



T-Seg

User guide

Content

1. Working with files	2
1.1 Open a file	2
1.2 Save a file	3
1.3 Save data in a new file.....	3
2. Exit T-Seg	5
3. Text segmentation	6
3.1 Manual segmentation	6
3.1.1 Define a simple unit	6
3.1.2 Define a embedded unit.....	7
3.1.3 Create an overlapped segment	8
3.1.4 Define an independent unit	9
3.2 Automatic segmentation.....	11
3.2.1 Segmentation of all files in a directory	12
3.3 Removal of units.....	15
3.3.1 Remove a unit	15
3.3.2 Remove all units	17
3.4 “Edit” Menu.....	19
3.5 “Statistics” tab.....	20
3.6 “Segmentation - Colors” tab	21
4. Settings.....	23
4.1 Colors of elements on “Segmentation - XML” tab	23
4.2 Colors of elements on the “Segmentation - Colors” tab.....	25
4.3 System language	27
5. Help	29
5.1 User guide	29
5.2 About T-Seg	30

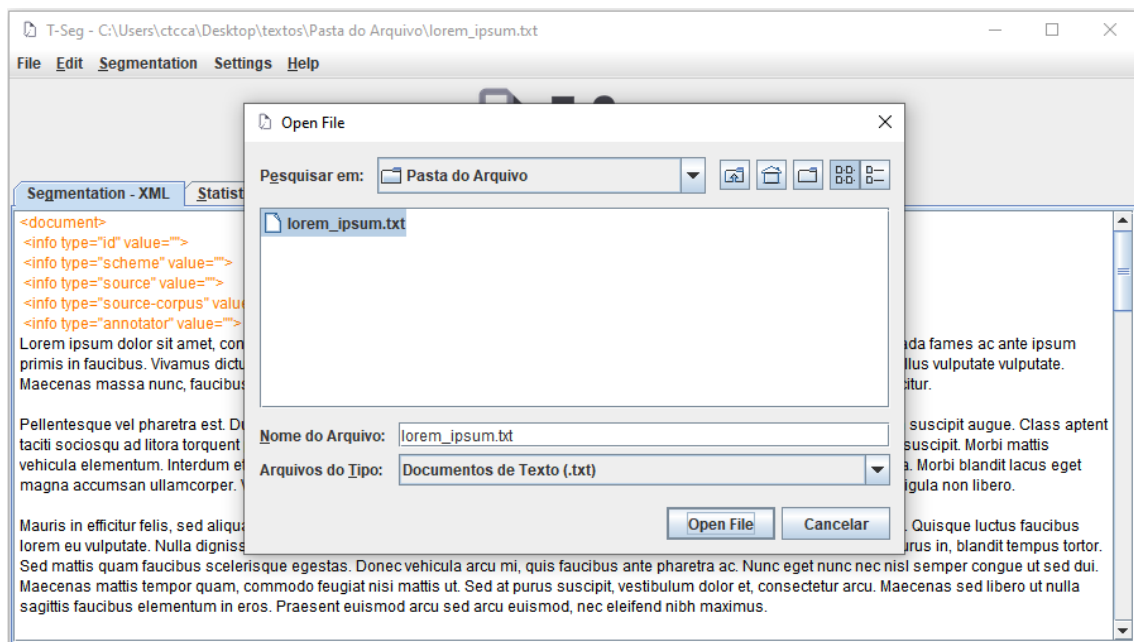
1. Working with files

1.1 Open a file

Opening a file on T-Seg is very simple task: left-click on the “File” menu and then click on “Open”.



After that, a file selection window will open. Choose the “txt” format file you want to open and click “Open File”.



You can also use the keyboard shortcut for this purpose, pressing together the “Ctrl” and “O” keys to display the file selection window.

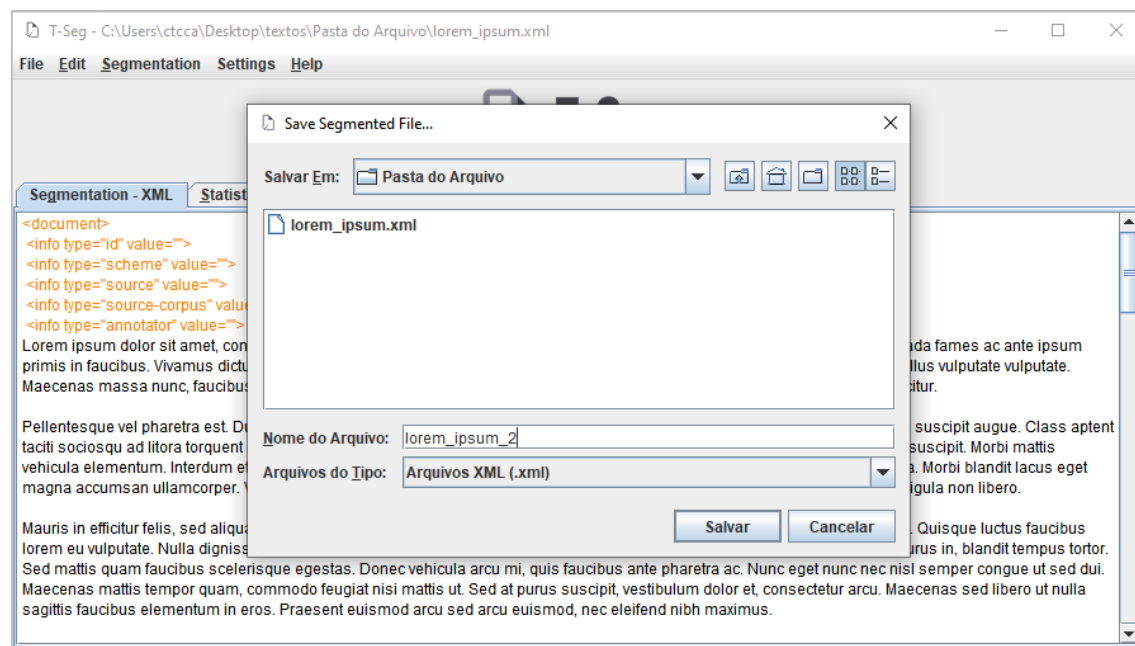
1.2 Save a file

To save the data of a file whose contents have been modified, go to the “File” menu and click on “Save”. You can also use the keyboard shortcut by pressing the “Ctrl” and “S” keys at the same time. In either case, a message will confirm saving the file:



1.3 Save data in a new file

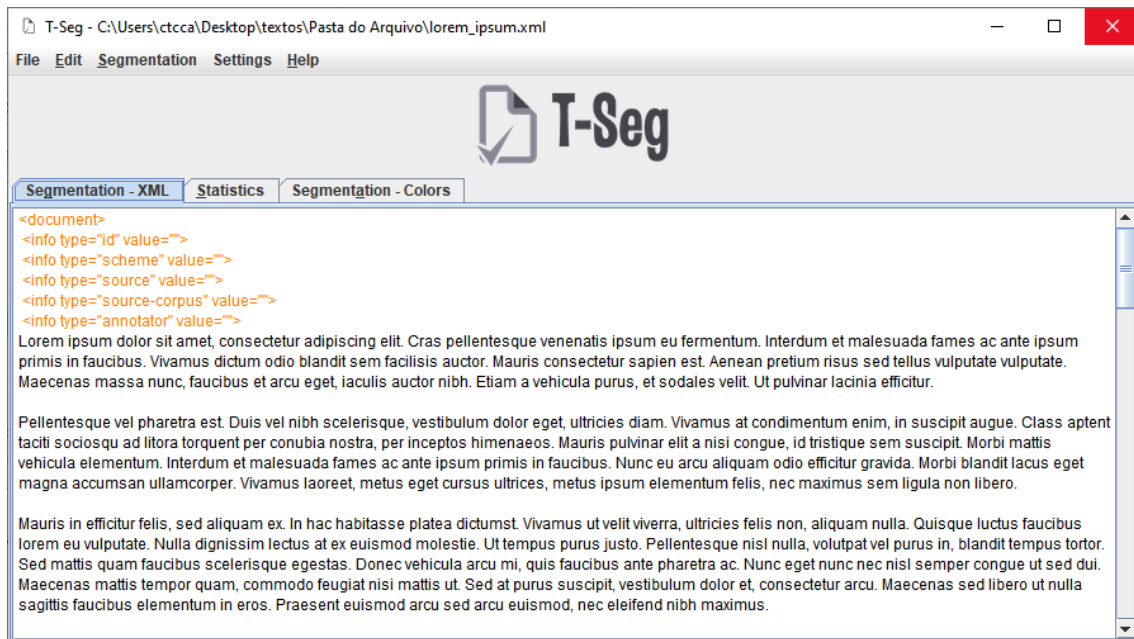
If you want to save the modified data in a new file, go to the “File” menu and click on “Save As”. A window will open to select the location where the file should stay:



2. Exit T-Seg

There are three ways of exit T-Seg:

- By clicking on the "Exit" option in the "File" menu
- By clicking on the "close" (X) option in the title bar
- Using the keyboard shortcut "ALT + F4"



3. Text segmentation

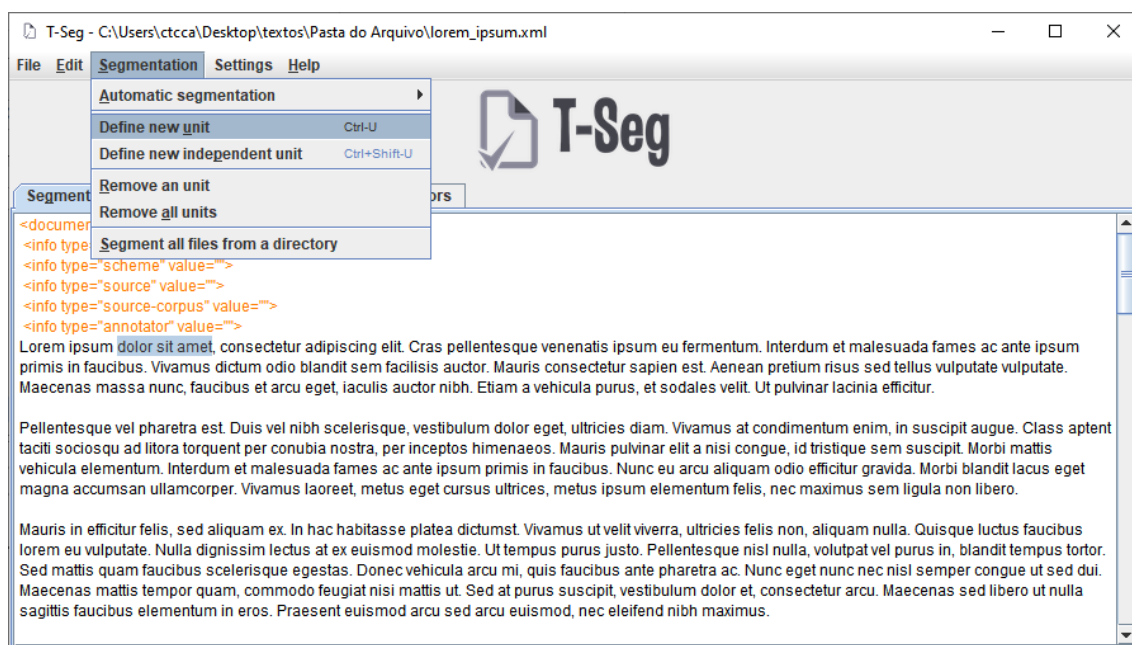
The “Segmentation – XML” tab is the main tab for annotation and segmentation work. In it, direct interaction is not allowed, that is, it is not possible to edit the text loaded within the T-Seg program. Changes allowed by the program are limited to annotation and segmentation only.

Segmentation tasks divide into two categories: manual segmentation – in which a user defines the basic units of annotation – and automatic segmentation – in which the program, through a set of rules, defines the units. The “Segmentation” menu brings together all tasks related to text segmentation.

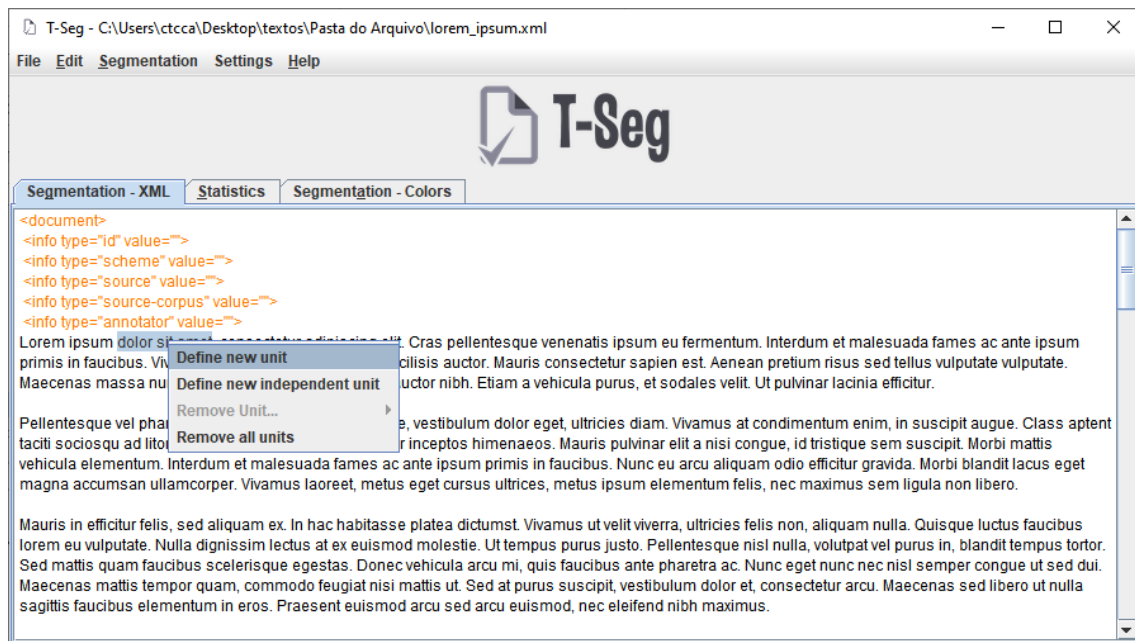
3.1 Manual segmentation

3.1.1 Define a simple unit

A simple unit is a basic unit of annotation. To define a simple unit, it is first necessary to select the text fragment that you want to define as a unit. With the text segment selected, just click on the "Segmentation" menu and then on "Define new unit":



There are another two ways to perform the same task: using the keyboard shortcut “CTRL + U”, or by right clicking on the selected area and then clicking on “Define new unit”:



After the action of defining a new unit, the selected text will be placed inside tags that identify the beginning and end of the unit:



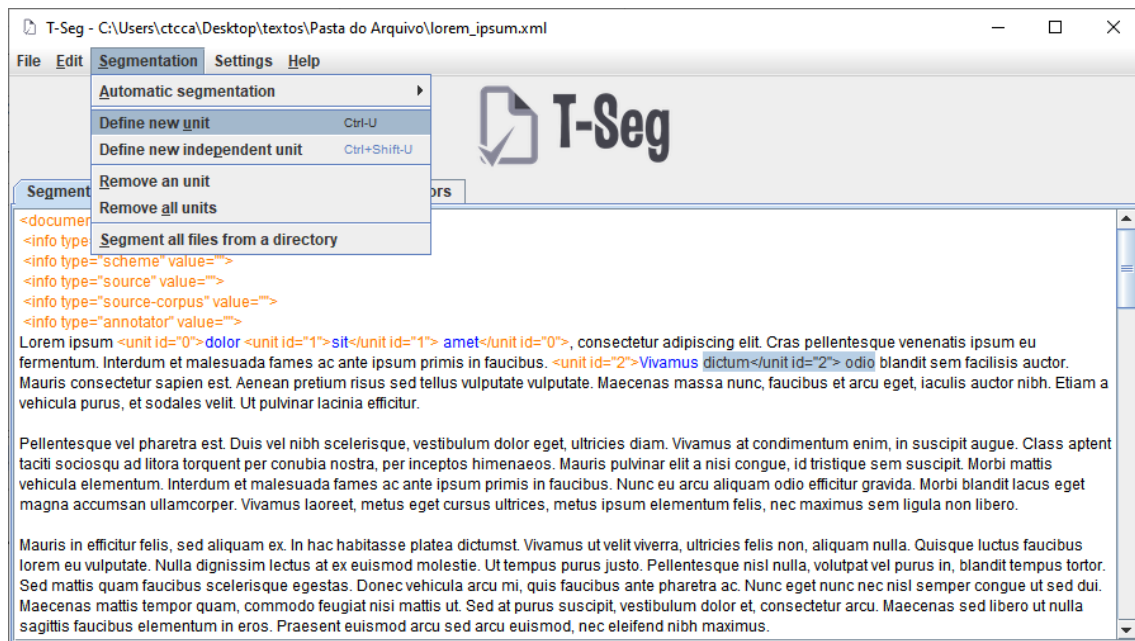
3.1.2 Define a embedded unit

A “embedded” unit is a unit that sits inside another unit. Defining a embedded unit, in T-Seg, just consists of defining a simple unit inside another simple unit:



3.1.3 Create an overlapped segment

An overlapped segment is a segment that belongs to two units, without one is inside another, different from an embedded unit. To define an overlapped segment, simply define a new unit containing a text segment that is already inside an existing unit:



3.1.4 Define an independent unit

An independent unit is a "special" embedded unit. It should be used whenever a text segment belongs to a unit, but does not belong to the longer unit that incorporates that unit.

These units are independent of the conditions that simple units have, such as: if an independent unit is inside a simple unit, the system does not count that as an embedded unit. Independent units are marked with the IND identifier in the tag that identifies the unit.

