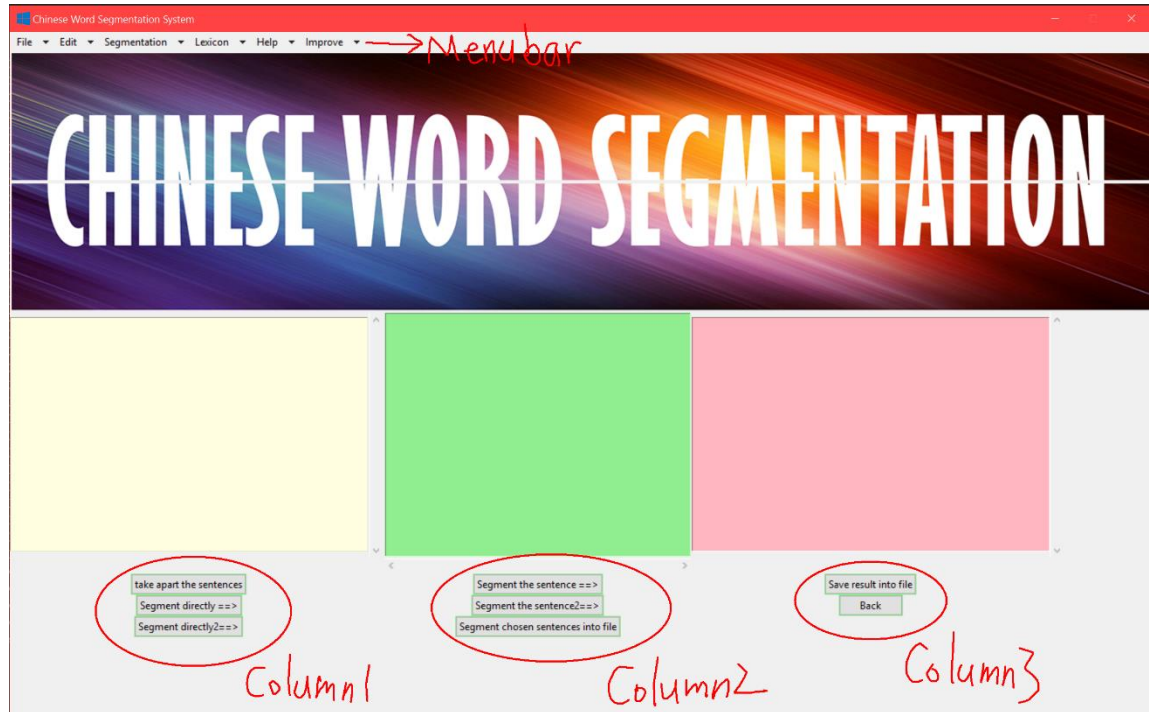


## Detailed instruction

### 1. The instructions for the main user interface



G-1.1. The main UI

The function of the buttons in the first column:

- 1) **take apart the sentence**: used to take apart the paragraph into separated sentences and show the result in the list box.
- 2) **Segment directly==>**: using the first algorithm to separate the whole passage without taking apart it.
- 3) **Segment directly2==>**: using the second algorithm to separate the whole passage without taking apart it.

The function of the button in the second column:

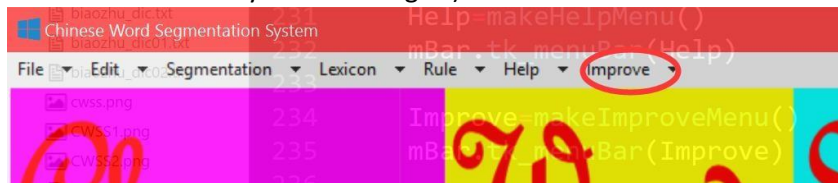
- 1) **Segment the sentence==>**: using the first algorithm to separate the chosen sentences in the list box.
- 2) **Segment the sentence2==>**: using the second algorithm to separate the chosen sentences in the list box.
- 3) **Segment and save into file**: using the first algorithm to separate the chosen sentences in the list box and save the result into a chosen file. (Why I chose the first algorithm? OK, the first is more accurate!)

