin of the text extracts, or point at any word, then press the right button of the mouse, and select the *[Context]* command in the context menu.

The display of the context is similar to that of the main window (syntactic coloring of words, etc.)



This tool proves useful when the display applies to a specific part of the text (i.e. when all the propositions are not displayed). It can also be used to store temporarily a passage of particular interest to you.

To quit this dialog, use the *[system menu]* in the top left corner of the window.

### What is the text Style?

Click on the *[Text Style]* line of the Result frame, or use the *[Show][Text Style]* menu.

The software makes a diagnosis of the **Text Style** and of its **Setting** according to the statistical indicators retrieved during the analysis. Here are the possible **Styles**:

Style	Explanation:
Argumentative	the speaker involves himself, argues, explains or analyzes in order to try to convince the interlocutor
Narrative	a narrator states a series of events, happening at a given time, and in a given place
Enunciative	the speaker and the interlocutor establish a mutual relation of influence, make their standpoints known
Descriptive	a narrator describes, identifies or classifies something or somebody

Here are the possible verbal **Settings**:

Setting	The Setting is expressed by means of:
Dynamic, action	action verbs
In the real	verbs conveying the idea of being and having
Involving the narrator	verbs helping to make a statement about a given state, an action
<i>Involving with "I"</i>	numerous pronouns in the first person singular ("I", "me", "myself", ...)

Also, you will immediately know whether or not: *notions (adverbs) of doubt* have been detected.

To obtain an explanation of the displayed Style (or of the *Setting*), click on the line concerned: a color display in the main window will show you all the words whose categories have been taken into account to make the diagnosis.

The study of the *Text Style* and of the *Setting* of a text written by you (directly or indirectly) is especially interesting when the software makes a diagnosis that proves contrary to your purposes. For instance, you will presumably want to find out why the Style has been detected as *[argumentative]* when it was not intended to be. Likewise, you will probably try to avoid as much as possible the *Settings [involving with "I"]* if your text is written on behalf of a group, or to rule out all *[notions of doubt]* from a financial offer, a contract, etc.

Important note: since the analyses are carried out on a statistical basis, the studied texts must be of sufficient length for the results to be significant.

## Most Characteristic Parts of text

To display the *Most Characteristic Parts of text*, click on the [nn Most Characteristic Parts of text] line of the [Text Style].

The screenshot displays the Tropes software interface. On the left, the 'Results' pane shows a tree view with 'Text style' expanded to '42 Most Characteristic Parts of text', indicating that 8 episodes were detected. The main 'Extract' pane shows a list of 42 extracted propositions from a text, each starting with a blue arrow icon. The propositions are:

- ↳ nu întorceam la Berlin acasă. Însănătoșirea mea se făcea cu anevoință cerând îngrijiri mari.
- ↳ ca să admir coțul cel mai frumos al pieței#un petec de pădure rămas neatins în plin oraș#cățiva bătrâni copaci frunzoși și sumbri,
- ↳ că se află nelipsit un tânăr, care acolo mai ales, ar fi atras privirile oricui,
- ↳ și oase o icoană din veacuri apuse? Cu doi ani înainte văzusem în sala franțuzească a muzeului o coconță care copia după Mignard pe Maria Mancini și avea o așa izbitoare asemănare cu modelul,
- ↳ în deplină frumusețe, pășea singur în viață, nepăsător, cu fruntea sus. L-am crezut, din capul locului, una din acele făpturi excepționale, străine de omenire,
- ↳ Alături de mine, pe singura laviță din pritoarea cămară, singuratică ziua, tânărul cu chip de portret vechi sorbea pe îndelete băuturile cele mai dulci și mai parfumate, asemenea unor nestimate topite, atâțătoare de visări
- ↳ Acolo, parcă nu mai eram străini unul de altul și, ceea ce e nostim, mai târziu,
- ↳ una e floarea de câmp, alta floarea de grădină. Acum, fie că trebuiseră veacuri ca,
- ↳ Numele lui normand #și până astăzi nu știu dacă astfel se cheamă într-adevăr#numi era străin,
- ↳ când la una când la cealaltă, sclipeau șapte inele, gemene toate#șapte safire de Ceylon.
- ↳ Cu brățara și cu parfumul, acea de neuitat mireasmă de garoafă roșie, inelele erau singurele lucruri
- ↳ Părea chiar să fi avut mai multe legături cu duhurile decât cu cei vii deoarece în povestirile sale nu venea niciodată vorba de ființe omenești.
- ↳ O veșnicie să ne fi întâhnit, tot mai lesne i-ar fi scăpat lui o destăinuire decât mie o întrebare.
- ↳ să intru să-mi înfioresc cheutoarea ca să aflu unde le trimesese și, punând astfel mâna pe fir,
- ↳ să fi avut un scop stăruința încăpățanată cu care își perdeluia scurtul lui trecut și viața de toate zilele se putea prea bine,
- ↳ să se risipească orice umbră de neîncredere sau de bănuială. Totuși, numi scăpase din vedere
- ↳ De la acea scrisoare, sir Aubrey n-a mai dat semn de viață. Nu era de mirare;
- ↳ Trecea o femeie înaltă, cu un bogat păr roșu sub o pălărie mare cu pene, o femeie slabă și osoasă, fără șolduri și fără săni, într-o rochie strâmtă de fluturi negri.
- ↳ cu nările pline de cunoscuta mireasmă#mireasma de garoafă roșie#am dat să o urmăresc.
- ↳ Era o noapte de catifea și de plumb, în care adierea molatecă a unui vânt fierbinte cerca în zadar să risipească păcla
- ↳ să cred că arătarea ce trecuse pe lângă mine era o femeie, acum mi se părea
- ↳ Rămase astfel înmărmurit câteva timp scrutând întunericul pe care îl biciu deodată cu mânușile lui albe ca și cum ar fi voit
- ↳ Trebuie să te simți și dumneata slăbit de zăduf, urmă el. Ai să-mi faci plăcerea
- ↳ poate mai mult de un sfert de ceas, oricum însă mai puțin de o jumătate.
- ↳ Cum așteptarea, ca toate neajunsurile, pare mai grea la început, am mai luat-o pe cheu în sus,

The contraction of the text reveals the *Most Characteristic Parts of text*. These are “propositions introducing main themes or characters, expressing events that are essential to the progression of the story (causal attributions of consequences, results, aims)”.

To extract these propositions, Tropes carries out a complex Cognitive-Discursive Analysis processing (CDA). To simplify matters, let us say that each proposition of the text is allotted a score, depending on its relative weight, its occurrence order and its argumentative role. The propositions are then sorted out according to their respective scores. To enable you to control the amount of displayed propositions and to insure that the result obtained reflects the text analyzed, Tropes provides the means to adjust the contraction rate of the text (see *Analysis options* below).

It must be stressed that the *Most Characteristic Parts of text* offer significance only when studying a monolithic and structured discourse, of moderate length. In no circumstances will they constitute a summary of the text (this would require a rewriting of the text).

### Reference fields

Click on the *[Reference fields 1]* line of the Result frame, or use the *[Show][Reference fields 1]* menu.

This function displays, in decreasing frequency, the *Reference fields* of the words in the text. Each line consists of a *field*, preceded by a counter showing the number of words contained within this particular *field*. Only significant *fields* are displayed.

The **Reference fields** represent the context and group together the main substantives of the text analyzed into *Equivalent classes*. The software detects the *Reference fields* by using two different representation levels of the context (*Reference fields 1 and 2*).

To view the content of a *field*, select this *field* and all the words that compose it will then be displayed in blue in the main window:

The screenshot shows the software interface with the following components:

- Results pane (left):** A tree view showing 'Reference fields 1' selected. Below it, a list of fields is shown, with '0077 corp' highlighted in blue. Other fields include '0192 timp', '0061 senzații\_și\_percepții', '0058 om', '0055 plante', '0052 existență', '0052 mediul\_inconjurător', '0044 mișcare', '0042 spațiu', '0032 intelect', '0025 îmbrăcăminte\_și\_încălțăminte', '0023 statele\_lumii', '0023 materie', '0019 ordine', '0018 cantitate', and '0017 sănătate\_si\_boală'.
- Text pane (right):** A text window showing a paragraph of Romanian text. Words from the selected field 'corp' are highlighted in blue: 'umeri', 'zâmbind', 'și', 'i-am spus', 'să fie pe pace', 'ochii', 'pe jumătate închiși', 'cum unduiau curcubeie în puberea fluidă a fântânii din largă piață-grădină', 'Lina boare a asfințitului legăna ciucurii purpurii ai trandafirilor agățaji pe terasa casei din față', 'împodobindu-l, propriul ei chip', 'Tot astfel semăna tânărul cu unii din acei lorzi', 'ale căror priviri, mâini și surâsuri Van Dyck și, după el, Van der Fae le-au hărăzit nemuririi', 'în deplină frumusețe, pășea singur în viață, nepăsător, cu fruntea sus', 'L-am crezut, din capul locului, una din acele făpturi excepționale, străine de omenire', 'Alături de mine, pe singura laviță din printoarea cămară, singuratică ziua, tânărul cu chip de portret vechi sorbea pe îndelete băutura cele mai dulci și mai parfumate, asemenea unor nestimate topite, ațățătoare de visări', and 'Sărea în ochi: una e floarea de câmp, alta floarea de grădină. Acum, fie că'.
- Graph pane (bottom right):** A bar chart titled 'corp' showing the frequency of each word. The x-axis represents the words and the y-axis represents the frequency. The bars are blue and have their values labeled below them: 3, 5, 9, 4, 3, 12, 9, 4, 11, 8, 2, 4, 3. A horizontal dashed line is drawn across the chart at the frequency level of 3. The total count '6389' is shown at the bottom right of the graph.

Note: to detect these *Reference fields*, the software uses a *semantic equivalents* dictionary which does not contain all of the Romanian words (see *Note on Equivalent classes* below); only the most significant substantives of your text will be displayed, along with some proper nouns.

### References: what is the text about?

Click on the *[References]* line of the Result frame, or use the *[Show][References]* command

This function displays, in decreasing frequency, the *References* of the words in the text. Each line consists of a *Reference*, preceded by a counter showing the number of

## Tropes – Reference Manual

words contained within this particular *Reference*. Only significant *References* are displayed.

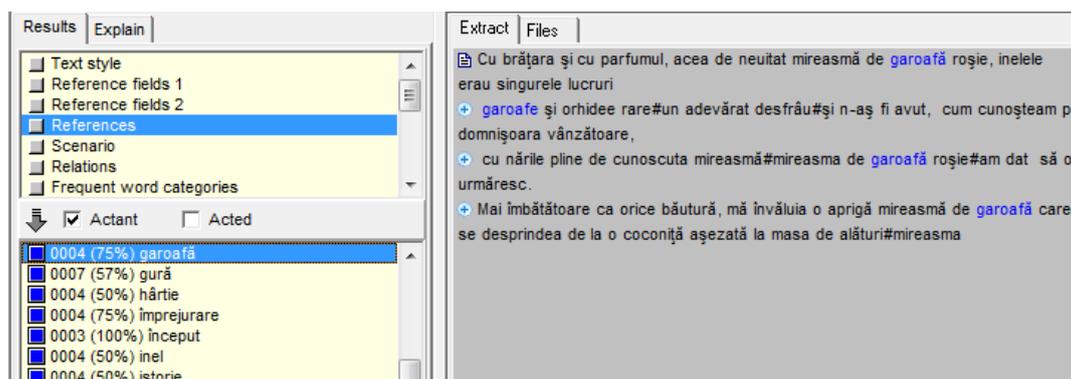
The *References* group together closely related common and proper nouns into *Equivalent classes* (for example, “ochi” and “mână” are grouped together into the “corp” *class*).

To view the content of a *class*, click on this *class* and all the words that comprise it will then be displayed in red in the main window.

### Actants and Acted

This option enables you to tell the position of the *Reference fields* (higher level *Equivalent classes*) and of the *References* (lower level *Equivalent classes*, see below); both being generally placed either as:

- Actant**, i.e. before the verb (and often subject of this verb),
- or as **Acted**, i.e. after the verb.



When checking one of these boxes only, the list shows the number of times (percentage between parentheses) the corresponding *class* has been in the position of *Actant* in the studied text. The same counter will appear when you print the *classes* list.

To get back to the standard display, check both the [*Actants*] and [*Acted*] boxes.