To define an independent unit, just select the desired text segment and click on the "Define new independent unit" option within the "Segmentation" menu:



It is possible to use the keyboard shortcut by simultaneously pressing the "Ctrl", "Shift" and "U" keys, or by right-clicking on the selection and then clicking on "Define new independent unit":



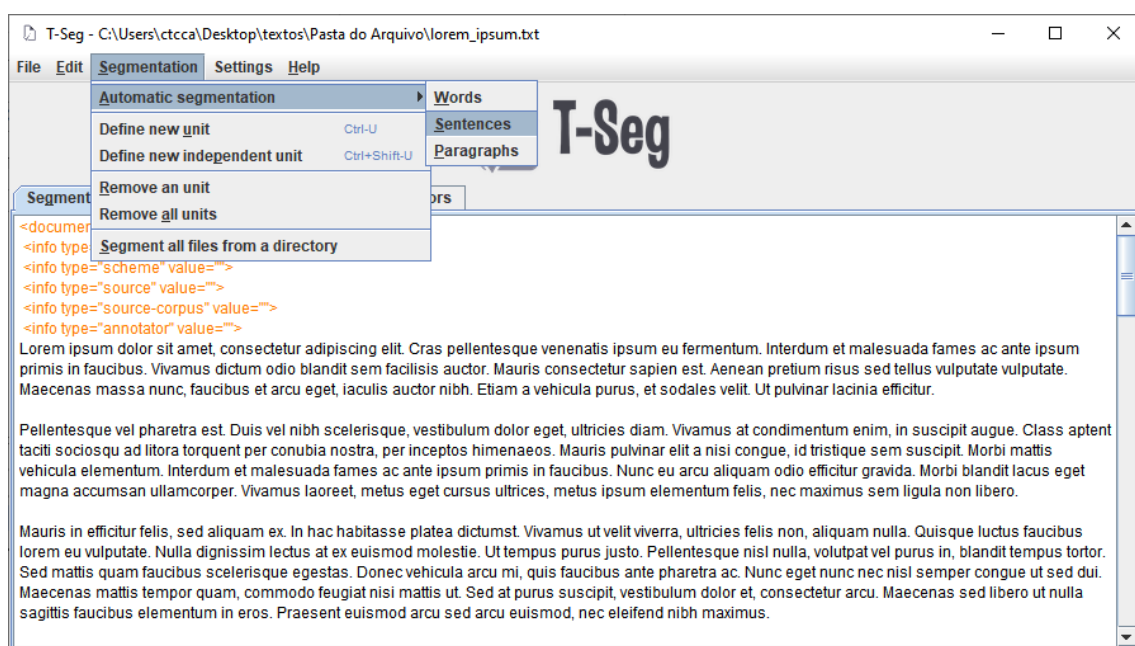
After adding the independent unit, the text segment is marked by tags that identify the beginning and end of the new unit:



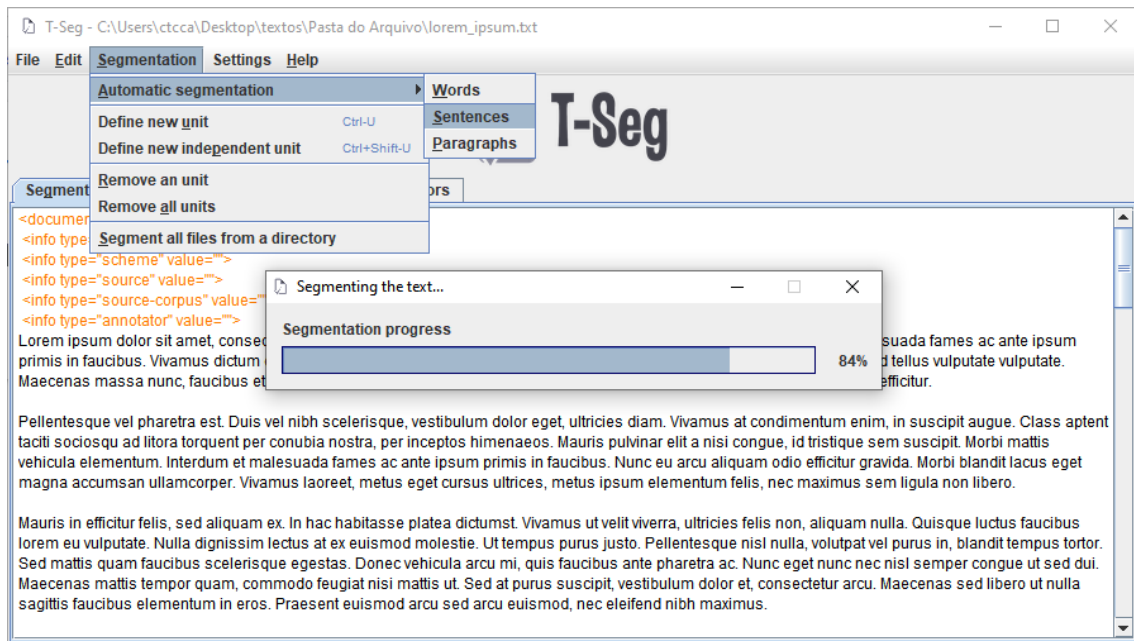
3.2 Automatic segmentation

T-Seg allows automatic segmentation, according to a predefined format. The user can choose to automatically segment by words, sentences or paragraphs. After segmentation, the user can edit the units as needed.

To segment the text automatically, just click on the “Segmentation” menu, point the mouse pointer over “Automatic segmentation” and, in the menu that opens, click on the appropriate segmentation format.



During the segmentation, a window will inform the progress of the task:



After finishing the segmentation, the progress window closes and the segmented text is displayed in the "Segmentation - XML" tab:

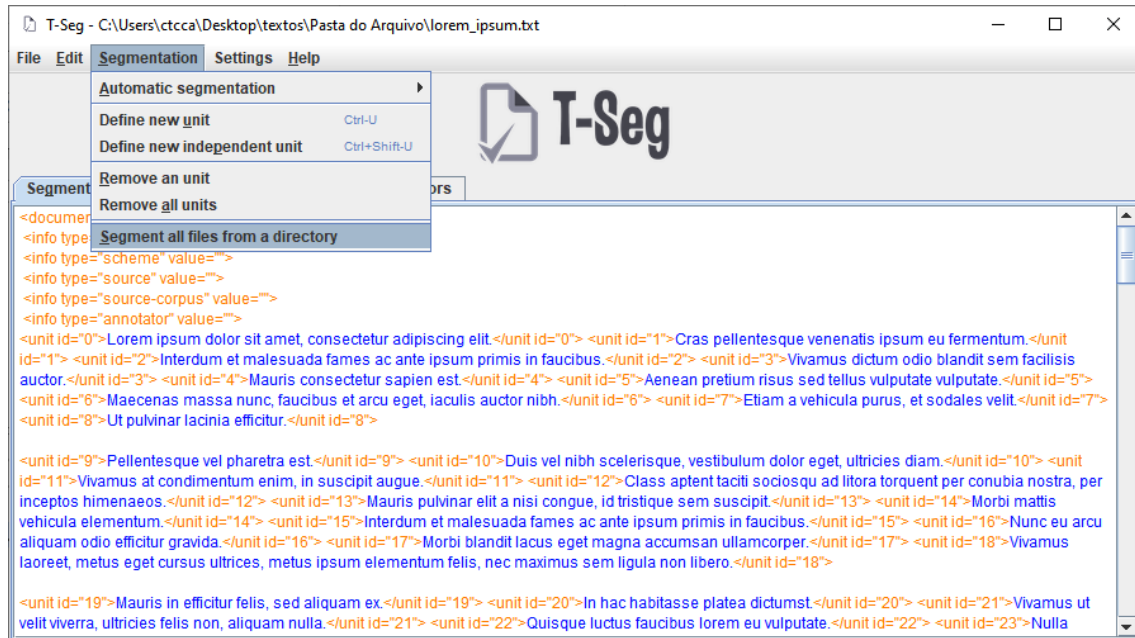


3.2.1 Segmentation of all files in a directory

There is also the option to automatically segment all text files in a directory. In this feature, the user enters a directory that contains text files, determines the target format, and the program takes care of selecting all text files and segment them,

producing the same result that would occur if a user had opened each file automatically segmented it.