The function of the button in the third column:

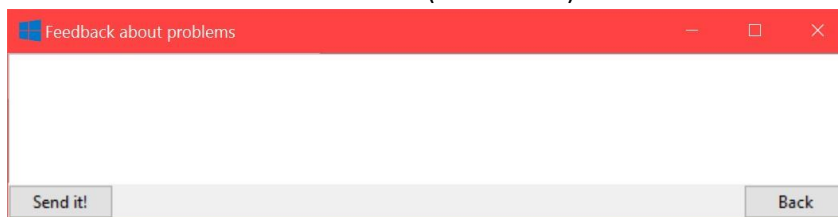
- 1) **Save result into file**: save the result into a file

2) **Back**: make the core UI hidden. And you can also just click at the menu and the back will automatically be activated.

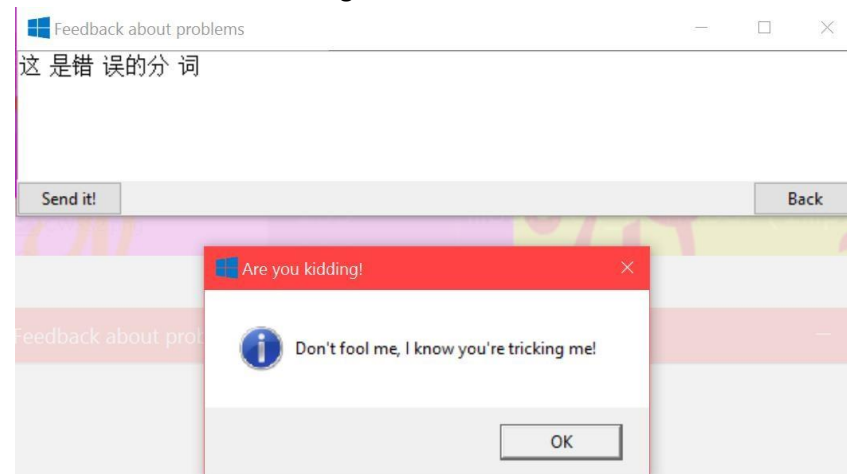
2. **The Improve function**: This can be used when you find our segmentation is not correct or not accurate. There'll be a dialogue window if you click the button, in which you can input a correct segmentation of the sentence. (Don't trick us. We have a check system to find out whether you're tricking us)



G-2.1. Where to find it (Old Edition)



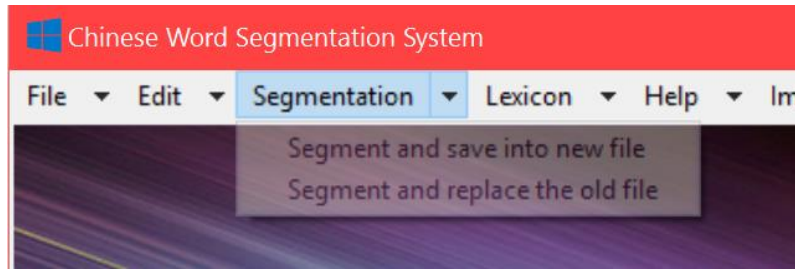
G-2.2. The dialogue window



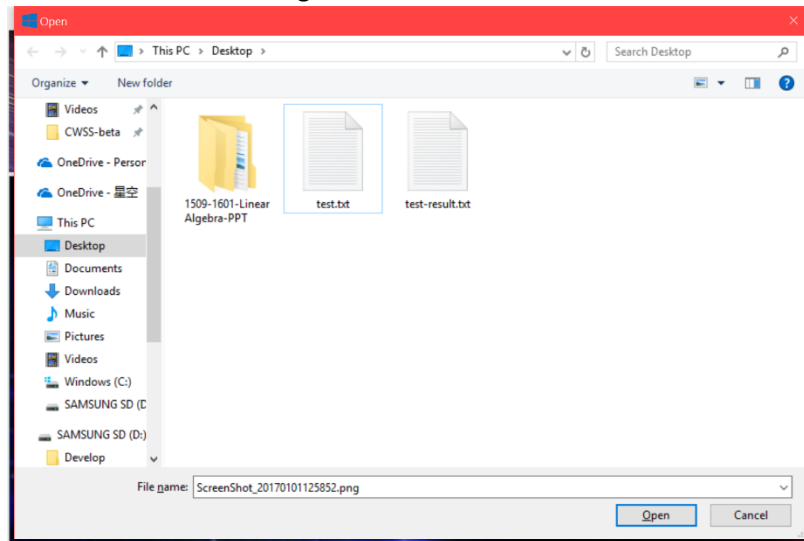
G-2.3. Don't trick us(Old Edition)

### 3. Segmentation:

- 1) **Segment and save into new file**: select a file which contains text. Then it'll be segmented and the result will be saved into a new file whose file name is the old one plus '-result'.
- 2) **Segment and replace the old file**: select a file which contains text. Then it'll be segmented and the result will be saved to the old file and cover the old text.