The identification of *Actants* and *Acted* constitutes one of the essential steps in text analysis. In fact, when a significant *Reference field* (or a *Reference*) clearly appears in the position of *Actant* (percentage above 60%), it is assumed that the notion it represents *carries out the action*. Otherwise, when a significant *field* (or a *Reference*) clearly appears in the position of *Acted*, it is assumed that the notion it represents is *subjected to the action*.

### Which elements are frequently connected?

Click on the *[Relations]* line, or use the *[Show][Relations]* command.

This function displays, in decreasing frequency, the strong *Relations* between the various *Equivalent classes*. Each *Relation* is preceded by a counter showing its occurrence frequency within the studied text.

The **Relations** show which *Equivalent classes* are frequently connected (i.e. found in the same proposition) within the text analyzed.

To display the propositions corresponding to a given *Relation*, click on the line of your choice.

*Relations* are oriented according to the occurrence order of the words comprising them.

By default, *Relations* are built on the *References*. It is possible to define the construction level of the *Relations* by using the *[Tools][Analysis options]* command (see *Analysis options* below).

There is little room for chance in the display of *Relations*: finding two *Equivalent classes* several times in the same text in the same order is indeed unlikely to happen. When it does, it means that these two *classes* are strongly connected, and this reveals the notions emphasized by the author of the text (but not necessarily what he intended to put into the text).

The display of the *Reference fields*, of the *References* and of their *Relations* brings you to the heart of the discourse: all the actors, objects, things and concepts presented in the text will appear before you in decreasing order of importance.

### Note on Equivalent classes

In this manual, the term *Equivalent classes* refers equally to *Reference fields* and to *References*. For further details about *Equivalent classes*, consult Chapter 4: "Introduction to text analysis".

Note about the dictionaries:

Since it is neither possible nor relevant to classify all of the English substantives (names, forenames and proper nouns), the software automatically generates *Equivalent classes* for all the words that are not referenced in the dictionary. Such *generated classes* are visible only in the *References*.

-The *generated classes* ("Other", for instance) are preceded by a blue square, whereas the *Equivalent classes* detected by the software are preceded by a red square.

-To group *generated classes* together with *Equivalent classes*, and so create your own personal classification, you have to use a *semantic Scenario* (see chapter 2).

### Which word categories are frequently used?

Click on the *[Frequent word categories]* line of the Result frame, or on the *[Show][Frequent word categories]* menu.

This function displays the most significant *Word (meta)categories* of the studied text.

The screenshot displays a software interface with two main panes. The left pane, titled 'Results', shows a hierarchical tree view of word categories. Under 'All word categories', the following categories are listed with their respective percentages and counts:

- \* Verbs :
  - Factive 58.4% ( 703 )
  - Stative 29.2% ( 351 )
  - Reflexive 12.4% ( 149 )
  - Performative 0.0% ( 0 )
- \* Connectors :
  - Condition 3.3% ( 15 )
  - Cause 44.1% ( 201 )

The right pane, titled 'Extract', shows a text snippet with several words highlighted in red and blue. The red highlights correspond to the 'Factive' category, and the blue highlights correspond to the 'Cause' category. The text snippet is: "ce se mai poate găsi de ars#hârțile încurcă#am dat peste o scrisoare care mi-a deșteptat amintirea unei întâmplări ciudate, așa de ciudată că, de când s-a petrecut, m-aș simți cuprins de îndoială, aș crede că într-adevăr am visat numai, sau că am citit-o ori auzit-o demult. Era în 1907. Fusesem greu bolnav în București și mă întorceam la Berlin acasă. Însănătoșirea mea se făcea cu anevoință cerând îngrijiri mari. să mă feresc până și de cele mai ușoare zguduiri sufletești. Bietul doctor! zâmbind și i-am spus să fie pe pace. După un surghiun de doi ani revedeam Berlinul. nici împrejurări foarte triste nu m-au împiedicat să-l revăd cu plăcere. L-am regăsit cum îl lăsasem: tot numai flori. Așa frumos chiar ca în acel început de iunie numi părușe totuși niciodată."

A *Word category* is deemed significant when its occurrence frequency is far above average. These results are constructed by comparing the statistics of the analyzed discourse (regarding the category distribution) with specific in-built tables.

When you select a *category*, the propositions concerned appear in the main window

### Displaying all word categories

Click on the *[All word categories]* line of the Result frame, or on the *[Show][All word categories]* menu.

The result obtained with this function is similar to that of the previous one, except that all *Word categories* are displayed, even those that are less frequent or not significant.

For further details about these *Word categories*, see chapter 4: "Introduction to text analysis".

To view the References found with these *categories*, we recommend that you use a *Star graph* (see *Graph properties* below).

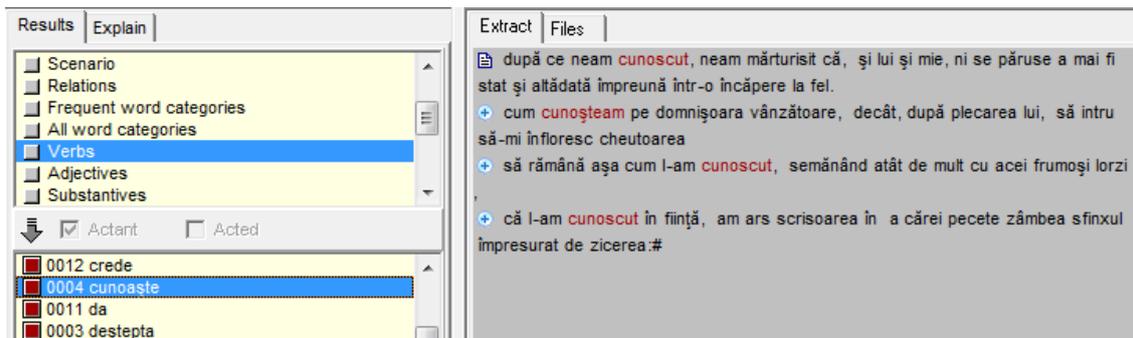
### Lists of verbs and adjectives

Click on the [Verbs] or [Adjectives] lines of the Result frame, or use the [Show] [Verbs]/[Adjectives] menu.

These two functions show, by lessening frequency, verbs and adjectives met in the analyzed text. Every line is preceded by a meter indicating the number of words (occurrences) found.

These words are reduced to their canonical shape (lemma):

- Verbs are reduced to the infinitive;
- Adjectives are presented in masculine singular.



You can add directly these verbs and adjectives in the Scenario Tool. These verbs and adjectives can be useful to complete certain analyses (in particular if you analyze a series of very short messages, where the least information must be got back), to define and to test your own "qualitative" classifications, or to make linguistic studies. To print these results, use the menu [File] [Print] and mark options [Verbs] or [Adjectives].

### Episodes and Bundles

Click on the [*nn Episode(s) Detected*] line of the [*Text Style*] in the Result frame, or use the [*Show*][*Episodes*] command.

This function enables you to study the chronology of a discourse. It is based on two notions, *Bundles* and *Episodes*:

-a **Bundle** groups together word occurrences (belonging to an *Equivalent class* or to a *Word category*) that tend to appear in a remarkable density within a limited portion of the text (somewhere near the beginning, the middle or the end, but not on a regular pattern throughout the whole text).

-an **Episode** corresponds to a part of the text in which a number of *Bundles* have been formed and completed. These are large blocks of arguments, quite representative of the structure of the discourse studied.

*Episodes* are displayed one after the other and numbered according to their occurrence order.

Inside each *Episode*, *Bundles* are sorted out according to their address (words position average) and prefixed by the occurrence frequency of the words that comprise each *Bundle*.

For example, the [*Remember.txt*] text begins with a first *Episode* including short *Bundles* about the “seară”, “muzeu”, “tânăr” References, but also *connectors*, *modalities*... Then comes a second *Episode* including a rather long *Bundle* (beginning in the first *Episode*) about “față” (containing 4 words), etc. The text ends with *Bundles* about “istorie”. This analysis brings out the construction of the discourse by the narrator, who starts his narrative (in this case a short story) by talking about “seară”, “muzeu”, “tânăr” before speaking of “lume”, “neam” and “inel”, and then ending his discourse by talking of “noapte”, “canal”, “Aubrey de Vere”, “istorie”, etc.

**Important note:** different *Bundles* about an identical Reference may appear in various parts of the text. To visualize these results, we recommend that you use an *Episodes graph* (see below).

It is possible to define the construction level of the *Bundles* (using *References fields*, *References*, or *Scenarios* as parameters) by changing the *construction base for the Relations* (see *Analysis options* below). Using *Reference fields* instead of *References* generally results in reducing the number of *Episodes* detected inside the text (in that case, the software uses a smaller amount of the generic concepts that have a high occurrence frequency in the text, which results in grouping together some utterances and reducing the accuracy of the analysis).

### Analysis options

Use the `[Tools][Analysis option]` command:

The screenshot shows a dialog box titled 'Options' with a 'Conversions' tab. It contains three main sections: 1. 'Class detection threshold:' with two radio buttons: 'Number of words' (selected) and 'Pertinence factor'. Next to 'Number of words' is a text box containing '3' and a spinner control. Next to 'Pertinence factor' is a text box containing '15' and a spinner control. Below this is an unchecked checkbox labeled 'Use ontologies on all words categories'. 2. 'Build relations on:' with four radio buttons: 'Reference fields 1', 'Reference fields 2', 'References' (selected), and 'Scenario'. 3. 'Quantity of characteristic parts of text:' with a slider control between 'More' and 'Less'.

With this dialog, you can both act on the analysis engine of the software and change some display options.

The **Class detection threshold** enables you to define the significance level of the *Equivalent classes*:

-When this threshold is based on a minimum **number of words**, all the *Equivalent classes* whose occurrence frequency is below this threshold will be ignored (i.e. they will not be displayed).

-When this threshold is based on a **pertinence factor**, all the *classes* whose *pertinence factor* is below this threshold will be ignored.

The *pertinence factor* is calculated in ten thousands of number of words. For instance, a *pertinence factor* of 10 corresponds to a minimum occurrence frequency of 3 words for a 3,000-word text.

You can change the *Class detection threshold* if you want the software to process only the most frequent *classes* or, conversely, to take into account the less frequent *classes*.

Important note: the higher you raise the thresholds, the more information you lose. And vice versa: when you lower the thresholds, you increase the amount of information taken into account by the *Equivalent classes*.

The analysis options dialog also enables you to change the **Construction base for the Relations** (using the `[Build Relations on]` box), i.e. the *Equivalent classes* level needed to build, display and print *Relations*, *Episodes* and *Bundles*.

For example, if you use *Scenarios*, it is possible to build the *Relations* from the content of the *current Scenario* (see Chapter 2 - *Semantic Scenarios*).

It is also possible to modify the contraction rate by using the cursor, which enables you to adjust the **[Quantity of characteristic parts of text]** to be displayed.

Since it is uncertain whether the software will or will not be able to detect, with the greatest accuracy, the essential propositions of a corpus, we recommend that you start with a rather low threshold (i.e. that you display many *Characteristic parts of text*), and then raise it gradually until you strike a balance between the amount of displayed propositions and the pertinence of the result.

The **[Use the Scenario on all word categories]** option enables you to require the software to convert all the words entered in the *semantic Scenario* into substantives, or References (see related chapter).

When this box is checked, it is possible to enter (as *items*) and display in the *Scenario* words that do not belong to the substantive category. For example, you can group together adverbs according to various themes. You can also use this option to force the software to consider as substantives words that are not filed under this particular category (for instance, if you analyze a text presenting two characters, “DI. Negru” and “DI. Roșu”, you can enter “Negru” and “Roșu” in your *Scenario* so that they will be counted as References).

To validate your choice, press the **[Accept]** button. Otherwise, if you do not wish to modify your analysis options, press the **[Cancel]** button.

**Caution:** when you check the **[Use the Scenario on all word categories]** box, all words subsequently entered in the Scenario will be converted into substantives (References) at the end of the automatic analysis of the text. This has several consequences on the operation of the software:

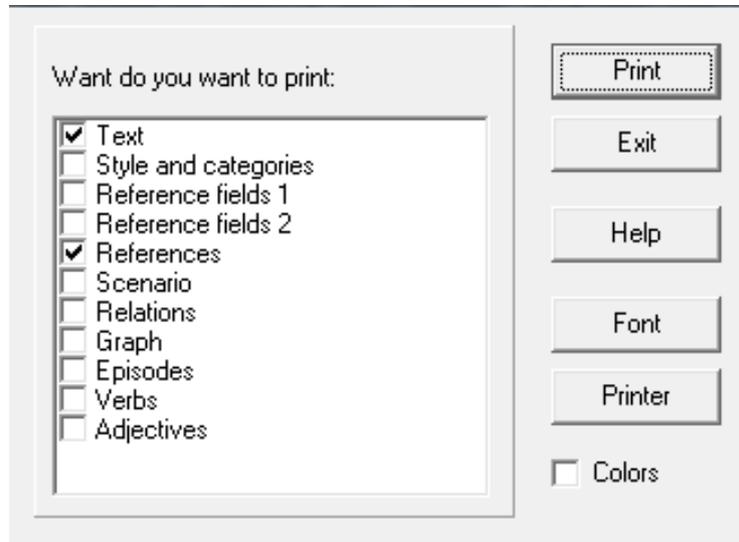
1 -The *Scenario* has priority, then, over the other classifications: all the non-substantives entered (as *items*) in the *Scenario* will be removed from their original categories. For example, if you enter the *item* “acum” in the *Scenario*, the corresponding adverb of place will no longer be displayed in the *Modality* category (in which it will nevertheless be counted).

2 -The *Scenario* takes precedence, then, over the syntactic analysis of the text; if an ambiguous word (simultaneously used in various grammatical categories) appears in a text, and if this word is placed in the *Scenario*, then all its various forms will be counted in the *Scenario*. For example, if you have entered the word “sare” in a *Scenario*, later used to analyze a text containing the following sentences: “Broasca **sare** în apă.” and “A fost primit cu pâine cu **sare**.”, then the two occurrences of “sare” (the verb and the common noun) will be counted in the *Scenario*.

3 -The above observations only apply when you enter words in the *Scenario*; if your *Scenario* is built from *Equivalent classes* (or verbs / adjectives / substantives lemmas), these will pose no classification problem (the *Equivalent classes* and lemmas have lexical and semantic ambiguity solving).

## Printing the results

Use the *[File][Print]* command:



To print results, check the related boxes, then press *[Print]*; otherwise press *[Exit]*.

Printing the report of the *Equivalent classes* and of the *Scenarios* discloses a *utilization rate*, expressed in a percentage that corresponds to the number of words contained in each *class* divided by the total number of words contained in the text.

Printing the report of the *Relations* discloses an additional item of information, not displayed in the *[Relations]* of the *results dialog*: the *connection rate*. This *rate* is obtained by dividing the number of observed *Relations* by the highest number of possible *Relations*. A *connection rate* of 100% shows that one of the two terms of the *Relation* is always presented with the other. A *connection rate* that is close to zero shows that the two terms are almost never presented together.

In the printing options, various buttons enable you to change the configuration of the printer and to select the font you wish to use for the printing.

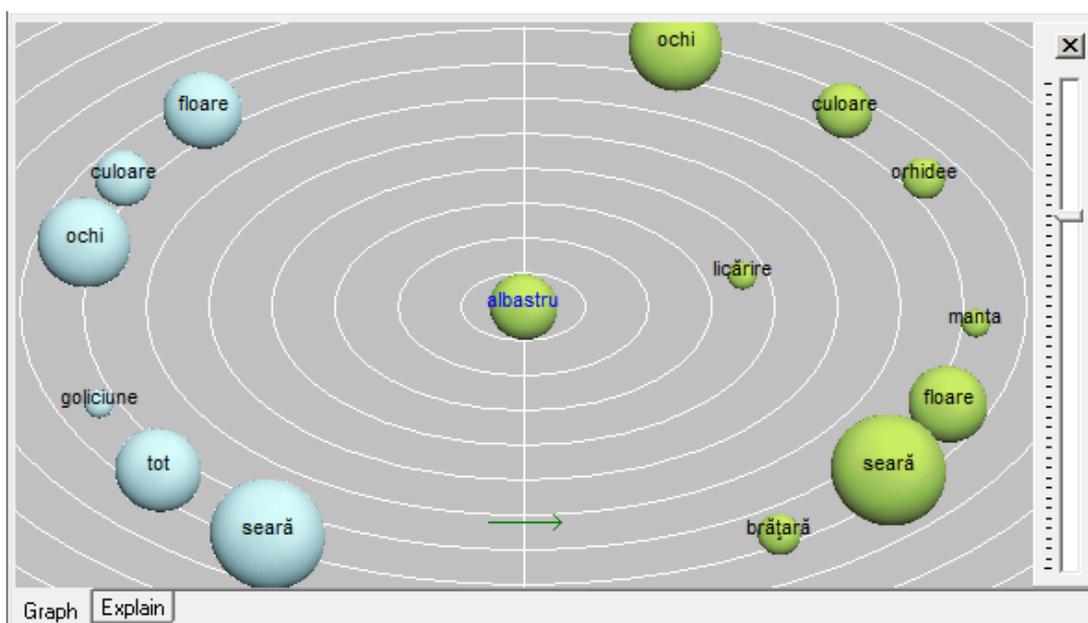
The *[Color]* box enables you to print in color, if you have a color printer.

## Equivalent classes and Relations graphs

For the graph display, you can choose between the following modes: *[Area]*, *[Star]*, *[Distribution]* or *[Episodes]*. Use the context menu that appears when clicking on an object of the graph with the right button of the mouse.

The first two graphs (*area* and *star*) show the *Relations* between the *Equivalent classes*. They are oriented: the *classes* on the left of the central *class* (in red) are its predecessors, those on the right its successors.

The first graph displays the *classes* in the form of **Areas**:



Each *Equivalent class* appears as a sphere, whose surface is proportional to the number of words it contains.

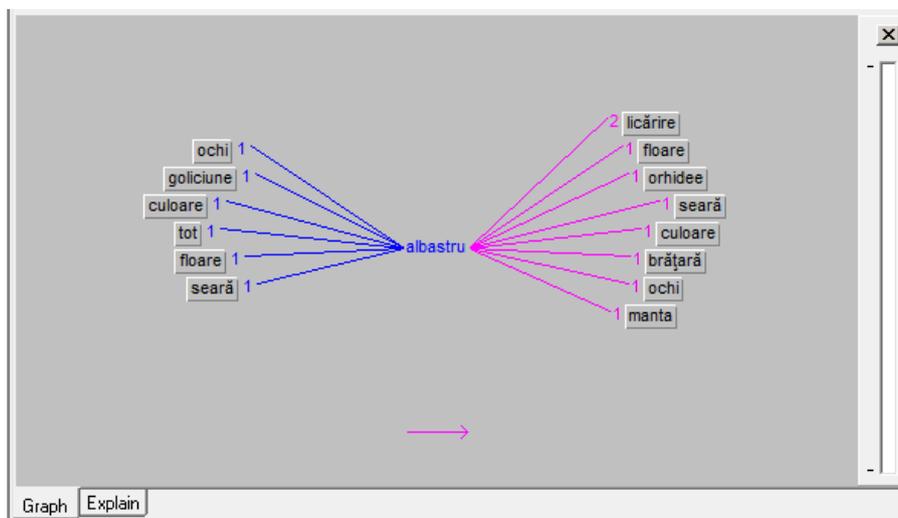
The distance between the central *class* and the other *classes* is proportional to the number of *Relations* connecting them: in other words, when two *classes* are close together, they share many *Relations*, and when they are far from one another, they share few *Relations*.

If we take up the metaphor about the planetary system, we can say that there is a central body (*Equivalent class*), around which planets (*Equivalent classes* having *Relations* with the central *class*) of various sizes (according to the number of word occurrences they contain) revolve more or less closely (are more or less frequently used together).

### Notes:

- 1 - The overlapping of two spheres has no particular significance.
- 2 - *Word categories* cannot be displayed in the area mode.

The **Star graph** displays the *Relations* between the *Equivalent classes*, or between a *Word category* and an *Equivalent class*:

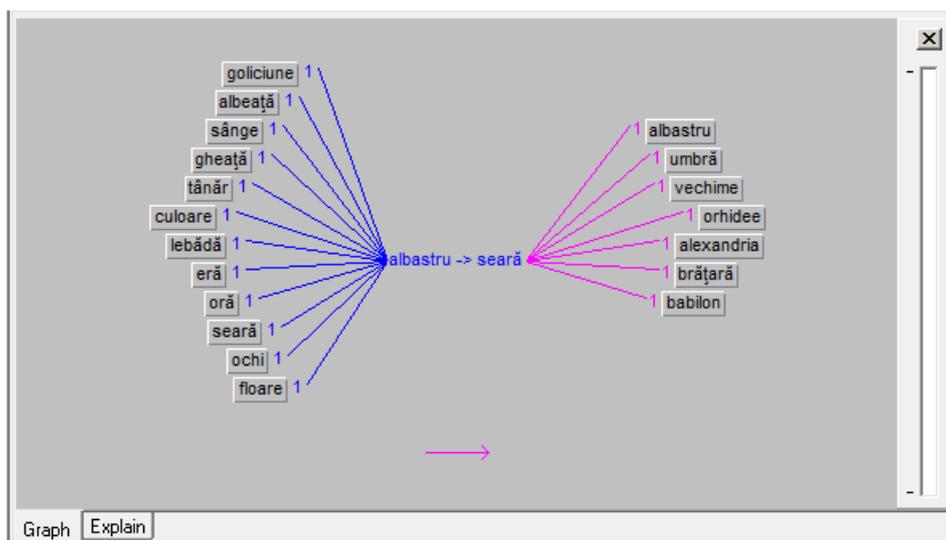


In the above example, the “albastru” *class* (central point) is preceded by the “ochi”, “goliciune”, etc. *classes*. On the right are many successors to “albastru”: “licărire”, “floare”, “orhidee”, “brățară”, etc.

The figures shown on the graph give the number of *Relations* (co-occurrence frequency) existing between the various *Equivalent classes*.

You can follow the *Relations* shown on the graph by clicking directly on the *classes* you wish to study. This very powerful function enables you to move through a text while viewing its micro-worlds and analyzing the connections between its various actors.

When you use the [*Relations*], the graph reveals two central *classes*:



In this example, we can see that there are only two *Relation* leading from the “albastru” *class* to the “seară” *class*, whereas there is one *Relations* leading from the “seară” *class* to the “umbră” and “vechime” *classes*.

The **graph of the actors** represents the concentration of relations between the main actors (actants/acted) in the whole of the text. It can be used to make a visual comparison of the "weight" of the Relations between the principal references (or between the groups in the Scenario).

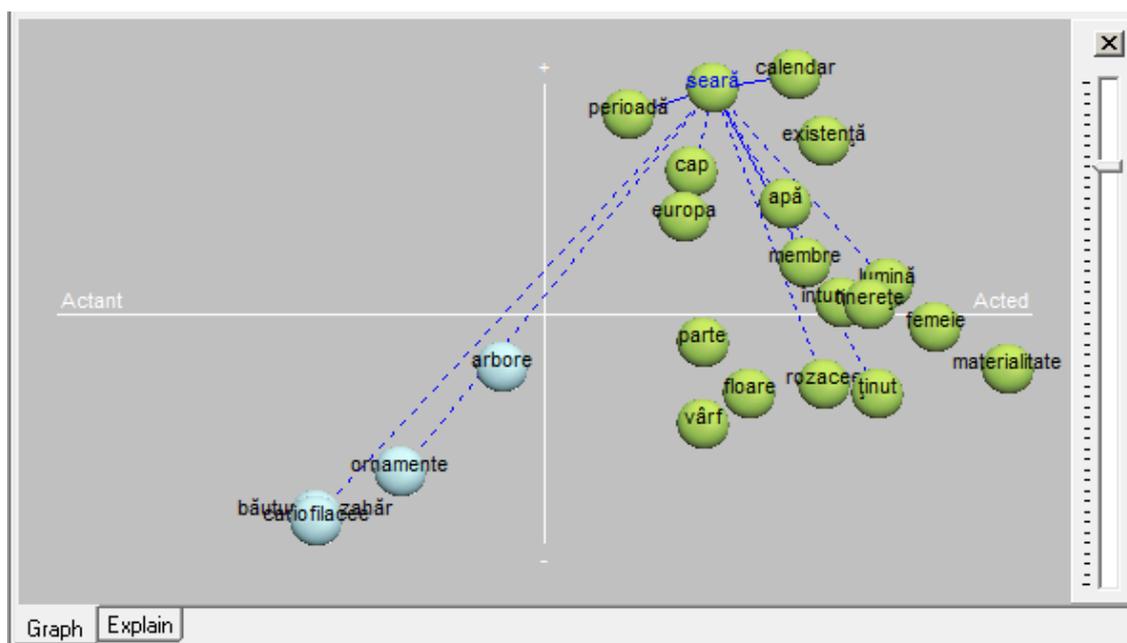
The references are displayed along two axes:

- the X axis (horizontal) shows the actant/acted ratio (from left to right),
- the Y axis (vertical) shows the concentration of relations for each reference (strong at the top of the graph, weak at the bottom).

The concentration of relations is calculated, for each reference, by dividing the total number of relations by the number of different relations.

The lines show the relations between the reference selected and the other references displayed. A dotted line shows an infrequent relation. A solid line indicates a frequent relation.

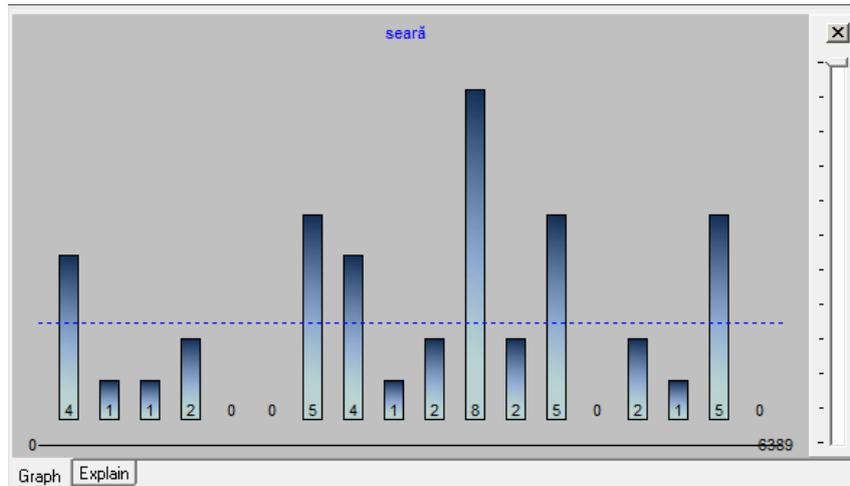
The example below is taken from the analysis of the [*Remember.txt*]. One of the most frequent reference ("seară") has been selected.



In this example, the software has positioned the principal concepts ("calendar", "perioadă", "apă", etc.) at the left (*actants* with a high concentration of relations), while the secondary concepts are positioned either at bottom left (*actants* with a low concentration of relations) or on the right (*acted*). Most of the other references are acted (on the right).

The frequency of relations and the number of different relations are indicators of the centre of interest of the text analyzed. If the author, or authors, of the text have associated a reference with many other references, it may be deduced that this reference is very important, or at any rate more important than others. On another level, this new graphical representation shows an overall graph, not depending on the selected reference alone (as is the case with the star and area graphs).

The **distribution graph** enables you to display a histogram showing the distribution of an *Equivalent class*, of a *Relation* (i.e. between two *Equivalent classes*), or of a *Word category*.



In the above example, we can see that the “seară” *Equivalent class* appears rather at the end (right side) than at the beginning (left side) of the text.

This graph is obtained by splitting the text into several sectors containing the same number of words, and by calculating the occurrence frequency of the selected *Equivalent class* within each sector. The histogram bars present each sector in chronological order, from the left (beginning of the text) to the right (end of the text). The dotted line indicates the average size of the histogram bars.

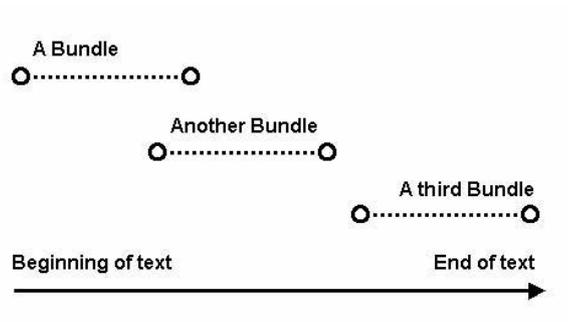
When clicking on a given bar, the display will automatically position itself on the propositions appearing from this point of the text onwards. To return to the standard display, make a second click on the same bar. When pointing at a given part of the histogram, the message line displays the number of words contained by the selected sector, as well as its place (beginning and end of the sector, calculated in number of words from the beginning of text).

The number of histogram bars and the sum of words contained in each bar are automatically determined by the software according to the total of words in the text, and to the size of the main window.

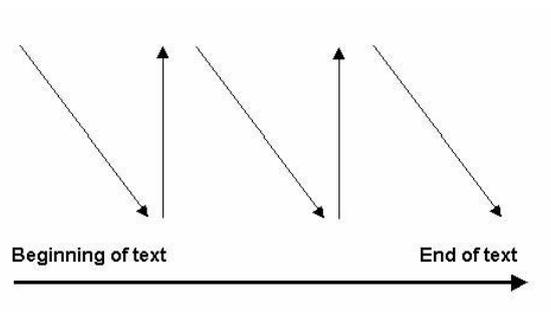
When the distribution graph concerns a *Relation*, the histogram gives the accumulated occurrence frequencies of the *classes* contained in the *Relation*.

## Bundles and Episodes graphs

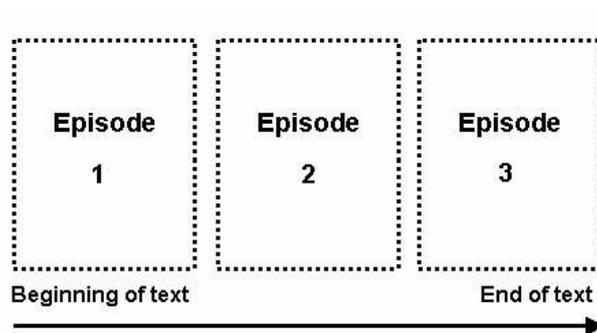
The *Bundles* graph is displayed when viewing an *Episode* and using a *distribution* graph (see *Graph properties* below). Each *Bundle* is displayed in the form of a horizontal dotted line indicating its extent (length) and its place in relation to the beginning of the text. As with the *distribution* graph, the chronological order is represented on the horizontal axis, from the beginning (on the left) to the end (on the right) of the text:



The *Bundles* are displayed from the top downwards and from left to right, according to their occurrence order in the text. Once at the bottom of the screen, the software makes a zigzag scan, in order to display as much information as possible:



The *Episodes* appear on the same graph. They are displayed from left to right, in chronological order, and in the form of large dotted frames:



Only the *Bundles* contained in the selected *Episode* (shown by a hatched frame) are displayed in color.



### Copying the graph

When a graph is displayed, you can use the *[Edit][Copy][Graph]* command to transfer the graph to the *Windows clipboard*. To retrieve this graph and include it in a report (for example), use the *[Edit][Paste]* command of your word processor.

### Copying the selected text

When the display applies to a subset of the text (for example when viewing an *Equivalent class*), you can use the *[Edit][Copy][Selected text]* command to transfer the selected propositions to the *Windows clipboard*. To retrieve these propositions, use the *[Edit][Paste]* command of your word processor.

### Graph properties

Use the *[Edit][Graph]* command of the main menu, or the *[Properties]* command of the menu that appears when clicking on an object of the graph with the right button of the mouse.

This dialog enables you to define the parameters of the graphs. You can:

- change the **[Type]** of graph to be displayed: *[Star]*, *[Area]*, *[Distribution]*, *[Episodes]*;
- modify the **[Maximum number of objects in the Area graph]**;
- show the **[Bundles on all word categories]** (including Connectors, Modalities, etc.), or if unchecked, on *Equivalent classes* only, which makes it easier to read when studying a very long text.

All modifications become immediately effective. If you wish to store them permanently, click on *[OK]*; otherwise, click on *[Cancel]*.

### Changing the colors

Use the *[Edit][Colors]* command of the main menu, or the *[Colors]* command of the menu that appears when clicking on the main window with the right button of the mouse.

This dialog enables you to change the colors:

- of each *Word category* that is displayed,
- of the borders framing the *classes* displayed **before** and **after** the central *class* of the graph,
- of the **background** of the areas (drawn on the graph),

## Tropes – Reference Manual

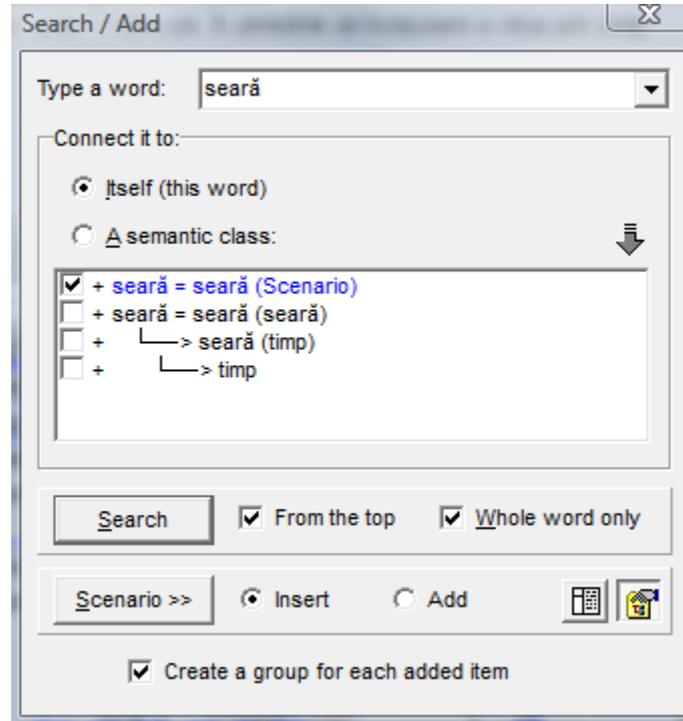
-and to choose between **Selective** (only the displayed elements are colored) or **Syntactic** (all word categories are colored) coloring.

It is also possible to choose the texture of the areas, which can be either **opaque** (full) or **striped** (transparent).

The *[Restore]* button cancels all color changes made after the initial installation of the software.

## Finding a word

Use the *[Edit][Search]* command or the *[Search]* command of the context menu that appears when clicking on a word of the text with the right button of the mouse:



To use this dialog, type a word, then press the *[Enter]* key or the *[Search]* button.

If the software finds the word sought, the propositions containing this word will be displayed in the main window; otherwise, you will hear a beep.

Depending on the chosen option, the search will be carried out either within the *text* (option **[Itself]**), or within the **[Semantic classes]** (or *Groups of the Scenario*). The second case requires that the corresponding *Equivalent class* exists in the text, and (incidentally) that the last modifications of the *Scenario* have been applied.

If the *[Start from the top]* box is checked, the search will be carried out from the beginning of the text; otherwise, it will start from the last searched word.

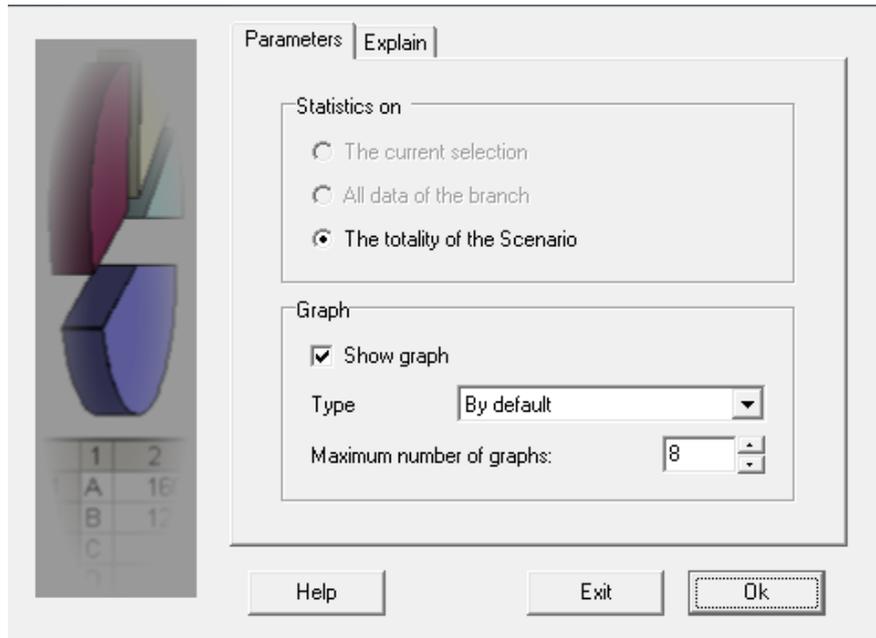
When the *Scenario* tool is opened, it is possible to add (create a new group) or insert (into an existing group) a word or an *Equivalent class* directly in the *Scenario* by pressing the *[Add][Insert]* button and checkbox at the bottom of this dialog.

The *semantic Groups* of the *Scenario* appear in the *Equivalent classes* list, and are displayed in blue (note: for this function to work, you must first save the *Scenario* on the studied text).

### Report writer

The report writer gives you the possibility to automatically build reports (containing statistical tables and graphs) in Microsoft Excel<sup>®</sup> by using the results obtained with the *Scenario*.

To use it, you must select a branch of the *Scenario*, then use the *[Tool][Report]* menu:



You can make statistical tables on:

- the selected element** (i.e. the semantic groups of the same level),
- all the data of the branch** that contains the selected element,
- the totality of the Scenario** (which can be very voluminous).

You can choose to **display graphs** for each generated statistical table, then choose their **type** (histogram, etc.) and control the **maximum number of graphs displayed** (bearing in mind that a huge number of graphs is sometimes difficult to utilize, and that your spreadsheet cannot deal with an unlimited amount of graphs).

This module requires a recent version of Microsoft Excel<sup>®</sup>, and is accessible both in Tropes (where the results are expressed in number of word occurrences) and in Zoom (where the results are expressed in number of documents).

Note that modifying the default type of graph in Microsoft Excel<sup>®</sup> enables you to have at your disposal numerous options to control the layout of the report.

# Tropes<sup>®</sup>

## CHAPTER 2

# Semantic Scenarios - Ontologies

### Introduction

*Scenarios* are designed to enrich and filter *Equivalent classes* according to an analysis strategy. With them, you can:

- define your own personalized classifications (ontologies),
- modify or restructure the software's dictionaries,
- replace a thesaurus and personalize your information retrieval functions,
- define an analysis grid for automatic generation of an analysis report (see *Report writer*, in this manual) or statistics when indexing your documents (see *Using a Scenario*, in the Zoom manual).

Why is it necessary to use *Scenarios* when it is tempting simply to make modifications directly to the Tropes dictionaries? Because you need to use classifications adapted to your analysis objectives, and these depend on what you want to do with your texts. For example, it is perfectly correct to classify “pepene roșu”, “castravete”, “pepene galben”, “dovlecel” in the “curcubitacee” family, if you want to conduct botanical analyses. If it is the domain of foodstuffs that interests you, on the other hand, it would be more appropriate to class “pepene galben” and “pepene roșu” as fruits, and “castravete”, “dovlecel” among the vegetables. In other contexts, you may wish to differentiate between large companies and SMEs, between toxic substances and medicines, or between political personalities and their opponents. Or indeed, between what interests you and what you consider of secondary importance.

It is essential to use a *Scenario* if a text is to be analyzed correctly. In most cases, you will need to rework classifications, correct any anomalies and personalize your analyses.

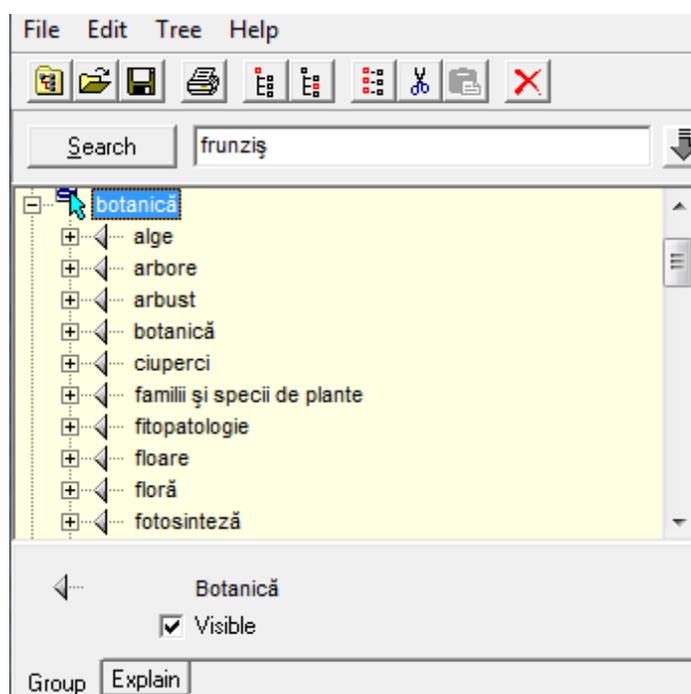
Furthermore, the dictionaries of Tropes are constantly evolving, and you can download new versions from the publisher's website. You then conserve your personalized *Scenarios*, even when you update your software. When you install a new version, analysis becomes more refined, and you have the benefit of new classifications, but you do not lose your work (which is preserved in your *Scenarios*).

Tropes is supplied with several predefined *Scenarios* which correspond to different approaches to the classification of your documents:

- 1 – comprehensive (with *Scenario Concepts*, which groups references under broad themes, similarly to a mass-market general thesaurus);
- 2 – detailed (with *Scenario Concepts detailed*, which groups references under a larger number of themes, on the lines of an encyclopedia);
- 3 – highly specialized (with the other *Scenarios*, or those available on request).

## Creating semantic Scenarios

To create a *Scenario*, use the [*Scenario*] command in the [*Tools*] menu.



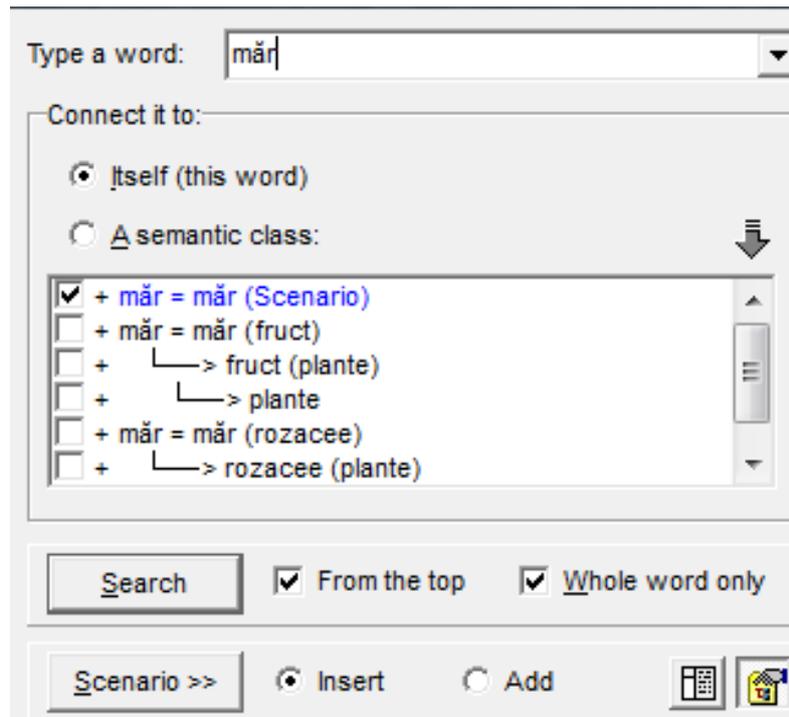
Tropes is supplied with a number of default *Scenarios*, rich in classifications, which can serve as a starting point for the rapid creation of a new classification plan. We advise you to be careful to keep the *Scenarios* supplied with the software, and to save them under a new name if you want to modify them.

A *Scenario* consists of a number of *Semantic Groups*, i.e. groups of words and/or *Equivalent classes*, which can be arranged in a hierarchy of nine ranks. Most combinations can be made by means of the mouse, either in the *Scenario* tool or interactively with Tropes' main window. Just create a group and put something into it.

When viewing a *Reference* in the Result frame, you can add it directly to the *Scenario* by dragging it with the mouse and dropping it on the *Scenario* tool. In Tropes Version 8 and later, you can also add the verb and adjectives lemmas (or their lemmas in languages other than Romanian) directly in the *Scenario*. It is also possible to grab a word in Tropes [*Text*] frame and drag it onto the *Scenario*. Use the [*Ctrl*] and [*Shift*] keys to control the addition or the insertion during this operation.

The Result popup menu (right click with the mouse) also contains two functions linked to the *Scenario*. The command [*Add to the Scenario*] creates an entry having the same group name as the selected *Equivalent class*. The [*Insert in the Scenario*] command creates an entry by using the group name currently displayed in the *Scenario*.

Lastly, you can create a new entry in the *Scenario* manually from the [*Search/Add*] dialog box. Just type a word, select a group and press the [*Scenario*] button. You can choose to add a word or an *Equivalent class* to the *Scenario*. It is possible to switch from the [*Insert*] function to the [*Add*] function (and vice versa) by checking a box located at the bottom of the [*Search/Add*] dialog.



All the parts of the Tropes Graphic Interface contain popup menus (right-click with the mouse) which enable easy addition/insertion of words or *Equivalent classes* in the *Scenario*:

The *[Save]* button in the *Scenario tool* enable you (without quitting this dialog) to take into account the modifications to the *Scenario*, save them and update the display.

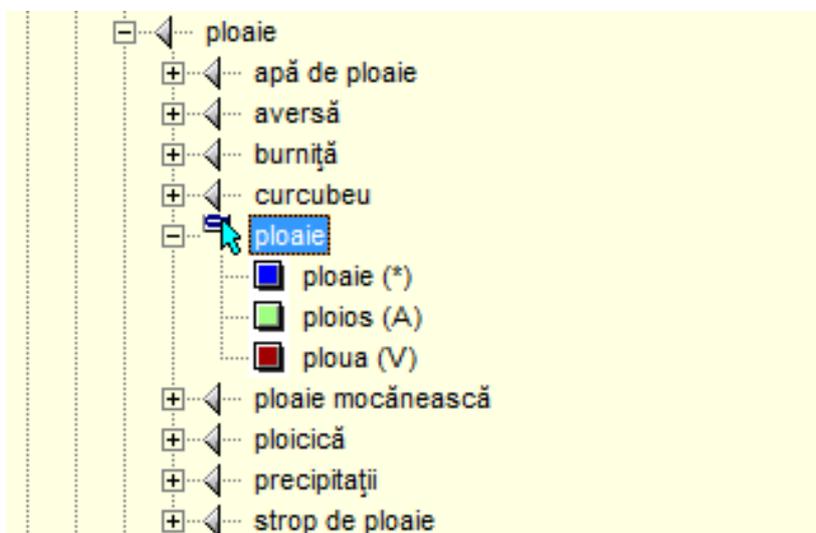
Most of the navigation functions in the *Scenario tool* are similar to those of Windows Explorer. For example, to rename a semantic group, just select it for a few seconds.

## Connecting an Equivalent class to a Semantic Group

When a *Semantic Group* refers to an *Equivalent class*, the *Scenario tool* displays the corresponding line and, between parentheses, a note containing:

- a star (\*), when the *item* includes a *Reference*,
- a number (1 or 2), when the *item* includes a *Reference field* (1 or 2),
- a letter (V), when the *item* includes a verb lemma,
- a letter (A), when the *item* includes an adjective (or its lemma).

In the example below, the equivalent class "ploaie" (which is one of the References), has been grouped together with the verb "ploaie" and the adjective "ploios".



Semantic groups can be designated [*Visible*] or [*Invisible*]. If the box is not checked (group invisible), none of the words or *equivalent classes* under the invisible group will be displayed in the results. This makes it possible to hide a branch temporarily without being forced to delete it.

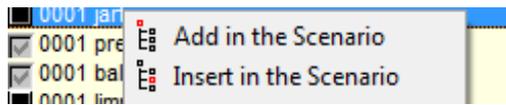
The Scenario has been designed to solve ambiguities manually, which means that when you add a word to a Semantic group, Tropes will cease solving ambiguities. For example, if you add the word "păr" to the [Botanică] group in the Scenario concept, you inhibit the ambiguity-solving mechanism ("păr" will then always be associated with [Botanică], whatever the text under analysis). To obtain ambiguity solving, use *Equivalent classes*. By virtue of these principles, you must avoid placing a semantic class or a word in two different groups (if you do this, the anomaly will be pointed out when you save the Scenario, and the second entry will not be taken into account).

### Notes:

- If you enter a noun, you must add it from *References* or *Substantives*. In this case, you add all its declination, in Romanian. In contrast, if you will add an

## Tropes – Reference Manual

unrecognized noun, you must decline it, if you will find all its forms.



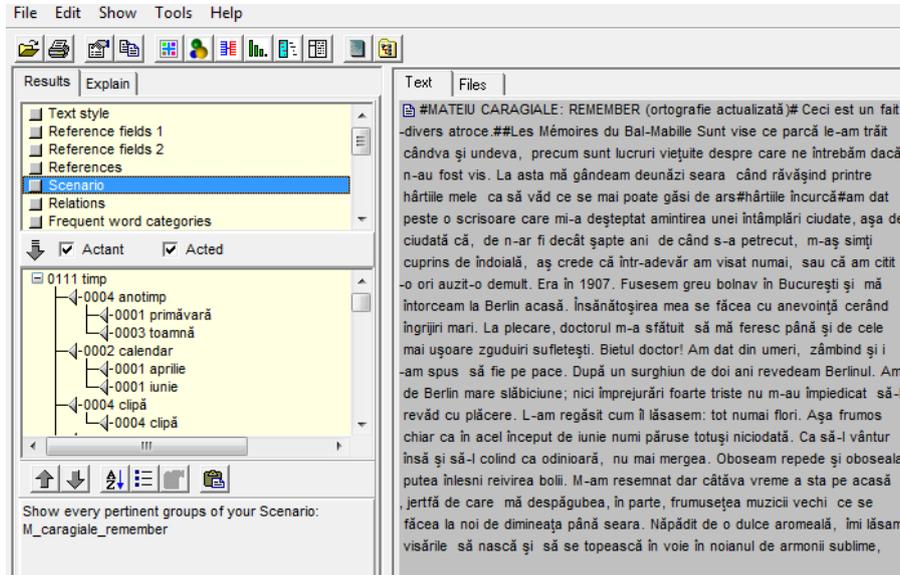
-If you enter a compound – unrecognized as such initially – in a *Scenario*, it will be added to the software’s dictionary. This means that the words forming this compound will not be taken into account independently any longer in the *Equivalent classes*. For example, if you enter “Doamna Steluța” in the *Scenario*, the software will no longer identify “steluța” as an “astru”; on the other hand, the “Doamna Steluța” compound will appear in the *References*.

-The lower level equivalentents have priority over the higher level equivalentents, i.e. if you create a *semantic group* containing a single word, this word will not be counted in another Semantic Group including an Equivalent class that contains this same word. For example, if you assign the word “ploicică” to the “ploaie” Group, this word will be excluded from the “ploaie” Reference field, even if this field is assigned to another Semantic Group.

If you wish to create a *Scenario* that is relevant to a series of texts, we recommend that you collect, inside a single file, a sample of texts representative of what you intend to analyze, and then build your *Scenario* from this sample.

## Applying the Scenario to the text

Once you have created the *Semantic groups* of your *Scenario*, and as soon as it has been saved, you can use the *[Scenario]* command in the Tropes main window.

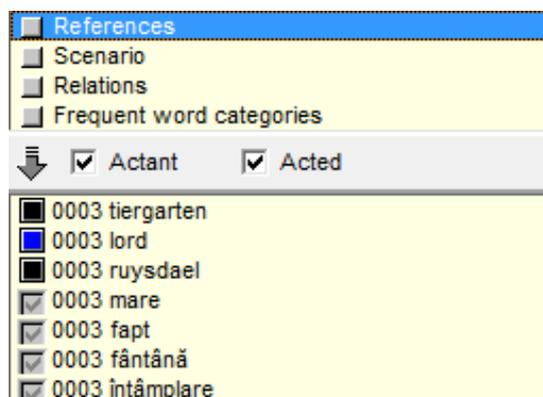


At this stage, you can use *Semantic groups* in the same way as *Equivalent classes*: you can visualize the various types of graphs, the *groups* placed in the position of *Actants* or *Acted*, etc.

When you display these results, you can use the toolbar at the bottom of the Tropes result window and a popup menu (right-click with your mouse) to move around in the arborescence by displaying or hiding the branches (commands *[Reduce all]*/*[Expand all]*) and to change the sorting order of the *Scenario*:



The semantic classes in a *Scenario* are indicated by a checkbox in the result lists. By this means, it is possible to see at once if a class has been taken into account in your classification. In the example below, it will be seen that "fapt" and "fântână" are classified in the *Scenario*, while "Tiergarten" and "Ruysdael" are not:



### The methodology of *Scenario* design

You may use a number of approaches when designing your *Scenario*. For example, you may use an existing *Scenario* (such as those supplied with the software), supplement it and refine it according to your working hypotheses. You can also start off with an empty *Scenario* which you will build up gradually on the basis of a text.

We suggest that you adopt the following methodology, which gives good results quickly:

1. open the *Scenario tool* ([Tools][Scenario] menu in the main window) and load one of the existing Scenarios ([File][Open] menu in the *Scenario tool*), which you then save under another name ([File][Save as] menu in the *Scenario tool*);
2. analyze a collection of documents representative of the subject that you want to deal with;
3. display the References;
4. supplement the *Scenario* by adding all the words that are not classified (that is to say, are not checked in the list of References) and which are relevant to the subject of your analysis;
5. display the Relations, or the expressions proposed by the terminology extractor, and use them to classify the relevant compounds in the *Scenario* (for example, if you are studying the Mateiu Caragiale files in the example texts, you should group "Aubrey " together with "Vere" to carry out a correct analysis of the text);
6. you may wish to study the lists of verbs and adjectives, some of which can strengthen the *Scenario*;
7. apply your *Scenario* to the text ([File][Save] menu in the *Scenario tool*);
8. move on to the *Scenario* in the Tropes results window (using the [Show][Scenario] menu in the main window if necessary);
9. examine the result obtained by going down the list of groups in the *Scenario*;
10. correct misclassifications by adding them directly in the *Scenario tool* (for example, if "the White House" is classed among colors, you must move it into another group);
11. remove the narrative branches (which are not useful for your analysis) from your *Scenario* (for example, the days of the week are generally of little interest when analyzing newspaper articles);
12. interpret the result and go back to step 7, if necessary;
13. when you have checked everything, your analysis is done.

## Tropes – Reference Manual

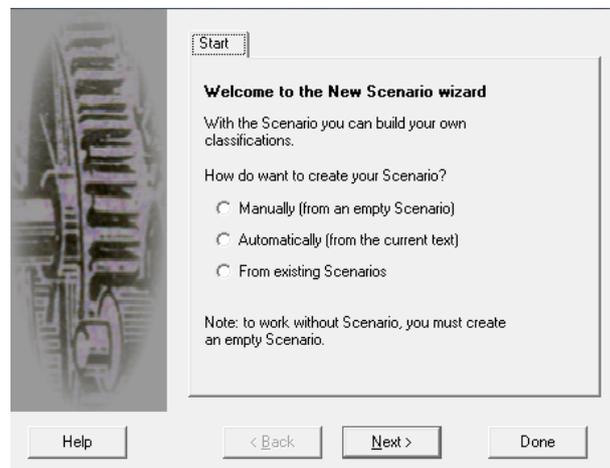
As the *References* are sorted in decreasing order of frequency, you can be sure that you have classified the main references in your texts when you employ this method, even if you do not incorporate all the semantic classes in the *Scenario*. For example, if you halt the construction of the *Scenario* when the *References* reach a frequency of two occurrences (i.e., generally not very relevant), you cannot be accused of making a hash of the analysis. You will simply have focused on the essentials, by halting the analysis at a particular moment.

### Scenario management

The [File] menu of the *Scenario tool* enables you to:

- create** a new Scenario,
- open** an already existing Scenario,
- save** the current Scenario under another name,
- print** the definition of the Scenario,
- find** the folder containing your Scenario files.

For the creation of a new *Scenario* ([New / Wizard] in the [File] menu), an assistant will help you create it automatically from a text, or from an existing *Scenario* (or several *Scenarios*):



If you do not want to use an existing *Scenario*, just create a new, empty *Scenario* and build it up. However, you are strongly recommended not to work without a *Scenario*.

Note that the [Edit][Cancel all changes] command enables you to cancel all the changes made since the *Scenario* was last saved.

#### Important:

- 1 -The names of files containing *Scenarios* must have the suffix [.SCN] (for example: *Remember.scn*).
- 2 -It is essential to store *Scenario* files in the [*Scenario*] sub-folder of the software installation folder (*c:\Program files\Tropes\Tropes...\scenario*), or in the folder that you designated when installing the software, if you opted for multiple-users mode.
- 3 -Save the contents of this folder regularly. Use the [File][Folder] menu in the *Scenario tool* to copy your *Scenario* files.

## Terminology extraction

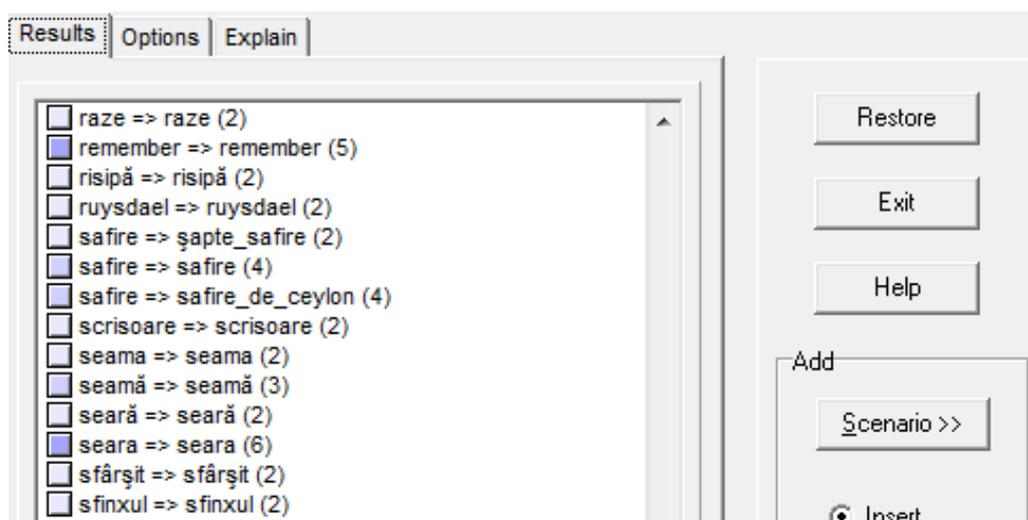
The terminology extractor is capable of automatically identifying most significant expressions and compounds, as well as all nouns not classified in the existing *Scenario*. This tool is useful both for rapidly enriching the *Scenarios* of the software (for example, by grouping together all acronyms and the expressions corresponding to them) and for obtaining a more precise classification (proposing, for example, to "hard-wire" terms causing ambiguity problems, and/or which might generate "noise" in the list of Relations).

The terminology extractor serves a triple purpose:

- 1 – It automatically extracts from the text all compounds (i.e. repeated sequences of terms which contain at least one noun and are linguistically coherent) which might prove interesting for purposes of analysis.
- 2 – It suggests a list of references to complement the *Scenario* and/or draws up a list of everything that has not been classified as yet.
- 3 – It speeds up the construction of the *Scenario*.

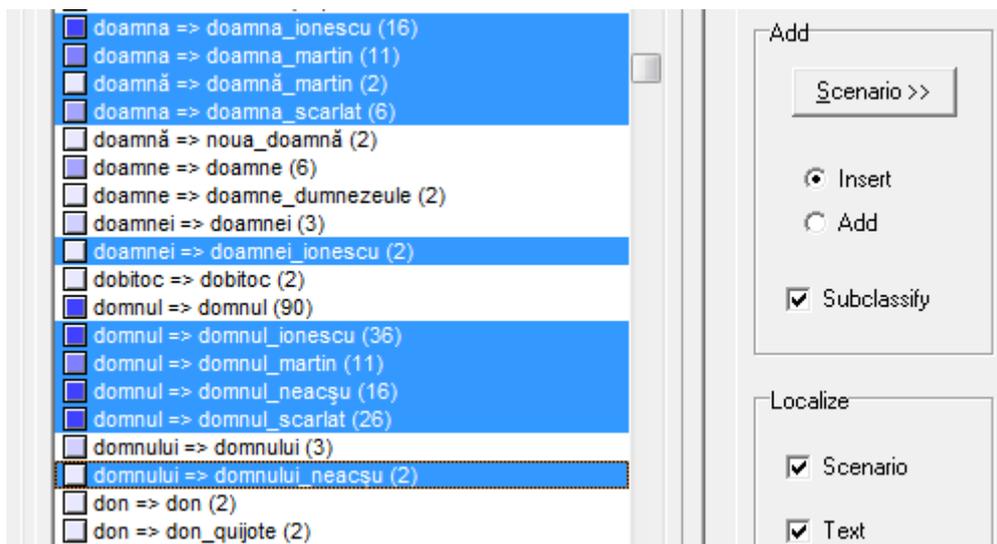
The terminology extractor is semantically linked to the *Scenario*, that is to say that when you select a family of terms, Tropes will automatically try to position the *Scenario* tool on the group that seems most suitable to accommodate the selected expression. The software can also perform an automatic search for a family of expressions in the text. These two semantic linking functions can be deactivated (cf. "Localize" options at the bottom right in the dialog box). The terms extracted are preceded by a color code showing which are the most frequent terms or expressions (dark = frequent, light = infrequent).

The extractor is run from the *[Tools]* menu of Tropes, when you have analyzed a text or (preferably) several texts, on the basis of which you want to build a classification.



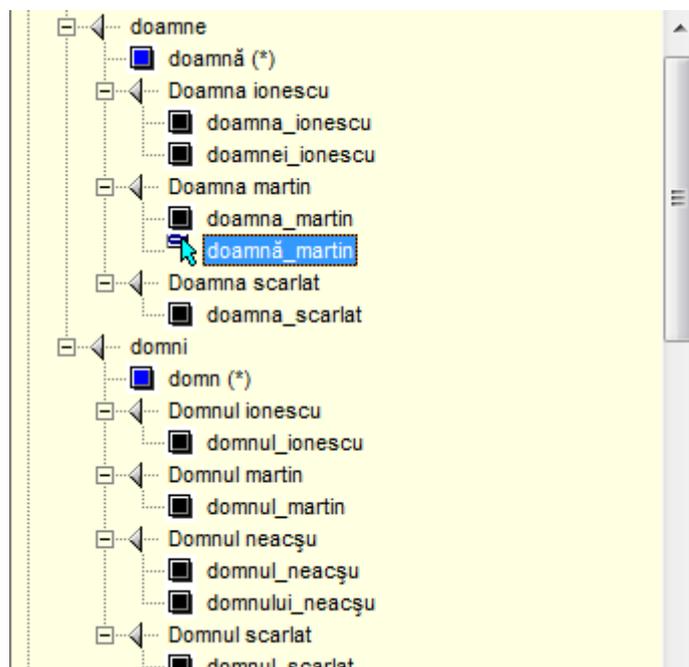
Each term (or expression) extracted is displayed on a separate line, with its membership group. It is preceded by a colored square showing its frequency of occurrence.

To enrich your classification, use the *Scenario* tool and select the terms that interest you in the terminology extractor.



We will then check the *Insert* box, and click on the *Scenario* button (on the right), to transfer them automatically into the *Scenario*.

The software will automatically create all the entries and corresponding group labels in the *Scenario* tool.

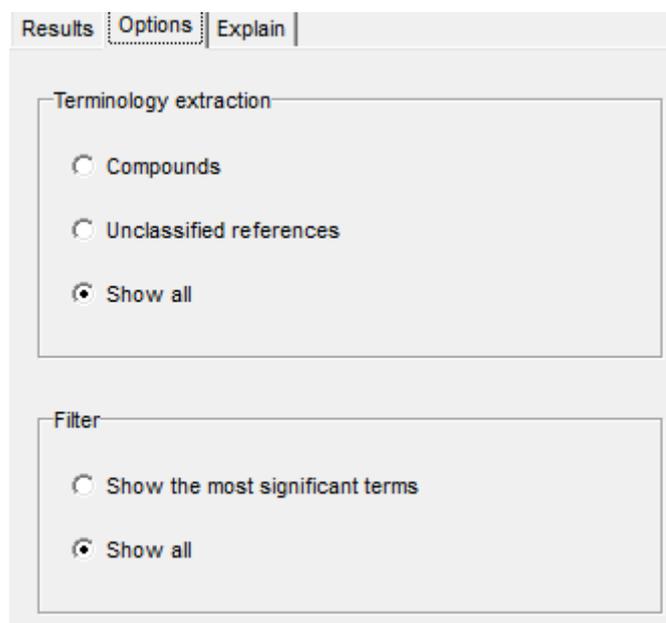


All you have to do then is save your *Scenario* ([Save] command in the [File] menu of the *Scenario* tool) to see the result in the text analyzed.

When you select a term, and provided the *Scenario* tool is open, the software will try automatically to position the *Scenario* on the closest semantic category. This behavior can be inhibited by removing the *Scenario* box from the [Localize] frame on the right in the dialog box.

## Tropes – Reference Manual

Lastly, the terminology extractor is provided with an *[Options]* tab with which you can fine-tune the level of extraction.



The screenshot shows a web interface with three tabs: 'Results', 'Options', and 'Explain'. The 'Options' tab is active. It contains two sections: 'Terminology extraction' and 'Filter'. Each section has three radio button options.

**Terminology extraction**

- Compounds
- Unclassified references
- Show all

**Filter**

- Show the most significant terms
- Show all

Over to you!

# Tropes<sup>®</sup>

## CHAPTER 3

# Borders

## Corpus segmentation

Tropes is equipped with a tool called “Borders”, designed to automatically segment a corpus. Within the same text, *Borders* can be used to automatically separate multiple actors, an interviewer and an interviewee, populations, chapters of a book, etc.

The use of *Borders* may require a preliminary coding of the documents.

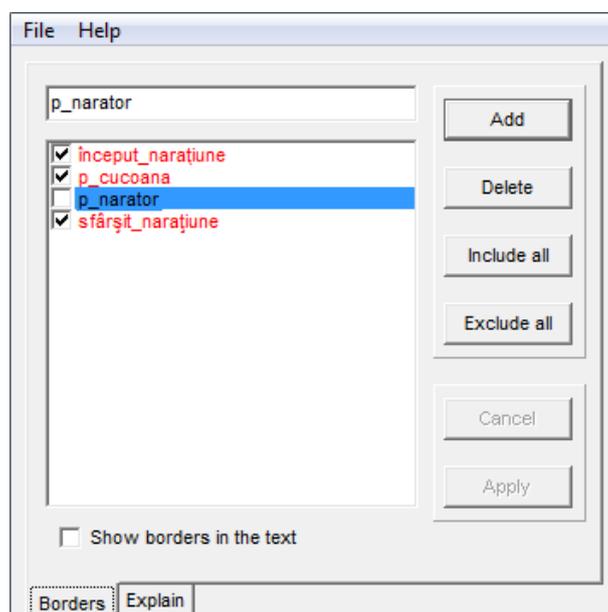
A *Border* starts from where it has been located in the text and ends where the next *Border* has been located. For example, if “p\_narator” and “p\_cucoana” are *Borders*, the sequence below delimits four parts of a text

	<b>Part</b>
p_narator – Frumuşel căţel aveţi, zic eu cocoanei, după câteva momente de tăcere; da' rău!	<b>1</b>
p_cucoana – Aş! nu e rău, zice cocoana; până se-nvată cu omul; dar nu ştii ce cuminte şi fidel este, şi destept! Ei bine! e ca un om, frate! doar ca nu vorbeşte...	<b>2</b>
p_narator – Madam! pentru Dumnezeu, tineţi-l să nu se dea la mine! eu sunt nevricos, şi nu ştiu ce-aş fi în stare... de frică...	<b>3</b>
p_cucoana – Vai de mine! cum crezi d-ta?... Noi suntem băieţi cumiţi şi binecrescuţi... Noi nu suntem mojiţi ca Bismarck.	<b>4</b>

These four parts correspond to a conversation between two characters, “*cucoana*” and “*narator*” (see the *Bubico.txt* file in the example texts supplied with the software). If you wish to automatically separate the discourse of “*cucoana*” (parts 2 and 4) from that of “*narator*” (parts 1 and 3), then you have to use *Borders*.

## Creating Borders

To create a *Borders* file, use the *[Tools][Borders]* command:



To create a new entry in the *Borders*, write a word in the upper field, then press the *[Add]* button. This word has to be representative of a sequence of the text analyzed. In the example shown above, the codes “inceput\_narațiune” and “sfârșit\_narațiune” have been used to identify the descriptive parts and the introduction of the play, while the various speech turns of the characters have been coded by putting the code “p\_” after the names of the characters.

To **delete** an existing *Border*, select it and press *[Delete]*.

Once you have created your *Borders*, you can choose which parts of the text you wish to ignore by checking the related *Borders*: when a *Border* is checked, then all the following text will be ignored; if not, the rest of the text will remain visible.

To disable all *Borders*, use the *[Include all]* button. To exclude all the text taken into account by the *Borders*, use the *[Exclude all]* button. This function enables you, for instance, to check whether or not the whole text is correctly delimited.

Once you have finished, use the *[Apply]* button to restart the analysis of the text. This time, the analysis will ignore all the parts of the text you have chosen to remove (when using *Borders* for the first time, the software will ask you to give a file name in order to save them).

When the *[Show Borders in the text]* option is checked, Tropes will count all *Borders* as words of the text and display them. Otherwise, if this option is not checked, Tropes will not display the related codes, which will not be counted as words.

**Note:** if your *Borders* are compounds, unrecognized by the software, you must link together the words forming these compounds with the underline character (for example “DI\_Negru”).

## Borders files

Use the *[File]* menu to:

- create a **new** (empty) *Borders* file,
- open** an already existing *Borders* file,
- save** the *Borders* file under another name.

*Borders* are automatically saved when using the *[File][Save]* menu or pressing the *[Apply]* button. To quit this tool without saving your modifications, click on *[Cancel]*.

# Tropes<sup>®</sup>

## CHAPTER 4

### **Introduction to text analysis**

## How Tropes operates

To process a text, Tropes operates in 6 stages:

- 1 -sentence and proposition hashing,
- 2 -ambiguity solving (regarding the words of the text),
- 3 - identification of *Equivalent classes*,
- 4 -statistics, detection of *Bundles* and *Episodes*,
- 5 -detection of the *Most Characteristic Parts of text*,
- 6 - layout and display of the result.

Words are grouped together in several main *Word categories*. Among these, six are of interest to us:

- 1 - verbs,
- 2 - connectors (conjunctions, conjunctive phrases),
- 3 - modalities (adverbs or adverbial phrases),
- 4 - qualifying adjectives,
- 5 - personal pronouns,
- 6 - substantives and proper nouns.

To achieve an analysis, the software carries out a complex processing aiming at: assigning all the significant words to the above categories; analyzing their distribution into subcategories (*Word categories, Equivalent classes*); examining their occurrence order, both within the propositions (*Relations, Actants and Acted*) and throughout the text (*Distribution graph, Bundles, Episodes, Most Characteristic Parts of text*).

### Propositional hashing

To simplify the analysis, Tropes divides the text into propositions (simple sentences). This first stage is based on a scrutiny of the punctuation, and on complex syntax analysis functions, which will not be detailed here.

Thus, you obtain co-occurrence statistics (*Relations*) of high reliability, since it is not possible for two words to fit into the same grammatical proposition if they are not closely connected.

Propositional hashing is bound to involve errors (propositions that are either too short or too long), but this does not alter the results.

### Ambiguity solving

The automatic interpretation of words in any living language, either written or spoken, requires the solving of numerous ambiguities:

-grammatical and syntactic (the word “mare”, for example, corresponds to an *adjective* in “un *mare* scandal”, and to a *noun* in “Vara mergem la mare.”),

-semantic (the word “lună”, for example, can refer to our *natural satellite* or to *month*).

On average, one word out of four is deemed ambiguous.

One of the main functions of this software is to solve these ambiguities by means of several problem-solving algorithms. Though a perfect result is impossible to achieve, the error rate is low enough to guarantee an accurate analysis of your text.

### Word categories

The **verbs** are either:

-*factive*, i.e. expressing actions (“a face”, “a munci”, “a umbla”, etc.)

-*stative*, i.e. expressing states or concepts of possession (“a fi”, “a sta”, etc.)

-*reflexive* (*a.k.a declarative*), i.e. expressing a statement about circumstances, beings, objects, etc. (“a gândi”, “a crede”, etc.)

**Connectors** (coordinating and subordinating conjunctions, conjunctive phrases) link together various parts of the discourse through concepts of:

- *condition* (“dacă”, etc.),

- *cause* (“fiindcă”, etc.),

- *goal* (“cu scopul de”, etc.),
- *addition* (“și”, etc.),
- *disjunction* (“sau”, etc.),
- *opposition* (“dar”, etc.),
- *comparison* (“ca”, etc.),
- *time* (“când”, etc.),
- *place* (“unde”, etc.)

**Personal pronouns** are displayed in gender (“eu”, “tu”, “el”, etc.) and in number (“ei”, “noi”, etc.).

**Modalities** (adverbs or adverbial phrases) enable the speaker to get involved in what he says, or to locate what he says in time and space, through concepts of:

- *time* (“acum”, etc.),
- *place* (“aici”, etc.),
- *manner* (“rău”, etc.),
- *assertion* (“da”, etc.),
- *doubt* (“probabil”, etc.),
- *negation* (“nu”, etc.),
- *intensity* (“mult”, etc.)

**Adjectives** are either:

- *objective*, i.e. enabling to characterize beings or objects, regardless of the speaker’s standpoint (color adjectives, for example),
- *subjective*, i.e. indicating judgment on something or on somebody, thus enabling to express the speaker’s standpoint (“frumos”, “mic”, “nostim”, etc.),
- or *numeral*, i.e. grouping together numbers (in letters or in figures), along with ordinal and cardinal adjectives.

**Other word categories** include *pronouns*, *articles*, *prepositions* and *non-qualifying adjectives*. Do not take these categories into account, because they are used only for

ambiguity solving.

Broadly speaking, we can say that:

- *time* and *place connectors* and *modalities* provide the means to locate the action,
- *intensity* and *negation modalities* provide the means to dramatize the discourse,
- *cause* and *condition connectors* provide the means to construct a chain of reasoning,
- *addition connectors* provide the means to enumerate facts or characteristics,
- *opposition connectors* more specifically provide the means to argue, to put things into perspective and to set out conflicting standpoints.

### Statistical, probabilistic and cognitive analyses

Tropes carries out different sorts of text analyses:

- statistics on the total occurrence frequency of the main *Word categories* and of their subcategories,
- statistics on the total occurrence frequency of the main *Word categories* and of their subcategories,
- a probabilistic analysis of the words occurring in *Bundles*, and a geometric analysis of the *Bundles* delimiting the *Episodes*,
- a Cognitive-Discursive Analysis (CDA), making it possible to detect the *Most Characteristic Parts of text*.

Among other things, statistics are used to build the graphs and to lay out the results.

The *Frequent word categories* and the *Text Style* are obtained by comparing the distribution of the occurrence frequency of the categories observed in the text with linguistic production norms. These norms have been elaborated after studying a great number of different texts. They are stored into specific in-built tables.

## Equivalent classes and Relations between equivalents

The *Equivalent classes* group together closely related References (common nouns, proper nouns, trademarks) appearing frequently throughout the text. For example, “tată” and “mamă” are grouped together into the “familie” *class*.

The *Reference fields* group together the words comprising the *Equivalent classes* in order to enable the software to elaborate a representation of the context. To achieve this, the *Semantic equivalents* dictionary of Tropes is composed of three different classification levels. At the lower level are the *References*, which are next merged more broadly into *Reference fields 2*, which, in turn, are merged into *Reference fields 1*.

animale	mamifere	irbis	irbis
animale	mamifere	irbis	irbiși
animale	mamifere	irbis	irbișii
animale	mamifere	irbis	irbișilor
animale	mamifere	irbis	irbisul
animale	mamifere	irbis	irbisului
animale	mamifere	irbis	leopardul_zăpezilor
animale	mamifere	irbis	leopardului_zăpezilor
animale	mamifere	irbis	leopardzii_zăpezilor
animale	mamifere	irbis	leopardzilor_zăpezilor
animale	mamifere	irbis	panthera_uncia

### Reflections on the size of the texts

Most of the analyses made by Tropes being statistical, the software is sensitive to the amount of text to be processed.

If your corpus consists of files containing **very short** texts (under one page), the results will probably not be significant and the software may have some difficulty in resolving semantic ambiguities (there will not be enough words to clarify the implicit concepts). If you work on a reasonable number of short texts, we recommend that you put them together in a single file. If you work on a great number of short texts, put them in several files, where you can group them in chronological order, for instance.

Otherwise, if your corpus consists of files containing **very long** texts (over 100 pages) and/or composed of utterances referring to completely different contexts, the pertinence of the results may be affected: the discourses get mixed up. When working on rather substantial corpora, try to divide them -as much as possible -into files that do not exceed a few hundred pages. You can use a *semantic Scenario* to merge the results obtained by the analysis of several files.

A few “symptoms” indicate that the corpus you are analyzing is too substantial:

- It takes more than one minute for Tropes to process the text.
- the *Equivalent classes* are hardly usable: they often contain words of different meanings.
- Tropes has not detected any significant *Actant*, and this fact cannot be explained by the intensive use of personal pronouns.

**Important note:** because of inner technical limits of the software, some *References classes* may not be displayed if the text is very long.

**Analysis of heterogeneous utterances: open questions, dispatches, enumerations, etc.**

The following notes concern all heterogeneous corpora, obtained by collecting within the same file utterances coming from numerous individuals, and without linear coherence (i.e. it is not the discourse of a single narrator, or an interaction between several interlocutors respecting a logical sequence or a strict chronological order).

Since the propositional hashing made by Tropes is based entirely on grammatical rules, you must use a non-ambiguous punctuation mark (question or exclamation mark) to force the software to separate the different utterances (for example, you can add an exclamation mark at the end of every answer to an Open question (Market research), in order to separate it from the next one).

If the corpus includes answers to several questions, you will have to use *Borders* to group the answers together and/or to separate the answers from the questions. You will then analyze each answer separately.

If you have an indicator enabling you to form your corpus according to an external variable (geographic area, type of population, period, etc.), you may use it to split up the utterances into several files (each file containing, for example, the utterances corresponding to only one variable), that you can analyze separately in order to compare them. You can also code these variables inside the texts and then use them as *Borders*.

If the corpus has no linear coherence (for example, when the utterances contained in each file have been compiled at random, without following a particular logic), the results depending on the chronological analyses of Tropes (*Most Characteristic Parts of text, Bundles, Episodes, Distribution graphs*) will not be significant; do not try to interpret these results.

### Analysis of discourses and conversations

When analyzing a text file containing the transcript of the discourse of several individuals, start with an overall analysis of the corpus, then use *Borders* to process the utterances of the various characters separately, and compare the results obtained. When comparing the results, ask yourself the following questions: have all the participants been talking about the same thing? Did they use the same *Actants*? If not, then why not? Has anyone refused to reply to certain questions? Has anyone been trying to convince another participant? Why? Have they succeeded? Etc.

If you have time, you can solve the anaphoras manually, i.e. replace each personal pronoun by who it refers to. Let us imagine, for example, that you wish to analyze the discourse of two characters - Petre and Paul - who have been talking about three other persons - Alin, Maria and Iulia -, and that the text contains many personal pronouns ("eu", "tu", "el", "ea", etc.) Use the *[find/replace]* command of your word processor to replace some of these pronouns as follows: "eu" and "tu" by "Petre" or "Paul", "ea" by "Maria" or "Iulia", "el" by Alin, etc. You will thus be able to count very precisely how many times this or that person has been mentioned, to know whether they are *Actants* or *Acted*, etc.

When you make transcripts from the spoken form, it is necessary that you include punctuation in the text, otherwise the software will not be able to carry out the propositional hashing properly and the processing of the analysis will be altered.

### Literary studies

To analyze a play, use the above method (analyzing the utterances of multiple actors is almost equivalent to analyzing a conversation between different interlocutors).

When studying long texts, such as an entire book, first analyze each chapter separately, then you can make a synthesis by processing the whole text (see *Reflections on the size of the texts* above).

## Comparing two texts

Comparing two texts comes down to making an analysis both of the contents (i.e. of the *Equivalent classes*) and of the *Setting* (i.e. of the *word categories*).

For example, you can compare:

- the respective weights (*weighted utilization rates*) and also the positions (*Actant/Acted*) of the *Equivalent classes*,
- the occurrence chronology of the main themes (*Bundles, Episodes* and *Distribution graphs*),
- the co-occurrence frequency of the References (*Star* and *Area graphs, Relations, Connection rate* of the *Relations, Scenarios*),
- the various types of action, through the analysis of the verbal categories (*Word categories, Styles* and *Settings*),
- the type of logic at work and of *involving*, through the analysis of *Connectors* and *Modalities* (*Word categories, Bundles, Episodes*).
- and, in general, all propositions enabling the introduction of the main characters and themes (*Most Characteristic Parts of text*).

# Tropes<sup>®</sup>

## CHAPTER 5

### Appendices

## Files conversion

Observations about files format conversions:

<i>Format</i>	<i>Extension</i>	<i>Description</i>
ASCII	*.ASC	The software uses Windows API function OemToChar to convert this format in ANSI ISO-8859-1.
HTML	*.HTM *.HTML	See XML and SGML remarks (below) for HTML (Hypertext Markup Language) file formats conversion.
Macintosh	*.MAC	The Apple Macintosh <sup>®</sup> Latin character sets are converted to ANSI ISO-8859-1 using the Windows API.
Microsoft Word	*.DOC *.DOCX	The software makes a conversion of Microsoft Word <sup>®</sup> 97-2003 files using the 32 bits IFilter drivers on your system. Besides this, the software can try extract the text by binary analysis, for problematic or not recognized files. The installation Microsoft Word <sup>®</sup> or Office <sup>®</sup> 2007 (or the equivalent 32 bits IFilter pack) is a prerequisite for DOCX documents.
PDF	*.PDF	Software uses a specific driver (Adobe <sup>®</sup> IFilter) to extract the relevant text of PDF <sup>®</sup> (Portable Document Format) files. If Adobe Reader <sup>®</sup> installation is not needed, it is necessary on the other hand to install a PDF IFilter driver to benefit from this file format. The usage of external character recognition software (OCR) may be necessary for some files (those that come from a digitalization by scanner).
RTF	*.RTF	The software interprets RTF (Rich Text File) formats containing characters in ANSI ISO-8859-1, Apple Macintosh, or ASCII IBM, coded on 7 or 8 bits. Unicode format is not accepted. Because disparities exist between RTF standards, parasite characters can appear in certain files.
SGML	*.SGM	The software discards the tags of SGML (Standard Generalized Markup Language, ISO 8879) and converts some HTML specific variables or entities (e.g. some characters or diacritics). It does not interpret the DTD. UTF-8 format (UCS transformation 8, ISO 10646 / RFC 2279) is converted in ANSI ISO-8859-x. The other Unicode formats (Universal Character Set, ISO 10646) are not accepted.
Text	*.TXT	No conversion is made on this file format, considered in ANSI ISO-8859-1 (a.k.a. ISO-LATIN1, or ANSI Windows) or in Unicode UTF-8 (if a Byte Order Mark is in the file header).
XML	*.XML	The software uses its SGML parser (see above) to read XML files (Extensible Markup Language). The same limitations are also applicable to this format. The conversion engine does not interpret scripts or style sheets.

The conversion of PDF<sup>®</sup> documents may require the installation of an Adobe<sup>®</sup> IFilter component on your system. You can download the latest version of IFilter on Adobe's site: <http://www.adobe.com/support/downloads/detail.jsp?ftplID=2611>. Note that the addition of this component does not require installation of Adobe Reader<sup>®</sup>.

The software uses the current IFilter components installed on your system to convert Microsoft Word<sup>®</sup> documents. In case of problems (corrupted file, etc.) conversion automatically switches over to binary analysis (heuristic), which attempts to recover the

## Tropes – Reference Manual

text by binary origin. If you notice that the software jams on certain Word files, you can deactivate conversion using the Analysis options dialog (tab-sheet [Conversions]).

It is never possible for password-protected files to be read and directly converted by the software. These files must be converted manually for analysis.

When the software performs a binary analysis (heuristic) to extract the text, parasitic characters may appear. By definition, this method of origin of the text cannot be perfect, because the software does not take into account the native file format.

For more information about the files supported, see:

- <http://www.apple.com>,
- <http://www.adobe.com>,
- <http://www.microsoft.com>,
- <http://msdn.microsoft.com>,
- <http://www.unicode.org>,
- <http://www.utf-8.com>,
- <http://www.macromedia.com>,
- <http://www.w3.org>,
- <http://www.w3.org/MarkUp>,
- <http://www.w3.org/MarkUp/SGML>.