To segment all files in a given directory, the user must first click on the "Segmentation" menu and then on the "Segment all files from a directory" option:

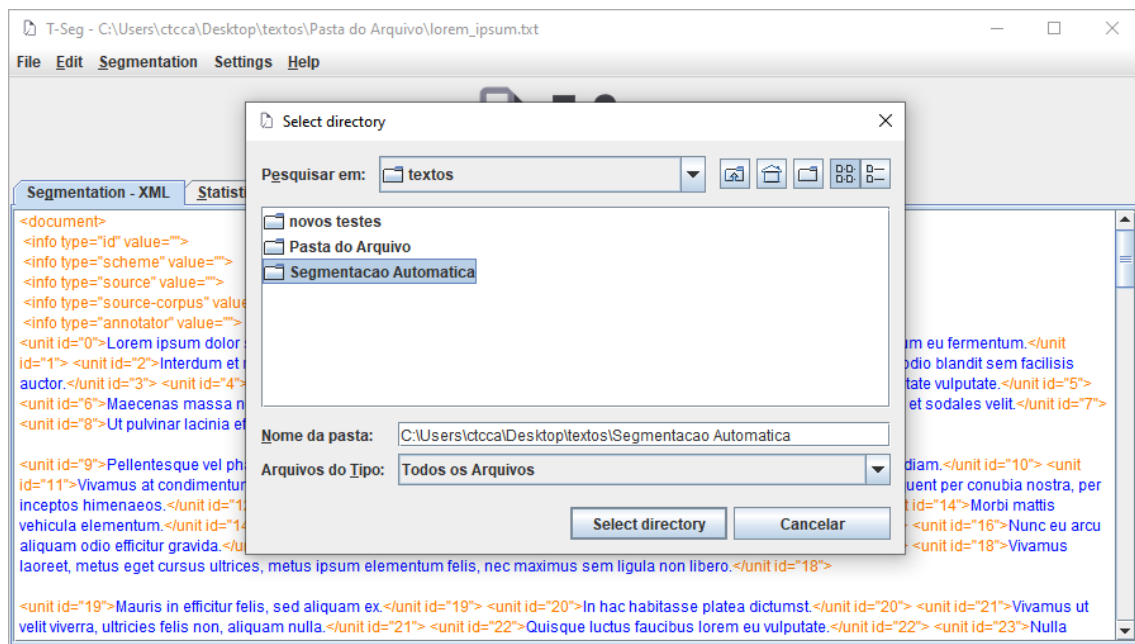


A window will open, containing the following elements for configuration:



- Directory: That's the directory that contains the files that will be segmented. To select the directory, the user must click on the "Find" button in the window, and choose the directory in the selection box that will open.

- Segmentation format: that's the format in which the files in the selected directory will be segmented, that is: words, sentences or paragraphs. The user must select a format. The option "Words" is chosen by default.



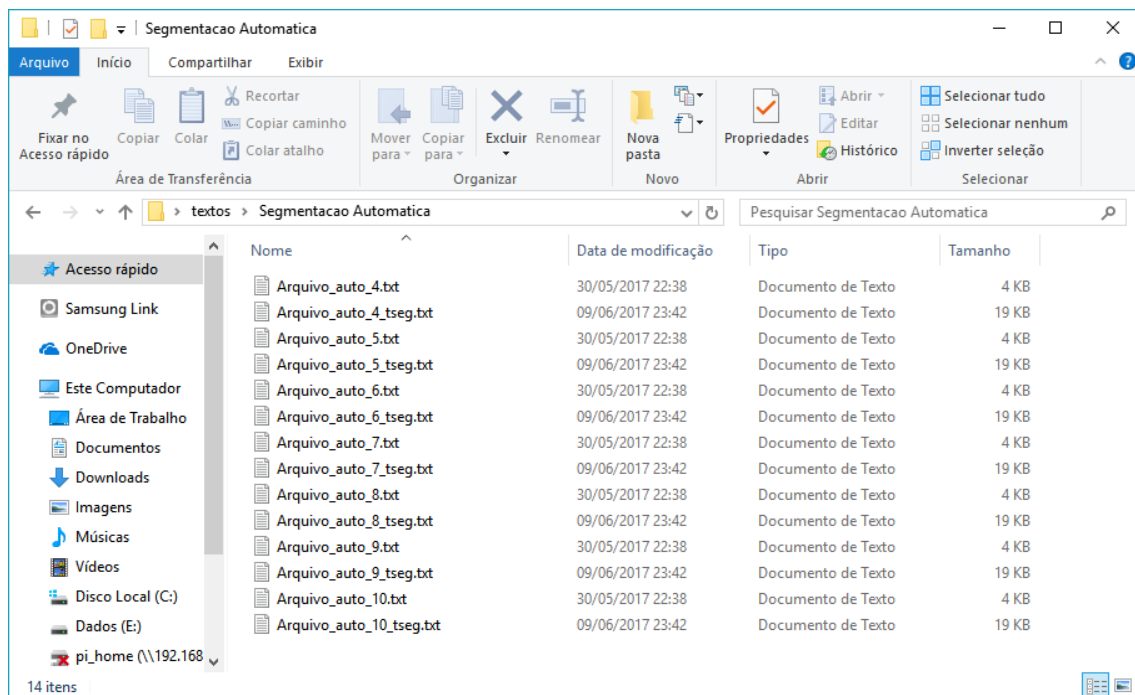
After selecting the segmentation directory and format, just press the "Segment" button to start the segmentation task. During segmentation the "Task status" field and the progress bar show the progress of the segmentation.



After the segmentation task ends, a message informs you that all files in the directory have been segmented and saved:



The program saves segmented files in the same directory, with the same name and extension, and concatenates the suffix “_tseg”:

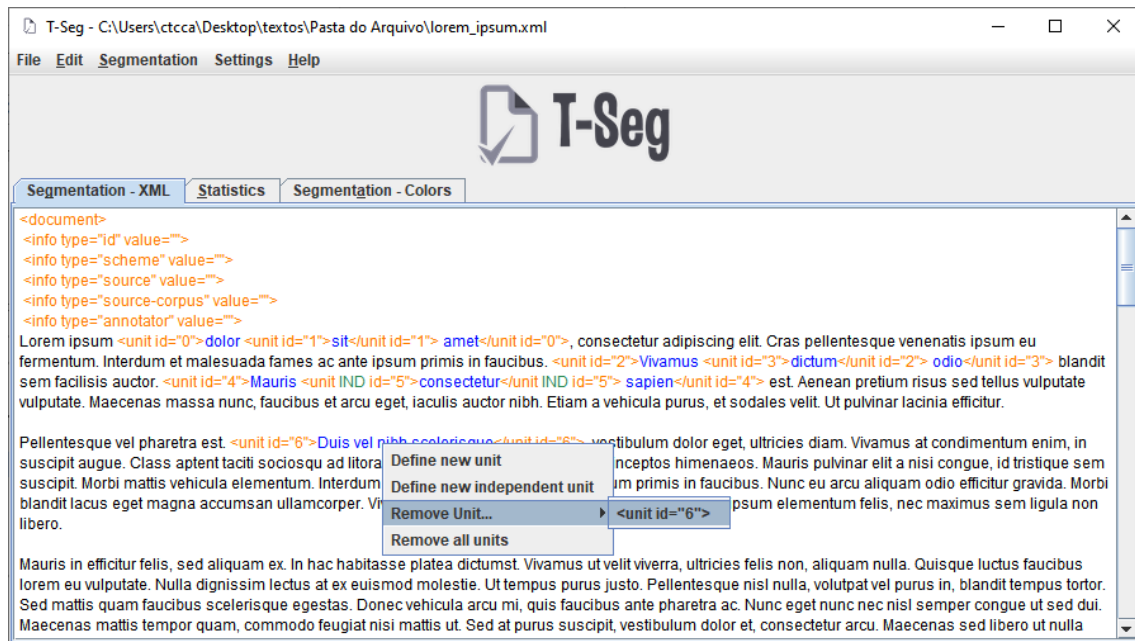


3.3 Removal of units

3.3.1 Remove a unit

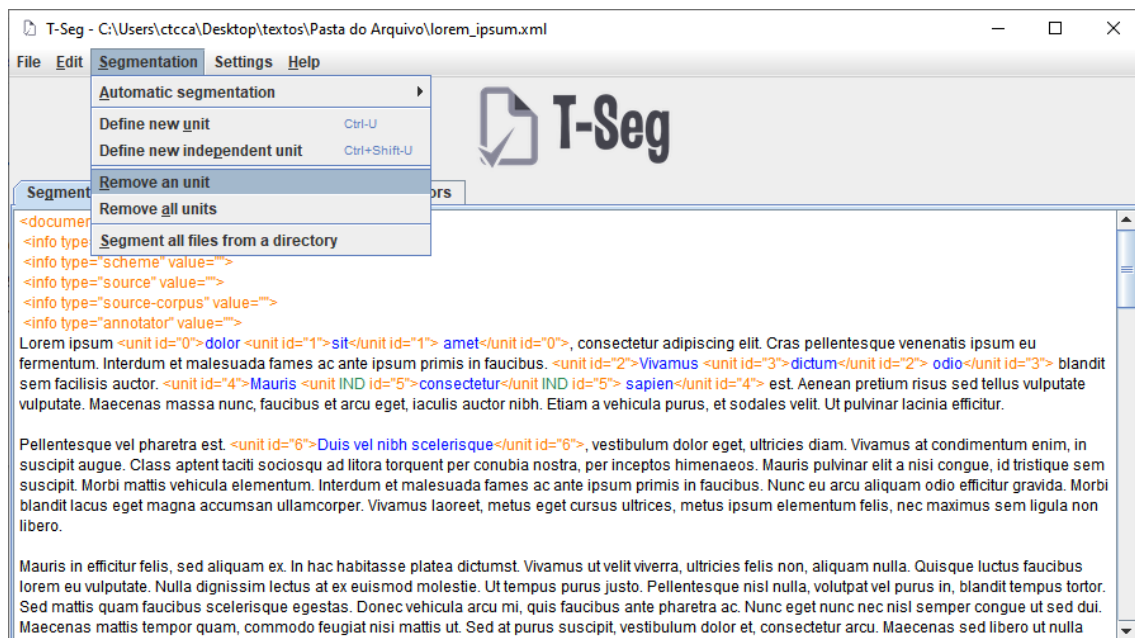
There are two ways to remove a specific unit.