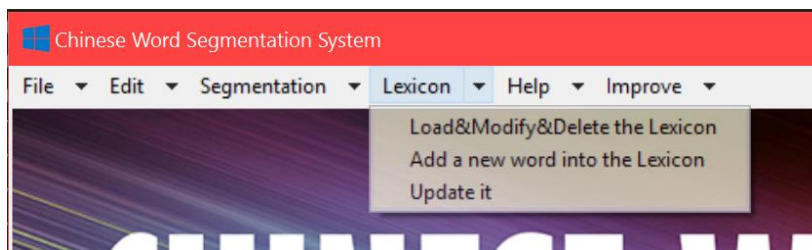
G3.1. Segmentation menu



G3.2. select a file

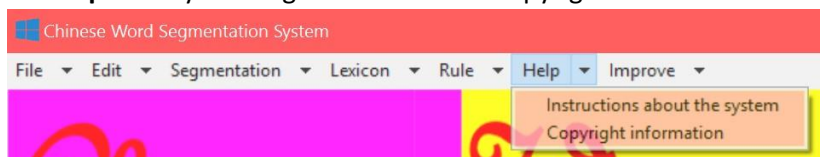
#### 4. **Lexicon:** operations about lexicon

- 1) **Load&Modify&Delete the Lexicon:** We'll open it as txt.
- 2) **Add a new word into the Lexicon:** an example will show in the information window, then just follow it.
- 3) **Update it:** update the lexicon so that the software can use the latest lexicon.



G4.1.Lexicon menu

#### 5. **Help:** Here you can get instruction or copyright information



G5.1. Help menu(old Edition)

## *Some other features*

### 6. The modules we use:

```
import threading
import multiprocessing
import pickle
import time
import __main__
from re import *
from string import *
from tkinter import *
from tkinter.filedialog import *
from tkinter.ttk import *
from tkinter.messagebox import *
from help import *
import segfunc
import segfunc2
from fileh import *
from edith import *
# from os import *
```

G6.1. Modules (1)

```
from tkinter import *
from tkinter.filedialog import *
from tkinter.ttk import *
from tkinter.messagebox import *
from segfunc import *
import os
import threading
```

G6.2. Modules (2)

**Note:** We're using multithreading to keep the fluency when CWSS is launched. There may be some bugs we haven't found.

### 7. Configuration file: Some configuration can be edited in this file, such as the title, dpi and so on

```
configuration.ini - Notepad
File Edit Format View Help
['title']='Chinese Word Segmentation System'
['dpi']='1350x810'
['xe']='+200'
['ye']='+200'
['picture1']='up.png'
['picture2']='down.png'
['picture3']='MacroHard.png'
['transparent']='1'
['topmostmain']='0'
['topmosthelp']='1'
['topmostlexicon']='1'
['gif0']='picture\\40.png'
['gif1']='picture\\80.png'
['gif2']='picture\\120.png'
['gif3']='picture\\160.png'
['gif4']='picture\\200.png'
['gif5']='picture\\240.png'
['gif6']='picture\\280.png'
['gif7']='picture\\320.png'
['gif8']='picture\\360.png'
['gif9']='picture\\400.png'
['gif10']='picture\\440.png'
['gif11']='picture\\480.png'
['main2transparent']='0'
['main2topmost']='1'
```

#### G7.1. configuration file

### 8. Log file: It'll record the important process of the program, including unexpected error

```
CWSS.log - Notepad
File Edit Format View Help
[2017-01-01 13:19:19]Loading configuration OK
[2017-01-01 13:19:19]CWSS starts
[2017-01-01 13:19:20]Loading UI OK
[2017-01-01 13:19:20]Loading matrices
[2017-01-01 13:19:20]Qi's algorithm begins to load
[2017-01-01 13:19:20]Shi's algorithm begins to load
[2017-01-01 13:19:21]Copyright information is loaded
Time cost 2.308
[2017-01-01 13:19:25]Shi's algorithm loads OK
Time cost 5.318
Error: Illegal characters occur in the text to be segmented!
[2017-01-01 13:19:57]CWSS exits
```

#### G8.1. An example of log file

### 9. The source code is on the GitHub:

<https://github.com/yyong119/Chinese-Words-Segmentation-System>