The first is to select a text fragment that contains or is contained in the unit you want to remove, and right-click on that fragment:

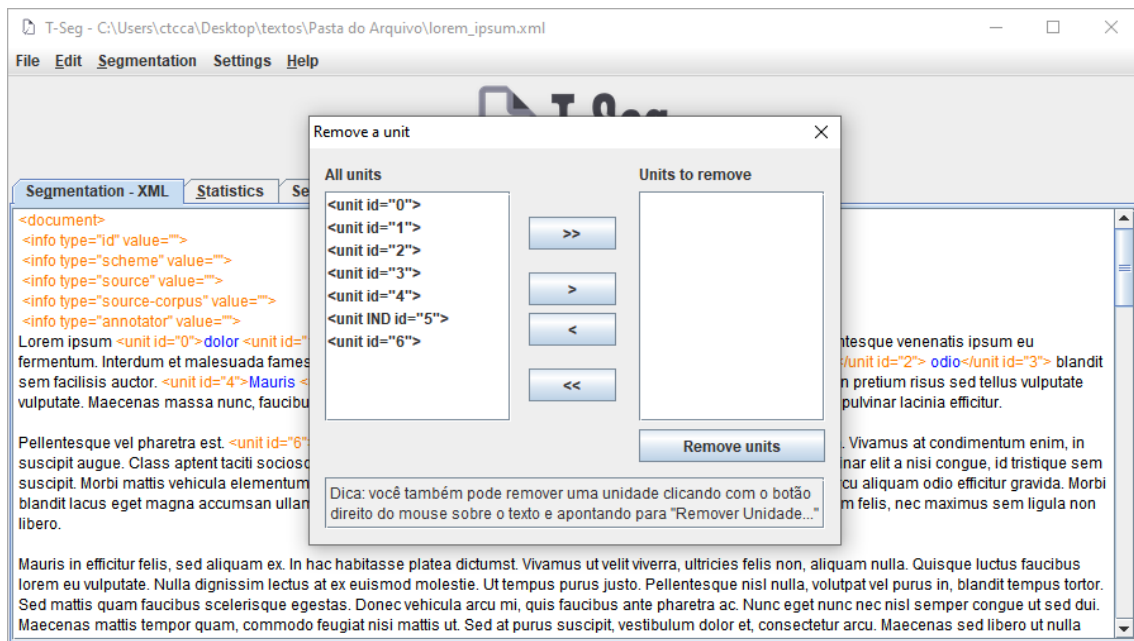


In the menu that opens, position the cursor on “Remove Unit...” and choose the unit you want to remove.

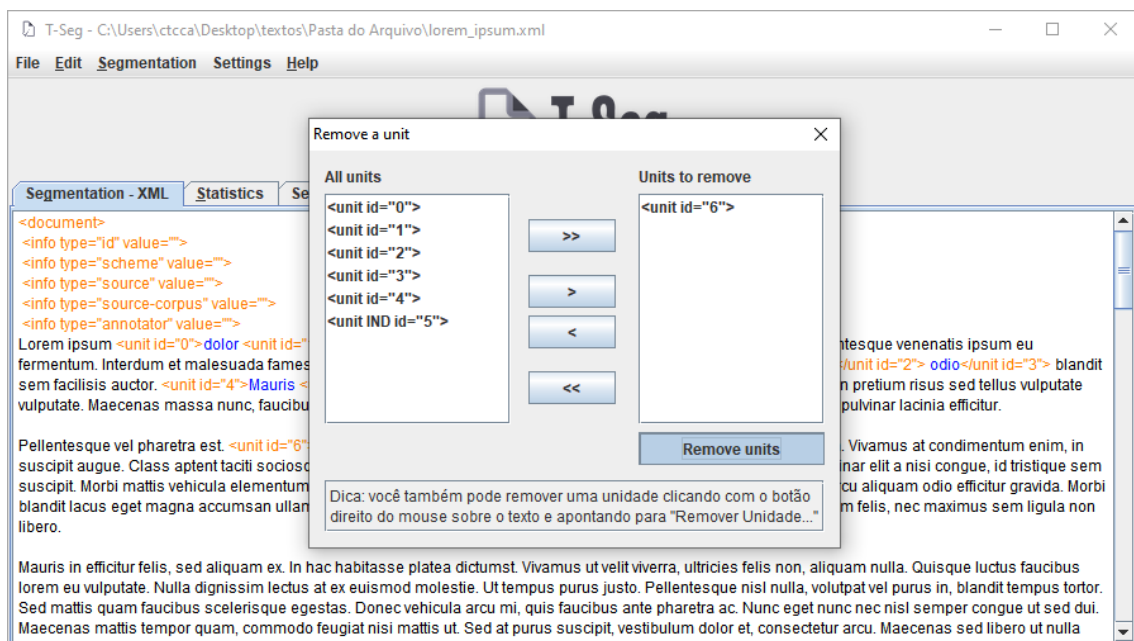
The second way is to click on the “Segmentation” menu and then on “Remove an unit”:



A window will open with two lists: “All units”, on the left, will contain all text units and the “Units to remove” list on the right will be empty.

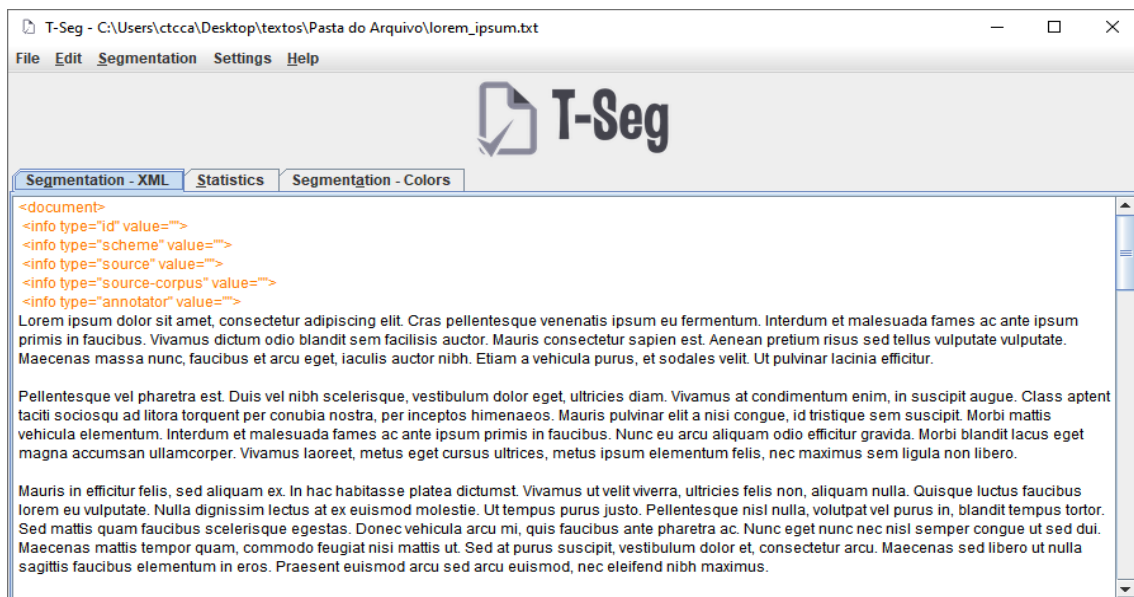


The user must select, in the first list, which units he wants to remove, send them to the other list, with the buttons “>”, “<”, “>>” and “<<”. Then, to perform the removal, just click on “Remove units”.



3.3.2 Remove all units

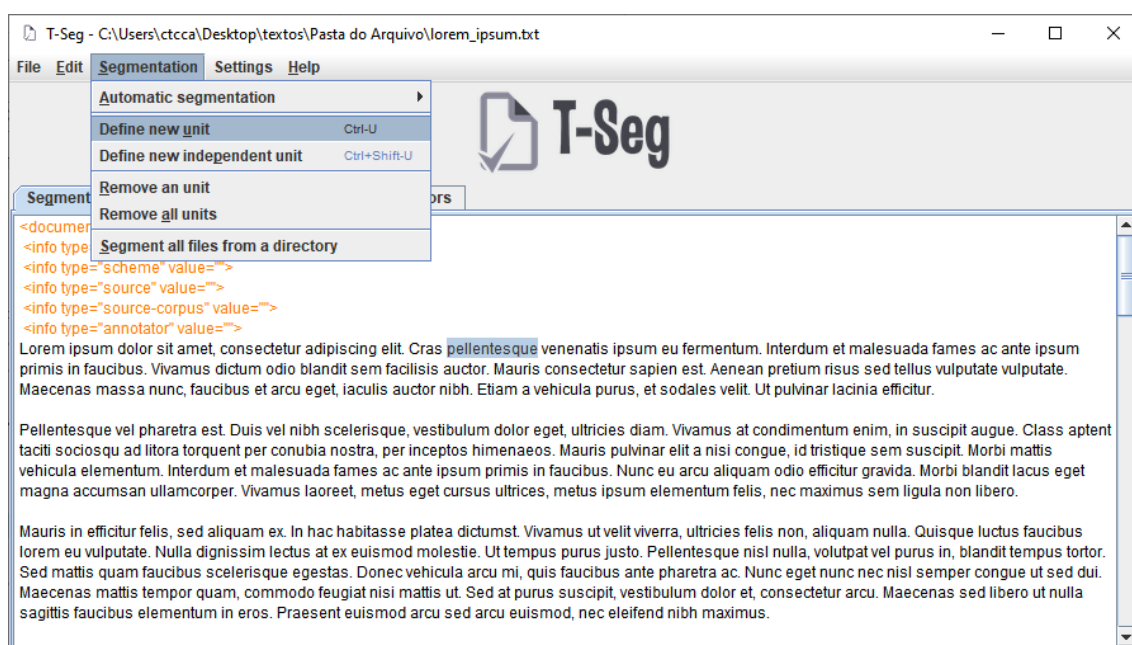
There is also a feature to remove all units of the text, leaving it in its original state, that is, without tagged segments. To remove all units from the text, just click on the “Remove all units” option in the “Segmentation” menu and then click on the “Yes” option in the confirmation window that appears:



3.4 “Edit” Menu

The “Edit” menu contains only two options: Undo and Redo, which can be accessed through the shortcut keys "Ctrl + Z" and "Ctrl + Y", respectively. As is the case by default on systems, these options are intended to undo actions or redo them after they are undone.

All segmentation actions of the current file, ie definition of simple and independent units, automatic segmentation and removal of units can be undone or redone using the options at this menu.

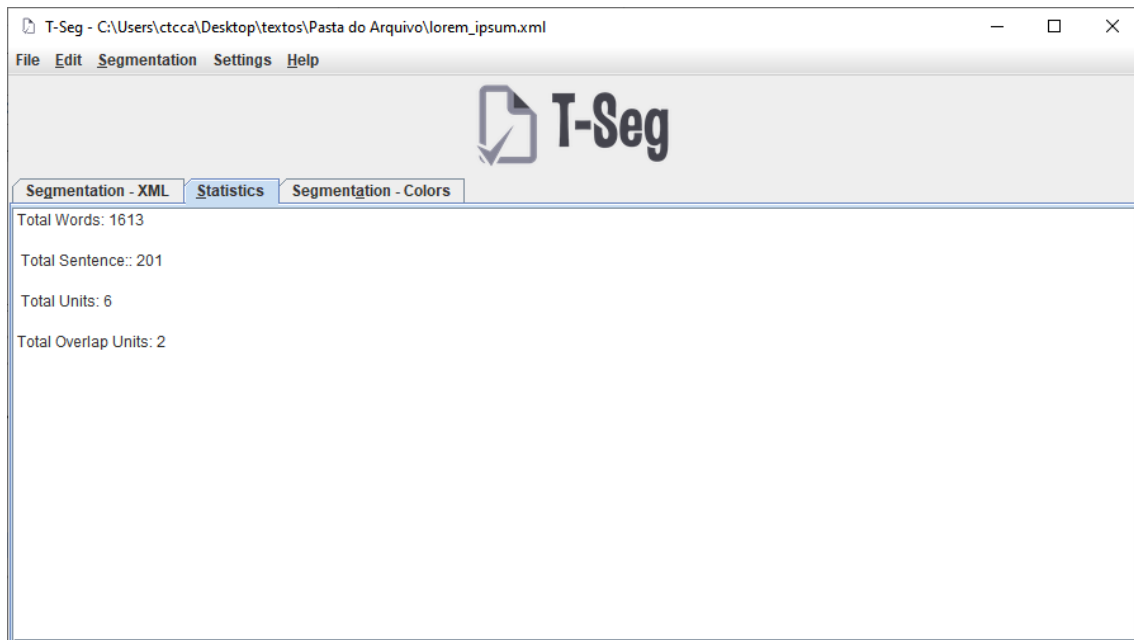




3.5 “Statistics” tab

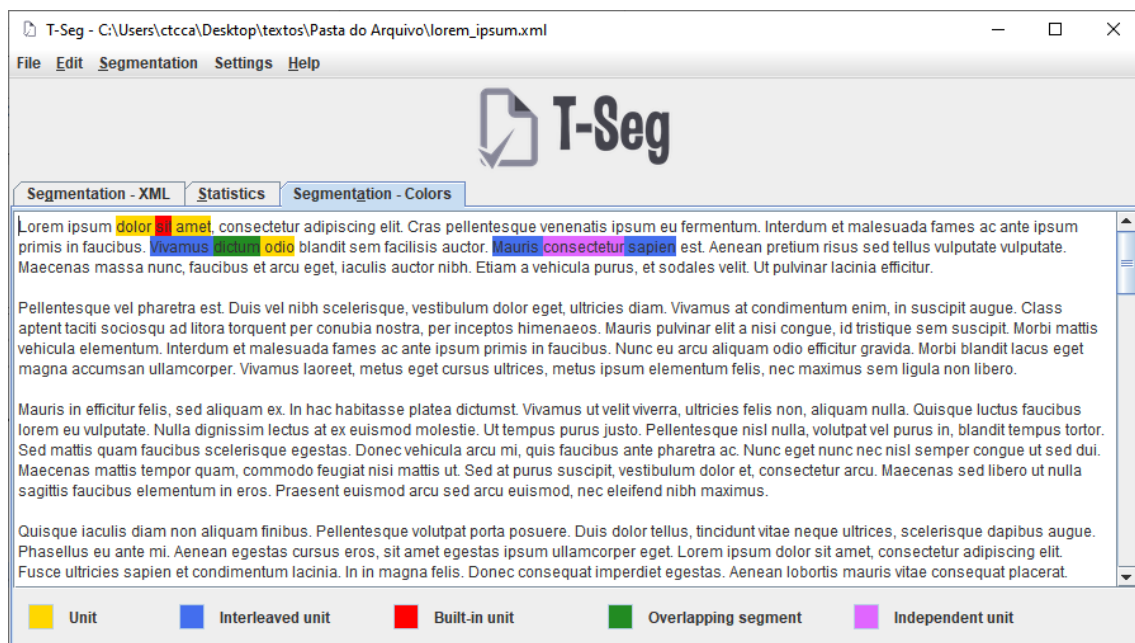
The main purpose of the “Statistics” tab is to present statistics in relation to the loaded text, that is, the amount of words, sentences, units and overlapping units. Important note: on this tab, both built-in units and overlapping segments are counted as overlap units.

To view the text statistics, just click on the “Statistics” tab on the right side of the “Segmentation – XML” tab.



3.6 “Segmentation - Colors” tab

This tab shows a segmentation scheme represented by colors instead of the traditional XML tags, unlike the "Segmentation - XML" tab. By clicking on the "Segmentation - colors" tab, the system generates and display the color scheme:



The legend in the lower corner of the window explains what each color represents in the segmentation scheme. By default, single units are represented by the colors yellow and blue, alternating to represent the beginning and end of each unit. Embedded units are marked with a red background, while overlapping segments, by default, have a green background. Finally, independent units are marked with a purple background.

These colors are default, but the user can change the background colors at will. This option can be seen in detail at item 4.2.

4. Settings

The “Settings” menu has basic options to configure T-Seg. The sections above explain these features.

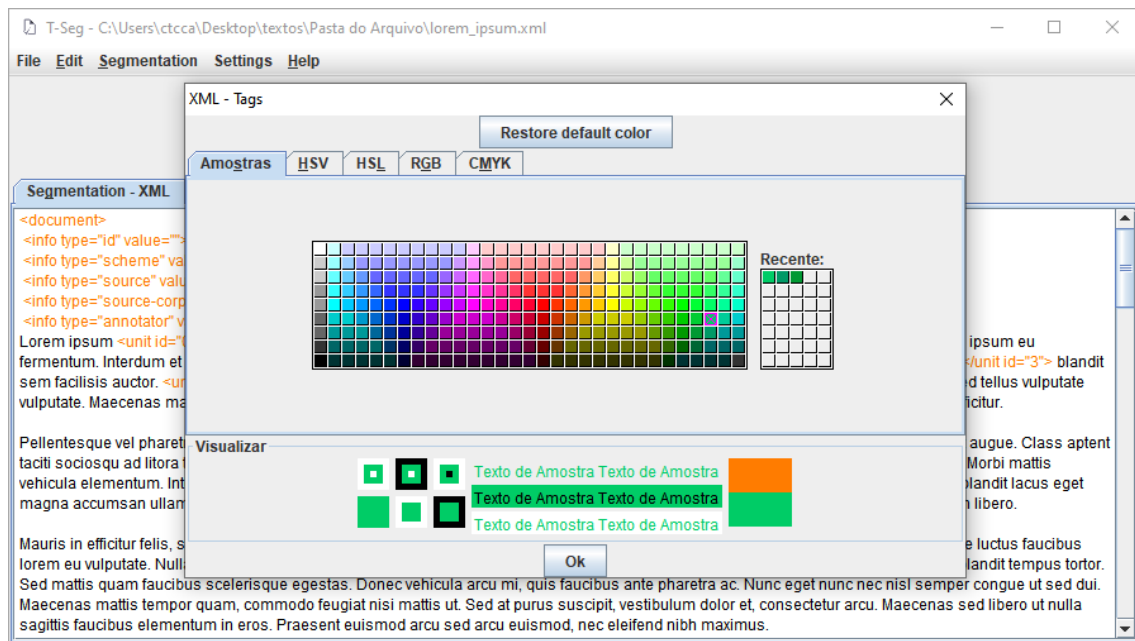
4.1 Colors of elements on “Segmentation - XML” tab

It is possible to change the colors of the elements that make up the units in the “Segmentation – XML” tab, that is, the tags that identify the units, the text inside them, and the independent unit identifier.

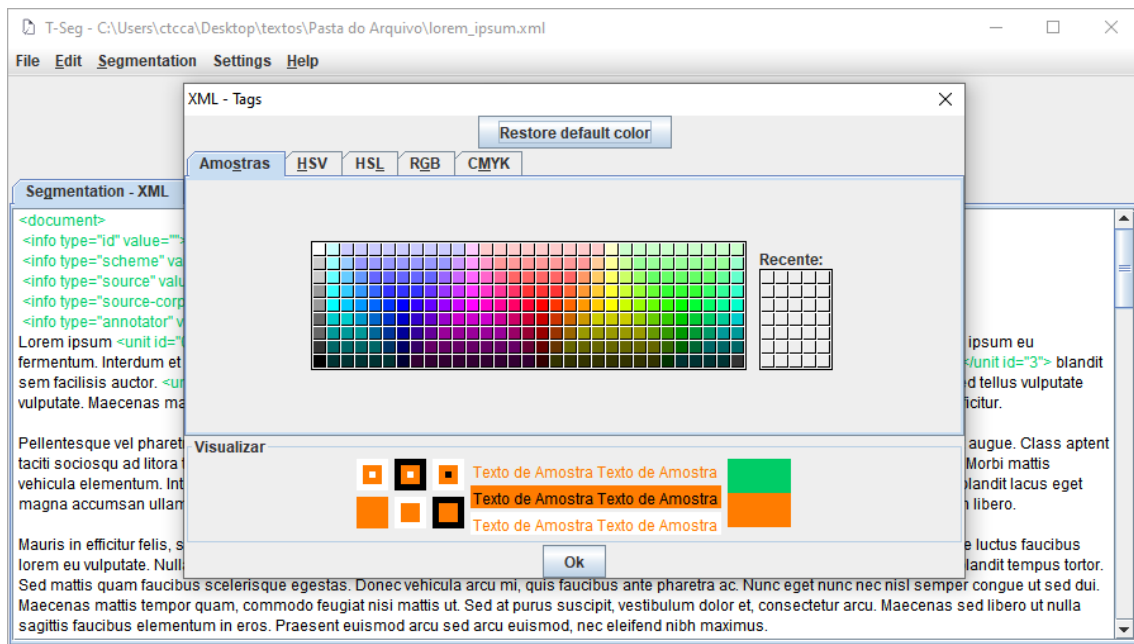
To change the color of one of these elements, just click on the "Settings" menu, place the mouse pointer on "Colors" and then on "XML" and, finally, click on the element whose color you want to change:



A window for selecting the new color will open. This window has five different color palettes: Samples, HSV, HSL, RGB and CMYK. The user simply chooses a color from one of these palettes and clicks "Ok" to apply the new color to the text elements:



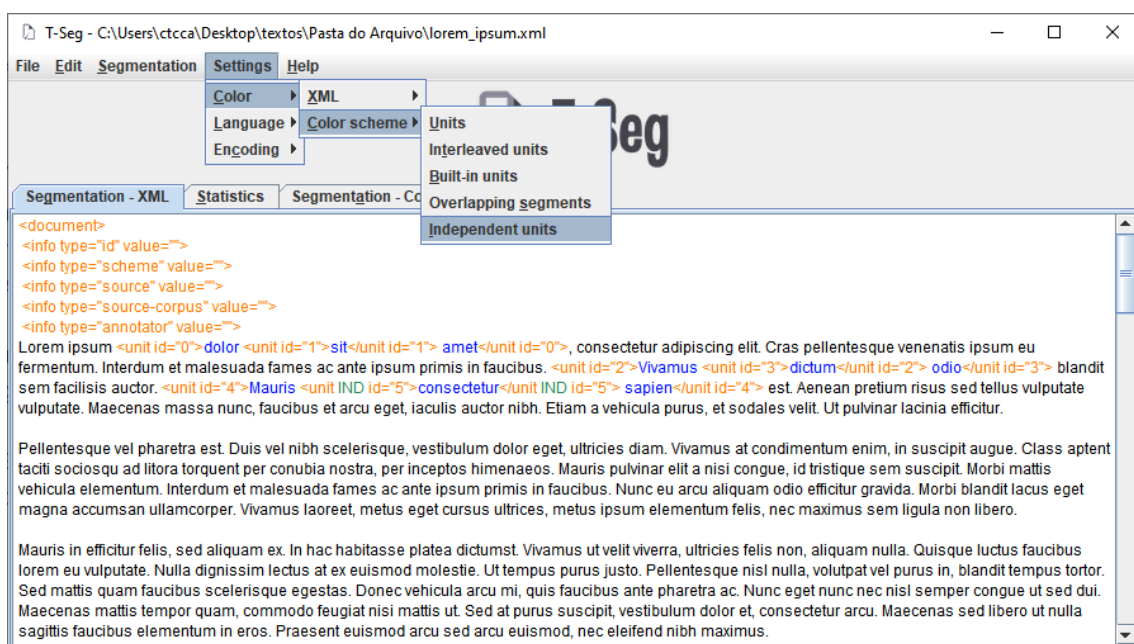
The “Restore default color” button has the effect of selecting the default color for the element, that is, orange for tags, blue for segmented texts and green for independent units. After clicking on this button, it is necessary to click on the “Ok” button for the changes to take effect.



4.2 Colors of elements on the "Segmentation - Colors" tab

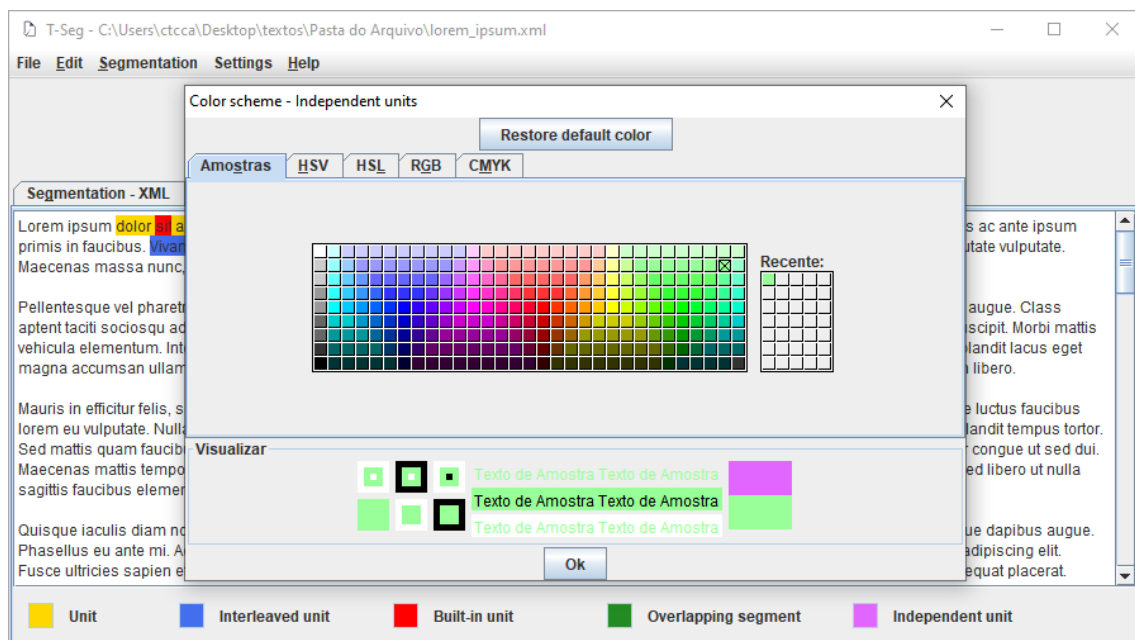
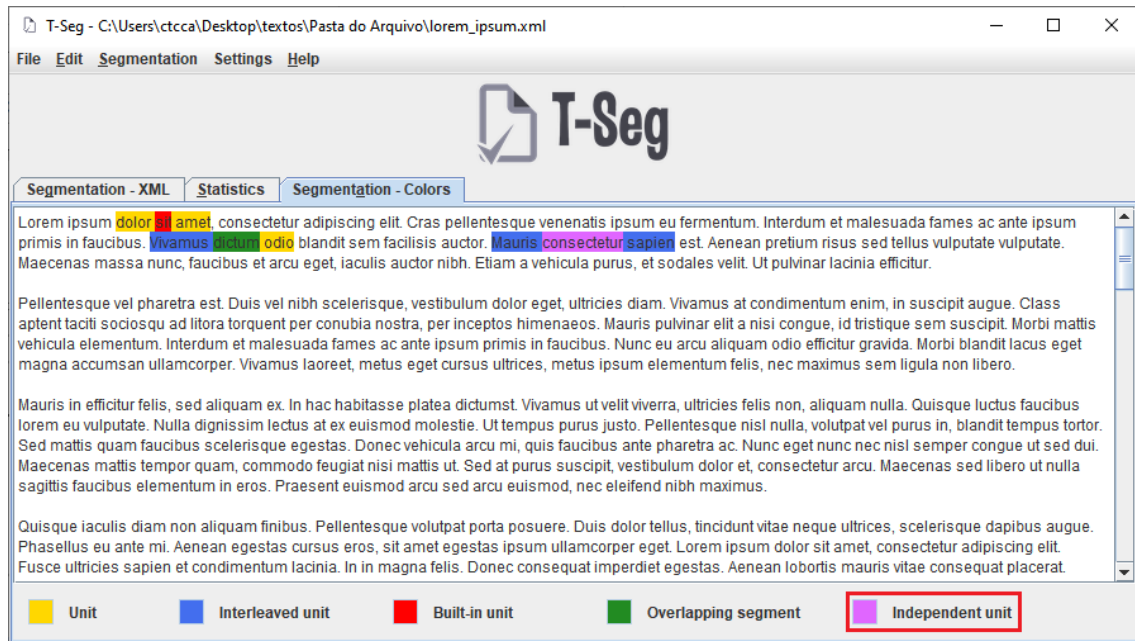
The elements of the "Segmentation - Colors" tab, similarly, can also have their colors changed in the same way as the elements of the "Segmentation - XML" tab.

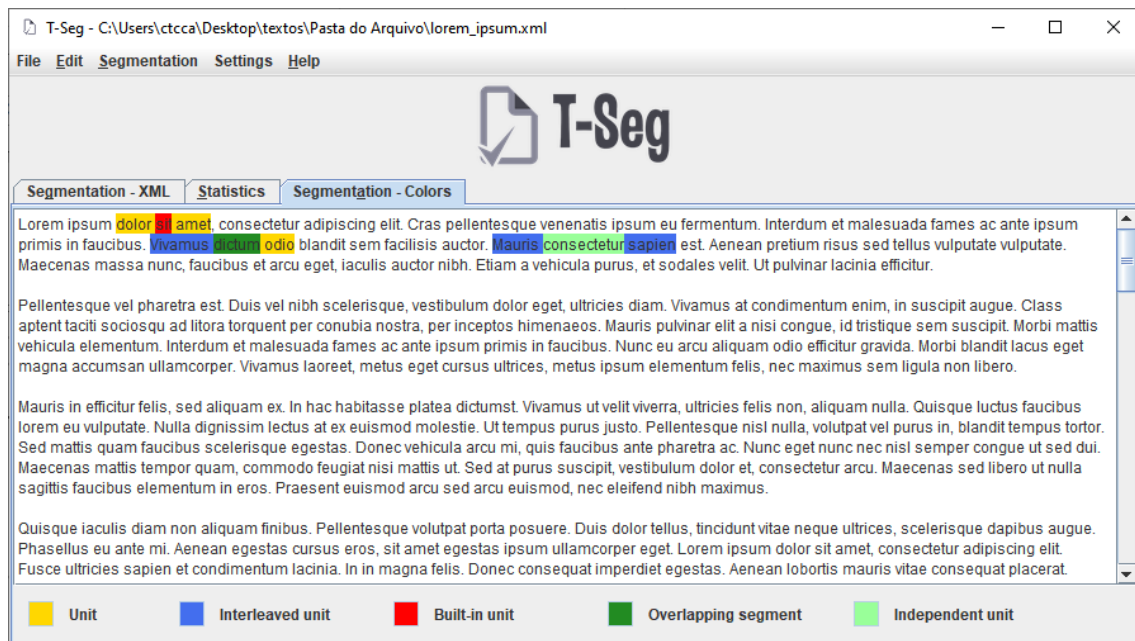
To change the color of an element on the "Segmentation - Colors" tab, just follow the same script, that is, click on the "Settings" menu, position the pointer on "Colors" and then on "Color Scheme" and , finally, select the element whose background color you want to change.



To understand the difference between each of these types of units, you can access item 3.1.

In addition to the menu, it is also possible to access the color choice window by clicking on the caption of the item that you want to change the color. For example, if you want to change the background color of independent units, you can click on the legend square or on the text “Independent Unit” to open the color selection box:





4.3 System language

T-Seg functions support both English and Portuguese languages. By default, the system always starts in the Portuguese language. To change the system language, just select the desired language, clicking on "Settings" and choosing the language in the menu that opens when you position the mouse pointer over "Language":



In this menu, the current language option will be disabled. For example: if the current language is Portuguese, it will not be possible to select the “Português” option, and if the current language is English, it will not be possible to select the “English” option.

When selecting a new language, all text in the program is automatically translated:

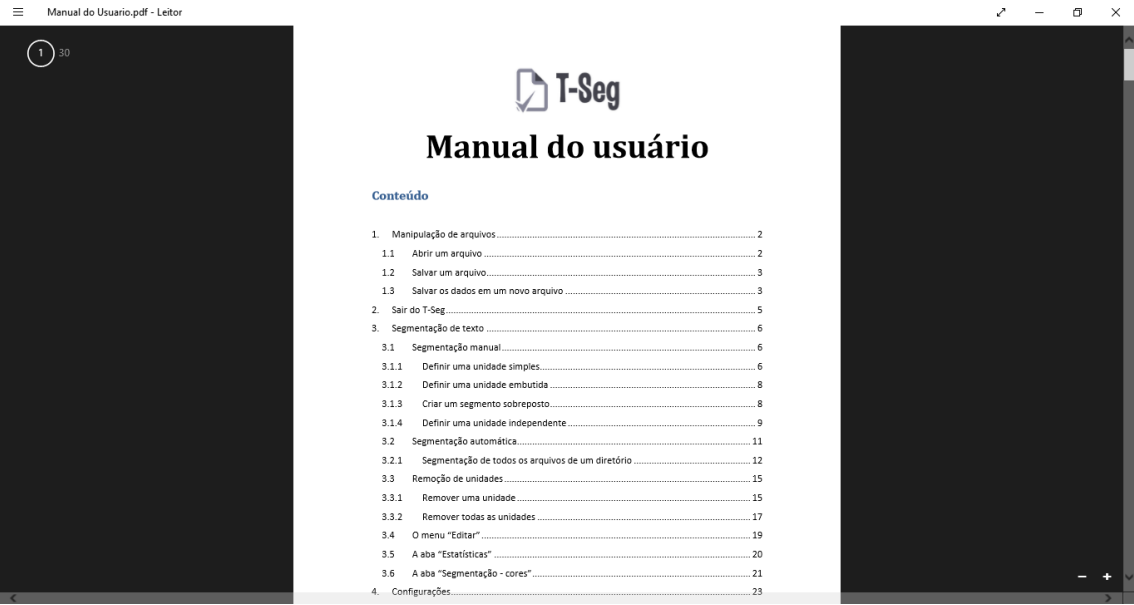


5. Help

“Help” menu contains the help topics related to the program.

5.1 User guide

The option “Manual Features” opens this guide. The same option can be also accessed through the shortcut key F1:



5.2 About T-Seg

The option "About T-Seg" displays information about the system:

