The UCLouvain Error Editor
User guide - Version 2.0

Sylviane Granger, Helen Swallow and Jennifer Thewissen

CECL
Centre for English
Corpus Linguistics

UCLouvain

2023

The UCLouvain Error Editor
User guide - Version 2.0

Sylviane Granger, Helen Swallow and Jennifer Thewissen

CECL
Centre for English
Corpus Linguistics

UCLouvain

2023

The UCLouvain Error Editor
User guide - Version 2.0 – 2023

Sylviane Granger, Helen Swallow & Jennifer Thewissen
Centre for English Corpus Linguistics, UCLouvain

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

The present document is the second version of the UCLouvain Error Editor (UCLEE) user guide. The software, which was developed at the Centre for English Corpus Linguistics (Université catholique de Louvain, Belgium), can be used with both raw and previously error-tagged texts and speeds up the insertion of error codes and corrections in learner texts. It also contains a concordancing function and an exercise generator.

1.1. Installation

Run the UCLEE.exe file in Windows, the UCLEE.dmg file in MacOS or the UCLEE file in Linux. This will open the "Tagset selection" window described in the following section.

1.2. Tagset selection

By default this version uses the Louvain Error Tagging Manual Version 2.0 (Granger, Swallow & Thewissen 2022[1]). It also contains a built-in converter which allows users to work on texts tagged with the previous version of UCLEE (version 1.3[2]). Alternatively, users can design their own tagset and use it to tag their data.

The "Tagset selection" window provides three tagset options (Figure 1):
1. Select uclee-en-2.0.tag if you wish to work with the new error typology described in Granger, Swallow & Thewissen (2022);
2. Select uclee-en-1.3.tag if you want to open texts previously tagged with the 1.3 version of the error annotation system;
3. Upload a new tagset: user-designed tagset.

Further guidance on Option 2 is given in **Appendix 1**.

---

[1] Granger, S., Swallow, H. & Thewissen, J. (2022). *Error tagging manual. Version 2.0.* CECL Papers 4. Louvain-la-Neuve: Centre for English Corpus Linguistic/Université catholique de Louvain. The manual is open access and can be downloaded from https://cdn.uclouvain.be/groups/cms-editors-cecl/Granger%20et%20al._Error%20tagging%20manual_v2.0_2022.pdf

[2] Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J., Thewissen, J. (2008). *The Louvain Error Tagging Manual Version 1.3.* Centre for English Corpus Linguistics, Université catholique de Louvain, Louvain-la-Neuve

> It is very important that you read and implement the guidance set out in Appendix 1 before starting to work on texts previously error-tagged with version 1.3.

Whichever tagset you choose to use, the tagged files will automatically be given an ERA file extension (e.g. ICLEFR001.era).
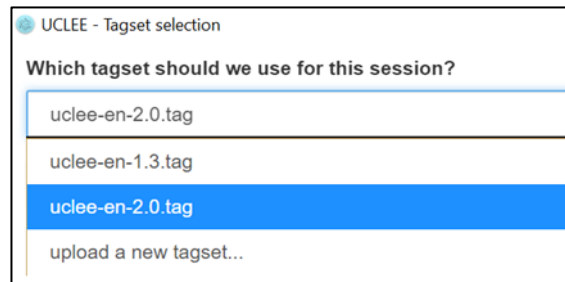

Figure 1: Tagset selection

Once you have selected a tagset, simply click on the "Open UCLEE" button at bottom right of the window. This will redirect you to the software's main interface (see Figure 2).

Note: If you want to select a different tagset, you need to close and reopen UCLEE.
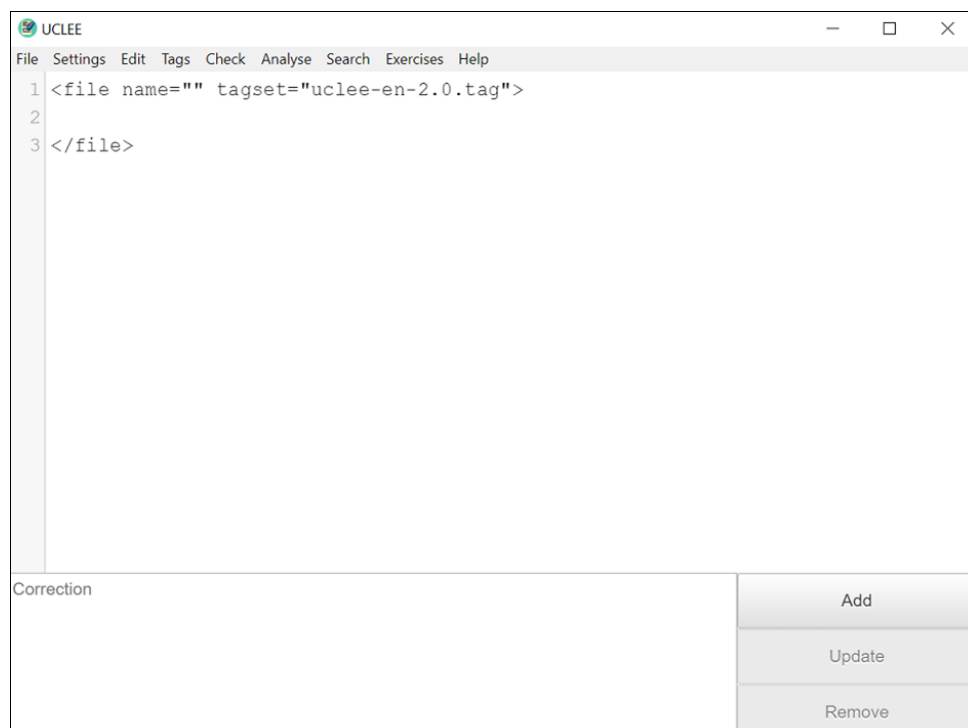

Figure 2: Main interface

1.3. Modification of tagsets

Existing tagsets can be modified, i.e. researchers can choose to add new tags, delete unwanted tags or modify existing tags to create a tagset better suited to their research purposes.

Tagsets can be accessed by clicking on the "Show tagsets" button in the opening "Tagset selection" box, or in Settings, Tags and Colors. To modify a tagset, open the corresponding *.tag file (e.g. uclee-en-2.0.tag) with a text editor such as Notepad or Notepad++ or any default text editor, and update the file in line with your preferences, paying particular attention to file structure. Save the new tagset with a new name so as to be able to return to the original tagset if you want to. As shown in Figure 3, the hierarchical structure of the tagset (e.g. the fact that the two error tags FS and FM belong to the broader category denoted by F) is represented by **[]** and **indents**:

- A broad category is enclosed in [] and headed with an error tag that directly follows [.
- Indents indicate sub-categories.
- There should be only one tag per line.

Sub-categories can be named as desired and do not have to start the same way as their parent. **Be careful not to use the same tag name twice, even in different categories.**

Warning! In order to apply changes and use the modified tagset, you must restart the software.

```
[F
    FM
    FS
]
[G
    GA
    [GADJ
        GADJCS
        GADJN
        GADJO
    ]
    GADVO
    [GD
        GDD
        GDI
        GDO
        GDT
    ]
    [GN
        GNC
        GNN
    ]
    [GP
        GPD
        GPF
        GPI
        GPO
        GPP
        GPR
        GPU
    ]
```

Figure 3: The hierarchical structure of a UCLEE-conformant tagset

1.4. Adding new tagsets

You can also add new tagsets to UCLEE. To create a new tagset, open a new text file and structure your tagset with [] and indents as explained above. Give your new file a name and save it in *.tag format (manually change .txt to tag when you save the file) in UCLEE's *resources/tagsets* repertory. All tagsets stored in this repertory will be made available to you when the program is launched.

Another way of creating a new tagset is to use one of the existing tagsets as a template, having saved it under a new name before introducing your changes; this means that you can use the existing structure as a basis rather than creating the new one from scratch.

To access the existing tagsets in order to use it as a template, go to UCLEE's *resources/tagsets* repertory (accessible by clicking on the "Show tagsets" button in the opening "Tagset selection" box).

The new tagset you have created will appear in the drop-down menu in the opening "Tagset selection" window once you closed and reopened the UCLEE software.

2. MENU BAR

2.1. File

Once redirected to the main interface, you can upload two types of text: either a raw (.txt) or a previously tagged (.era) text file by selecting the "Open" option in the "File" menu.
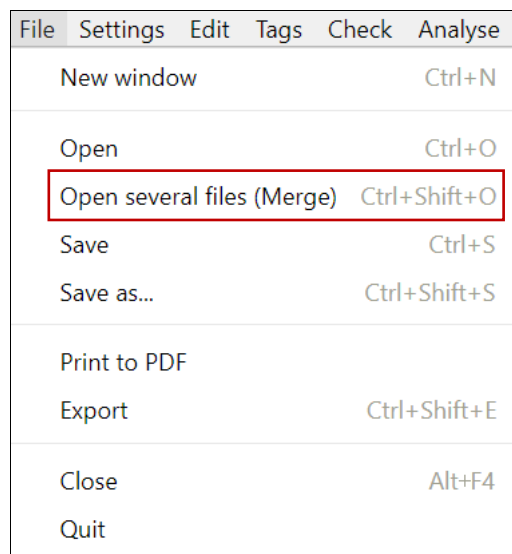

Figure 4: Open several files (Merge) option

Multiple files can be uploaded by clicking on the "Open several files (Merge)" option (Figure 4). This will merge the uploaded documents into one. Further files can only be added subsequently by performing a new upload.

Note: If you do *not* want to merge your files, select the "New window" option; this will open a new workspace in which you can work on texts separately.

Note. *.txt files should preferably be encoded in UTF-8 format.
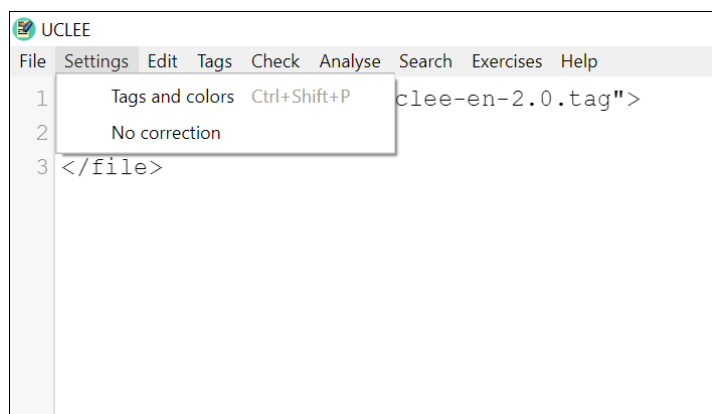
2.2. Settings

Figure 5: Settings

2.2.1. "Tags and colors"

This submenu allows you to view your current tagset and customise the colours you are working with. Clicking on this option opens a box with two tabs: "Tags" and "Colors".

- The "Tags" tab contains the tag hierarchy. This is the list that is used to generate the menu-driven tag insertion function; it lists the tags that will be recognised by the parser. (For instructions on modifying tagsets or creating new ones, see 1.3 Modification of tagsets.)

- The "Colors" tab specifies the colours used by the tag-highlighting system in the "Check" and "Exercises" modes. Each line must be in the "T: RRGGBB" format, in which **T** corresponds to the tag that will be coloured in this way. **RRGGBB** corresponds to a certain colour in the RGB hexadecimal format (see http://www.javascripter.net/faq/colornam.htm for examples) The line enclosed by "_" defines the default colour for the tagset.

2.2.2. No correction

If you don't want to insert any corrections in the texts you are error-tagging, you can apply the "No correction" setting (see Figure 5). This will result in the following tag format: 'it's my great <FS>presure</FS> to share my college life with you.'

Do not use this setting if you only wish to omit corrections in the case of certain specific errors (see Section 2.4).

2.3. Edit

Clicking on "Edit" opens the editing menu – Copy/Cut/Paste, Select All, etc. (Figure 6). Each function has a corresponding keyboard shortcut that is displayed next to it, e.g. <CTRL+F> (or <cmd ⌘+F> on Mac) activates the Find function, which allows you to find specific words in the text.
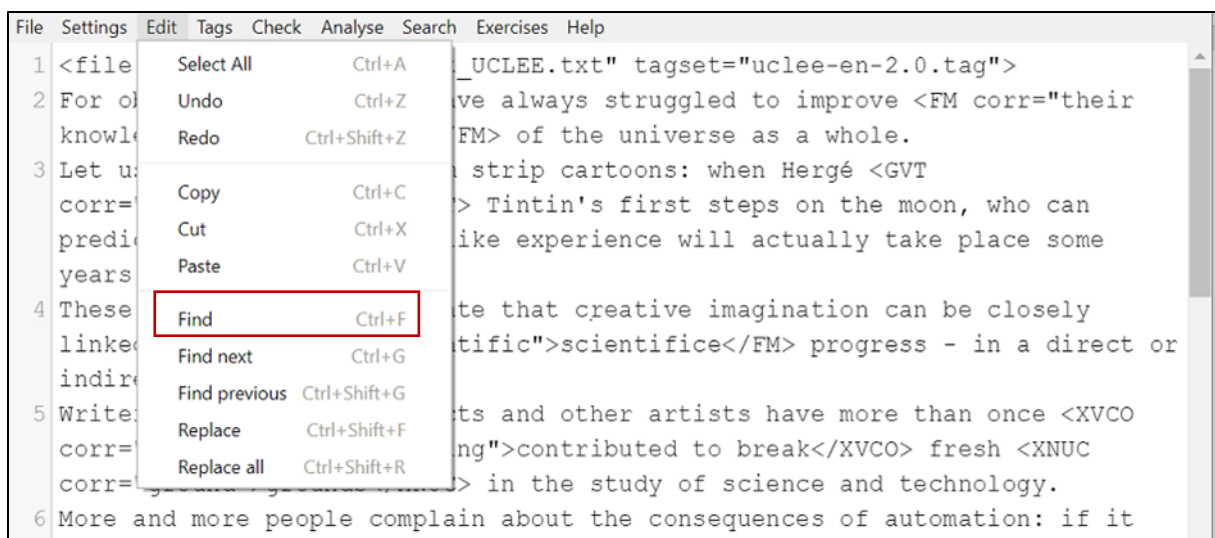

Figure 6: "Edit" menu functionalities

2.4. Tags (Error tagging)

This functionality allows you to insert, update or remove error tags in your texts. Proceed as follows to tag detected errors (Figure 7):

1. Highlight the erroneous word/word string in the learner text;
2. Insert the correction(s) in the box situated at the bottom of the window;
3. Select the tag that applies to the error. This can be done in one of two ways:
    a. Go to the 'Tag' menu at the top of the interface and select the appropriate tag or
    b. Click on the 'Add' button at the bottom right of the dialogue box.

The error taxonomy is described in the Louvain Error Tagging Manual, which can be found by clicking on "Help" in the menu bar (cf. Section 2.9).
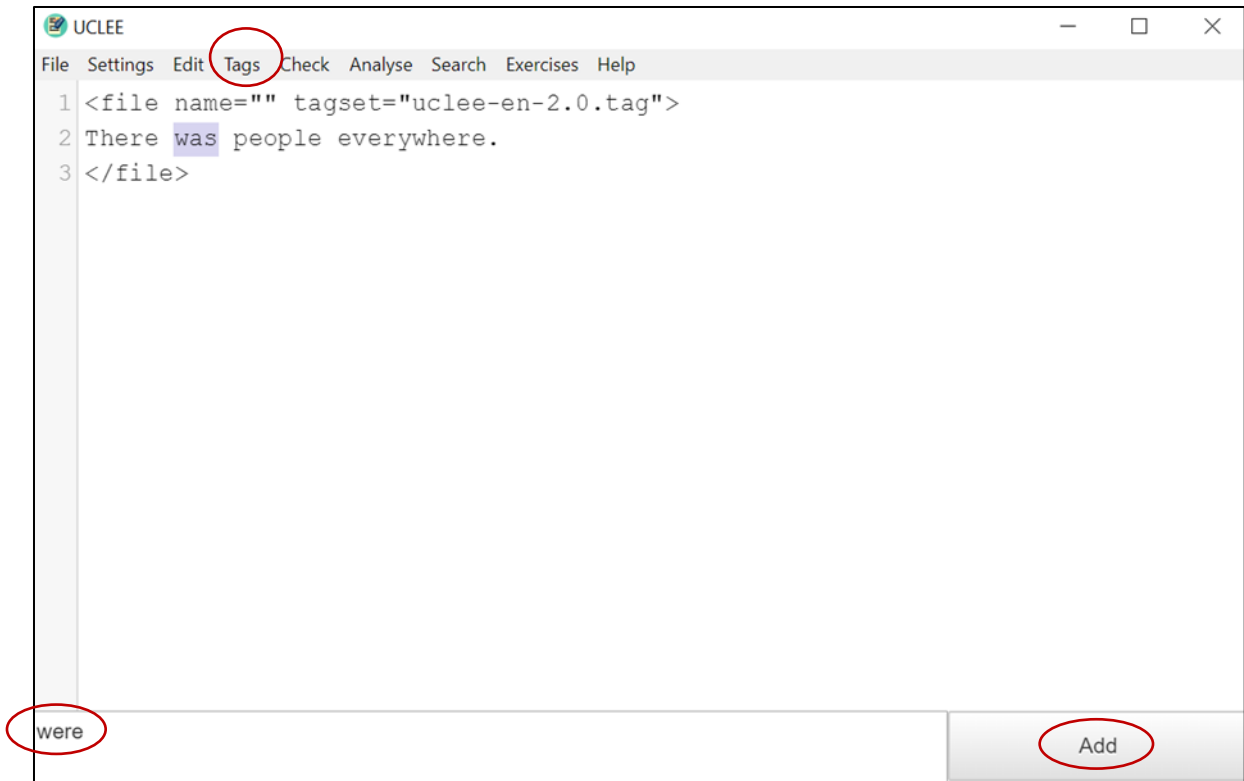
Figure 7: Highlighting erroneous words/strings of words in the text

In the example above (Figure 7), the concord error *was* is corrected to *were* and tagged as GVN (Grammar Verb Number). The tag assigned by the user appears on the screen with the markup illustrated in Figure 8.
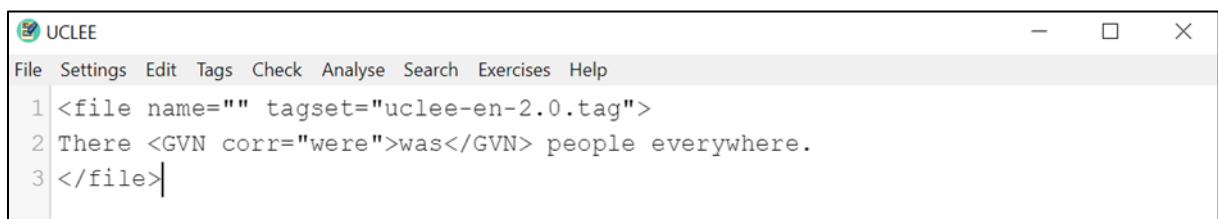

Figure 8: Error code format

In some situations you may need to insert a tag to mark a missing word. To do this, place the cursor at the omission point, enter the missing word(s) in the Correction field and insert the appropriate tag (cf. first sentence in Figure 9).

Conversely, to mark a word or words as superfluous, select it/them, enter **\0** (backslash followed by 0) in the Correction field and insert the appropriate tag (cf. second sentence in Figure 9).
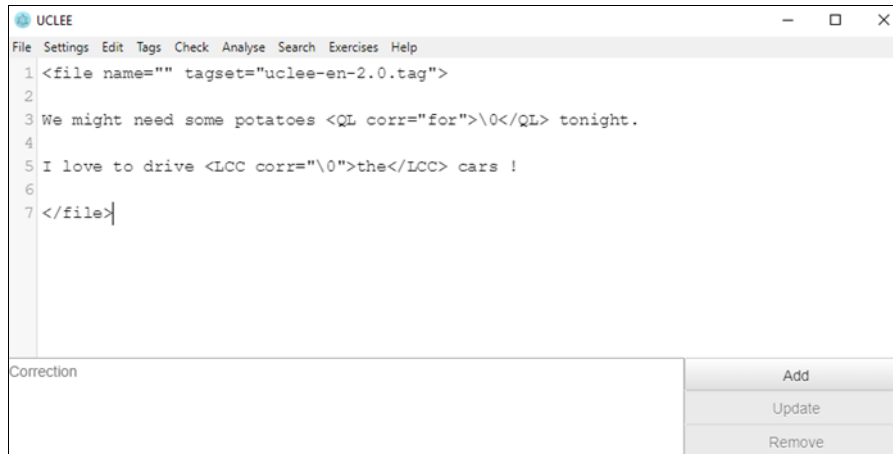

Figure 9: Error tagging missing and superfluous words

If you only want to tag an error without providing a correction, you should select the appropriate tag and leave the correction box empty. This will ensure that the attribute corr="" is automatically inserted beside the error tag (see Figure 10).

Note: If you don't want to add any corrections at all in the texts you are error-tagging, select the 'no correction' option in Settings (see Section 2.2.2).
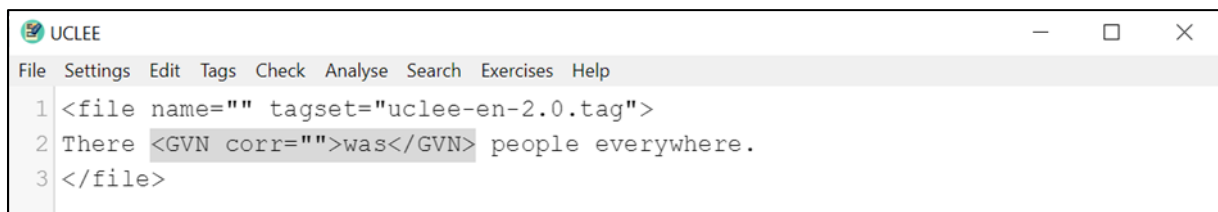

Figure 10: Error tag with no correction

2.5. Check

This menu contains two functionalities: "Check all" and "Check only" (Figure 11). The "Check all" option gives a broad overview of the corrected error(s); it can be used to check whether you have been consistent in your error tagging system throughout a given text. "Check only" allows you to review a limited number of tags by selecting them from a scroll-down list.
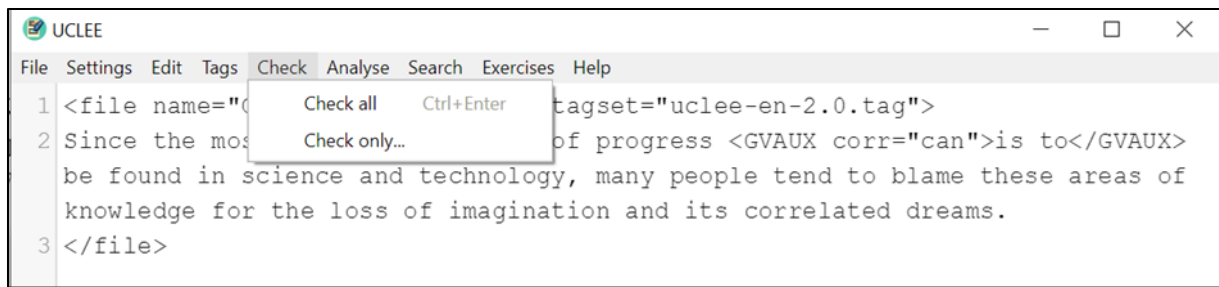
Figure 11: "Check" menu

2.5.1. "Check all"

You can review your corrections at any time using the "Check all" function in the "Check" menu (Figure 12). When you click on "Check all", all the tags used and corrections made are displayed in colour-coded boxes (one colour per error category). To return to the raw text view, deselect the "Check all" option in the "Check" menu.
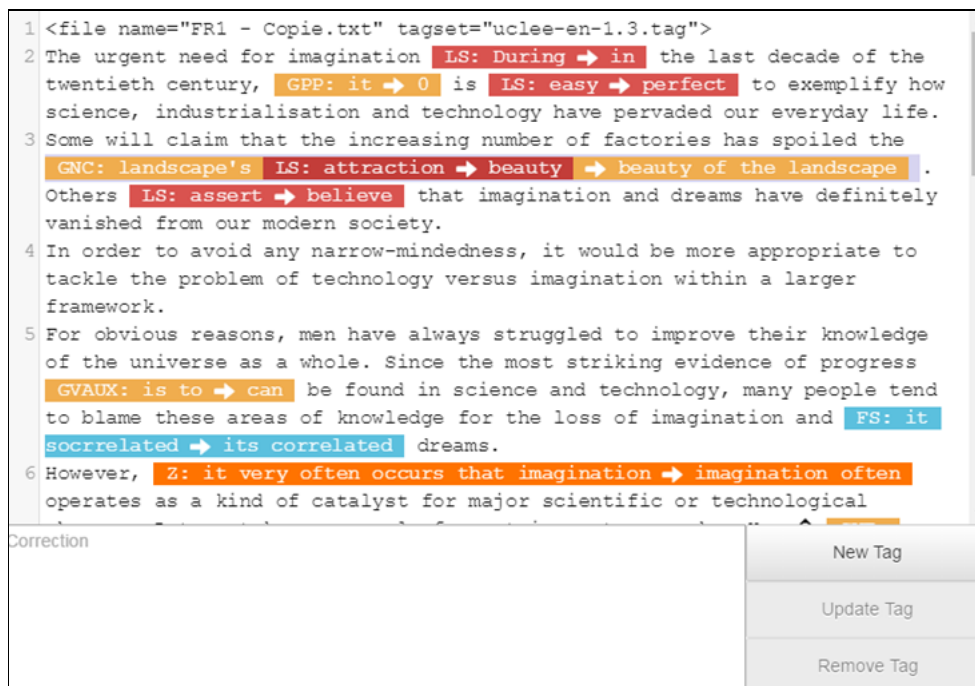

Figure 12: Error tagging in the "Check all" display

To change the error tag on a given word/string of words (for example, in Figure 13Figure 13 the user wants to replace the GVM tag with GVAUX), click on the coloured box you want to update and then on the "Update tag" button. You can now select a new error tag.
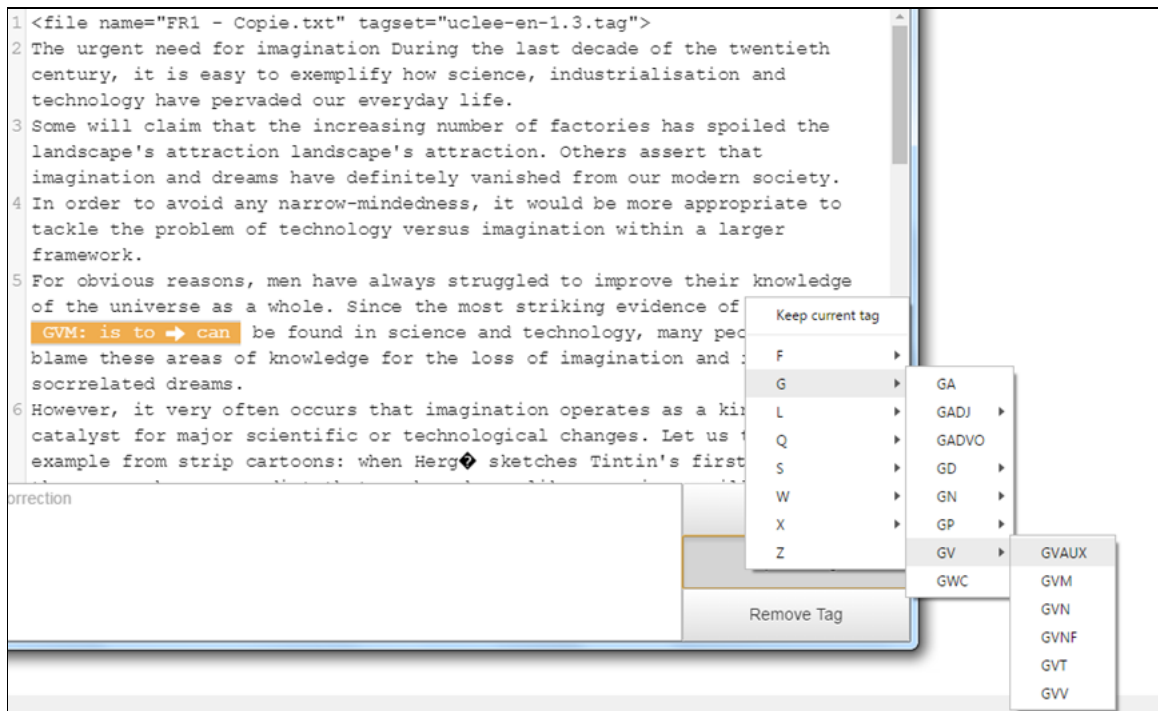
Figure 13: "Update tag" option

To update a correction while keeping its tag, click on the box with the correction you want to change, click on Update at bottom right, and select the "Only update correction" option that you will see above the list of tags. (Figure 14). Insert the new correction in the correction box and click on "enter". This allows you to replace the correction you previously entered (*scientifical*) with a new one (*scientific*).

Figure 14: "Only update correction" option

To remove a tag and the corresponding correction, select the relevant box and click on the "Remove" button at bottom right.

2.5.2. Review specific tags ("Check only")

To review only certain tags, select them in the scroll-down list in the "Check only" submenu. Multiple tags can be reviewed at the same time by holding the <CTRL> (or <cmd ⌘> in Mac) key down while (de)selecting the relevant tags (Figure 15).
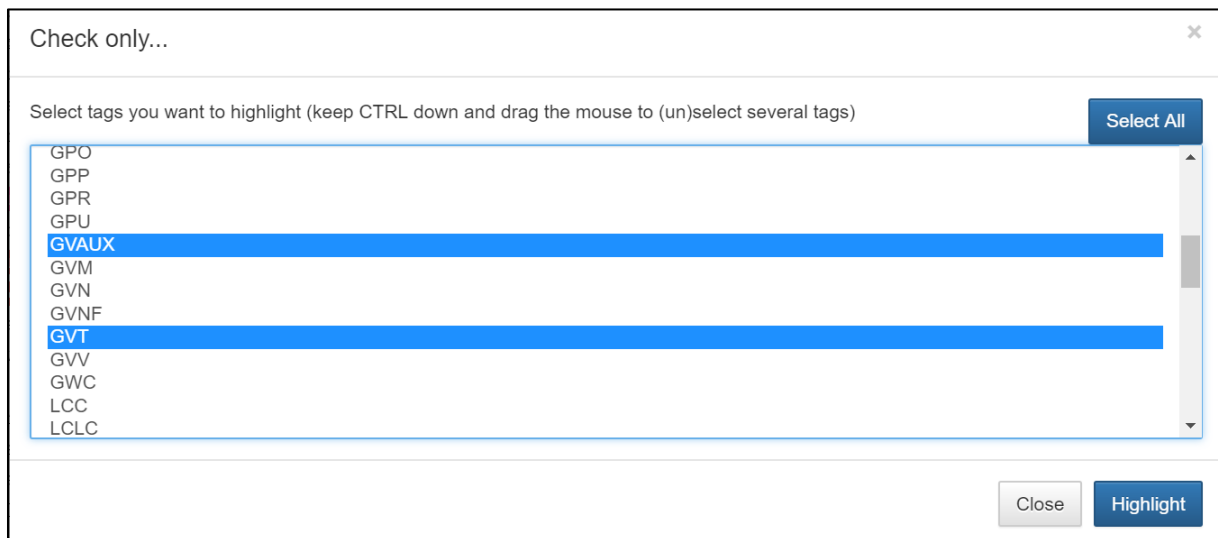

Figure 15: Review specific tags only

Once the specific tag(s) have been selected (e.g. LS, see Figure 16), the main window will foreground the selected tag(s), while greying out the others.

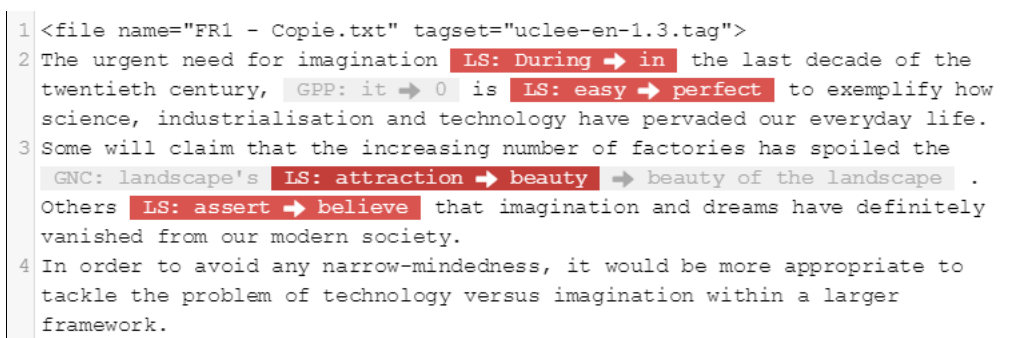*Figure 16: "Check only" option after the selection of a specific tag*

2.5.3. Nested error tags

To correct more than one error in the same string of words, repeat the steps detailed above, following the order of your own thought processes. For example, let us imagine that you want to correct the erroneous string of words "landscape's attraction" (see Figure 19). You first notice that "beauty" would be a better choice than "attraction" and therefore highlight the word "attraction", adding a new tag (in this case LS), and the correction "beauty" by following the steps described in Section 2.4.

Then you realise that "beauty of the landscape" would be a better formulation than "landscape's beauty". Select the same string of words (so 'landscape's beauty') (as if there were no previous correction), add the correction in the correction box and select a new tag (here GNC). This will produce the result shown in Figure 17, where the red tag (LS) is nested inside the orange tag (GNC).
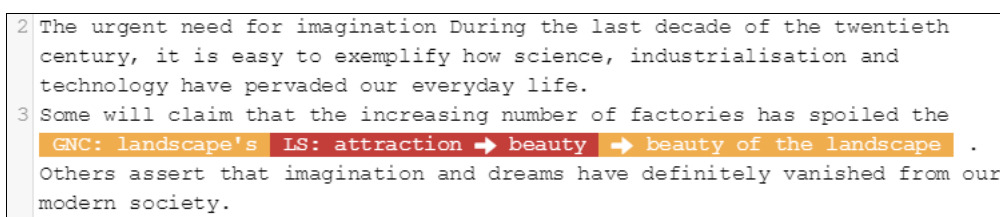

Figure 17: Nested error tags

2.6. Analyse

The built-in concordancer in the "Analyse" tool allows you to count, sort and analyse errors, together with their corrections. It also offers the possibility of drawing up an error profile for specific users or user groups. To open the concordancer, click on the "Analyse" menu (Figure 18).

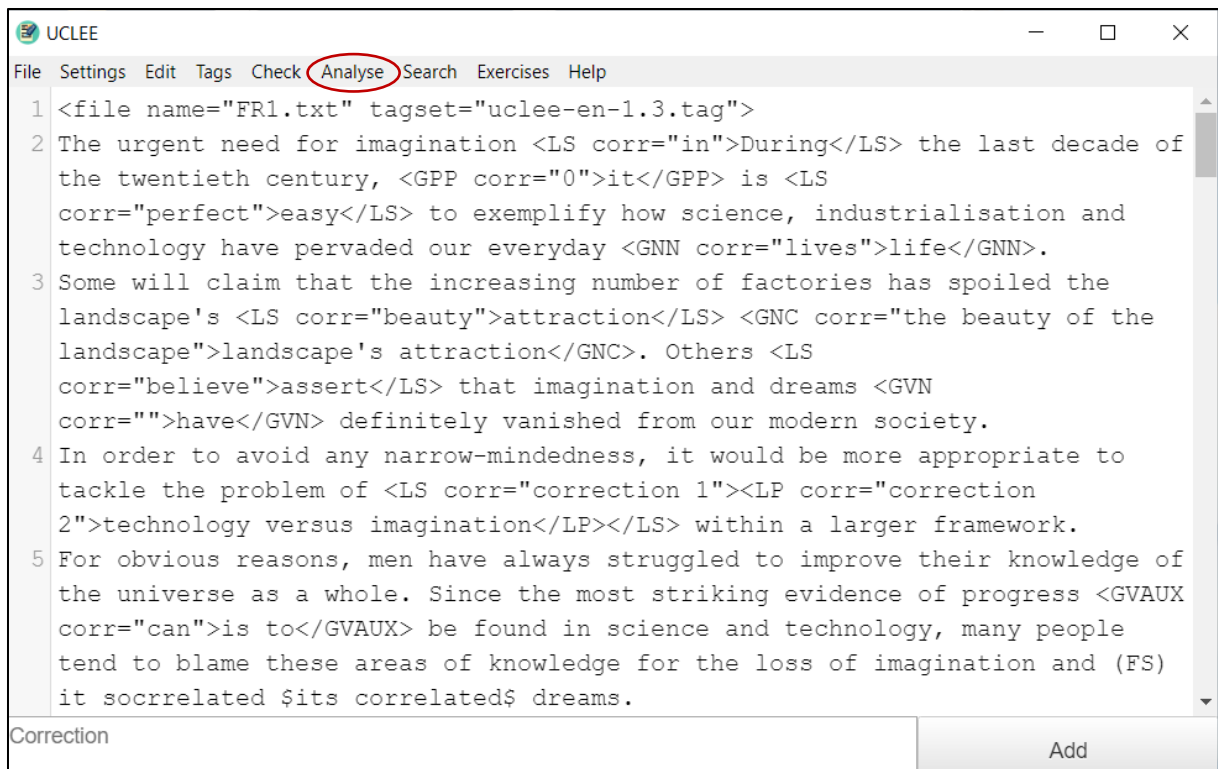**Note:** The "Analyse" function only works with error-tagged texts (i.e. files with the .era extension).

Figure 18: "Analyse" menu

Clicking on "Analyse" opens a separate window with a concordance tool (see Figure 19).
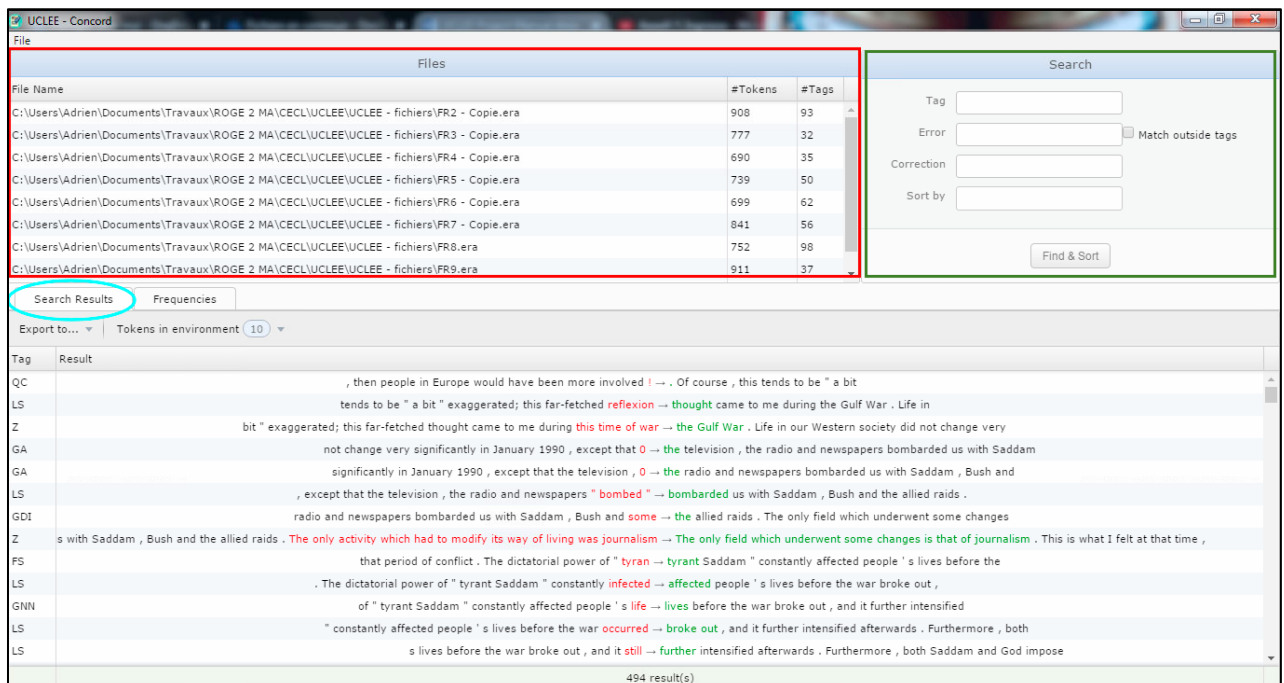

Figure 19: Overview of the UCLEE concordance tool

"Files" (red box, top left) displays the names of the .era files that have been uploaded, with their respective numbers of tokens and tags.

"Search" (green box, top right) allows you to launch a linguistic query by entering a tag, an error and/or a correction and to filter the query results. It is composed of four fields (Tag, Error, Correction, Sort by), three of which are *matchers* (Tag, Error, Correction). *Matchers* are currently not case-sensitive. Queries are performed by entering a search target into one matcher (or more) and clicking on "Find & Sort". Queries in the Error field can be extended to non-tagged instances by checking "Include correct uses". An empty *matcher* field means that the field will accept all results. If all three *matcher* fields are empty, the concordancer will retrieve all error tag instances. The wildcard symbol (*) can be used at the beginning or end of a word to specify that the word can be followed or preceded by anything (e.g. *the * of*). A *matcher* can contain several words or tags separated by spaces (e.g. GVAUX GVT). The "Sort by" field defines the order in which the search results are reported in the concordancer. There are three possible sorting types (tag, error, correction), which can be combined.

It is possible to combine multiple search criteria. These are a few examples:
- you can search for all cases of missing personal pronouns by entering "\0" in the Error box and "GPP" in the Tag box;
- the wildcard (*), representing a sequence of zero or more characters, may be used to group multiple Tags, Errors or Corrections. For example, querying "un*" in the Error box will yield all instances of error-tagged text that includes a word beginning with "un";
- you can search for all tags denoting grammatical verb errors by querying "GV*" in the Tag box.

The concordance lines (Figure 19, blue circle) are sorted by order of appearance in the selected file(s). Concordance lines can be exported to PDF or CSV[3] formats by using the "Export to" option situated on the left-hand side of the window (see Figure 20). You can adapt the length of the concordance lines displayed on the screen (ranging from 10 to 50 tokens) by selecting the desired number in the "Tokens in environment" scroll-down menu ("Search Results" tab).

---

[3] The CSV format is better suited to subsequent data processing in Microsoft Excel.
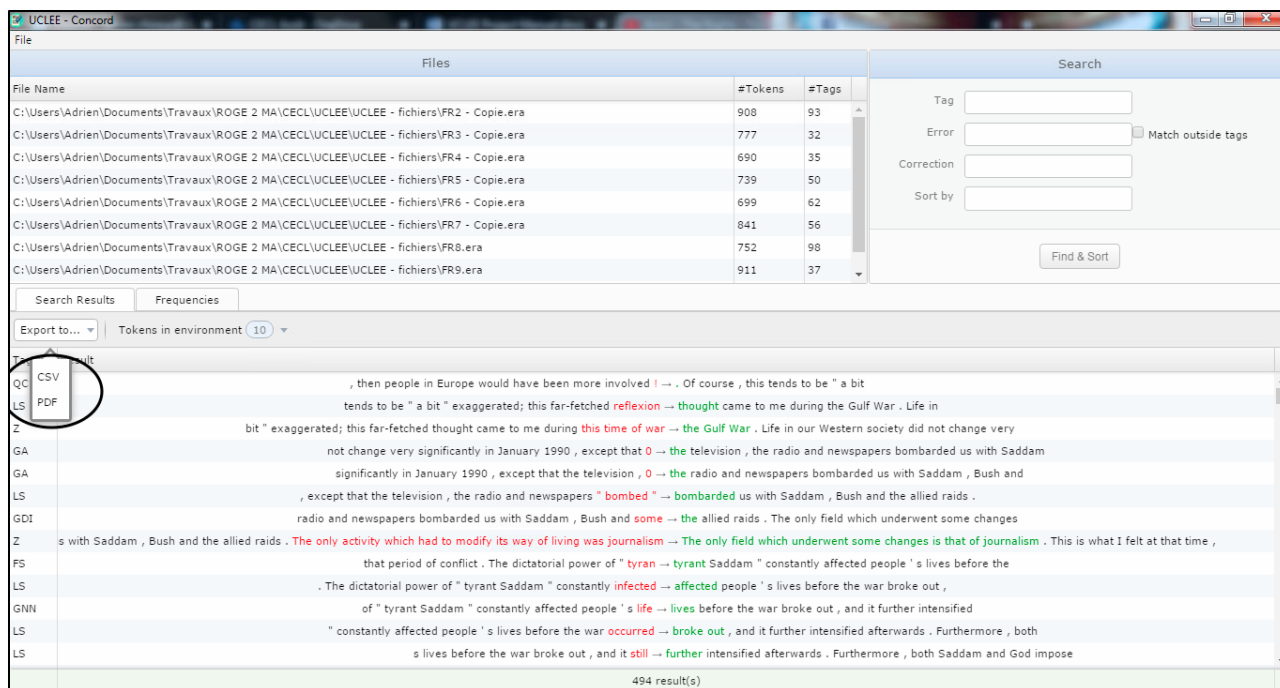
Figure 20: Exporting the concordance lines

The "Frequencies" tab ([Figure 21](#)) displays the following results:

1. **"Absolute frequencies"**

This option indicates the raw number of errors sorted by tag type. For example, the first file on the list (indicated in the first column of the window) contains 93 error tags (see second column). Among these 93 error tags there are four FS errors (fourth column), six GA errors (fifth column), one GADJN error (sixth column), etc. If the grid is read vertically, the results found at the very bottom of the page indicate that if all the sub-corpora are put together, the FM error tag represents 1% of all the errors found in the entire corpus, with a total of three occurrences. Similarly, the FS tag represents 7% of all the error tags in the corpus, with 35 occurrences.

Note: For all error types to appear, the search boxes need to be empty.

Figure 21: Absolute frequencies in the concordance tool

2. "% by tag"

The results displayed in this sub-header should be read vertically. They indicate the relative distribution of individual error tags across corpus files (with 100% representing the whole corpus). For example, Figure 22 indicates that 33.3% of the FM error tags are found in the fourth file, while the remaining 66.7% are in the seventh file.

Figure 22: Detailed percentages of errors sorted by tag

3. "% by file"

The results displayed in the option "% by file" ([Figure 23](#)) should be read horizontally. They indicate the relative distribution of error tags in a single corpus file (as opposed to the previous "% by tag" option). In this case, 100% represents the total number of errors in an individual file (and not the entire corpus).
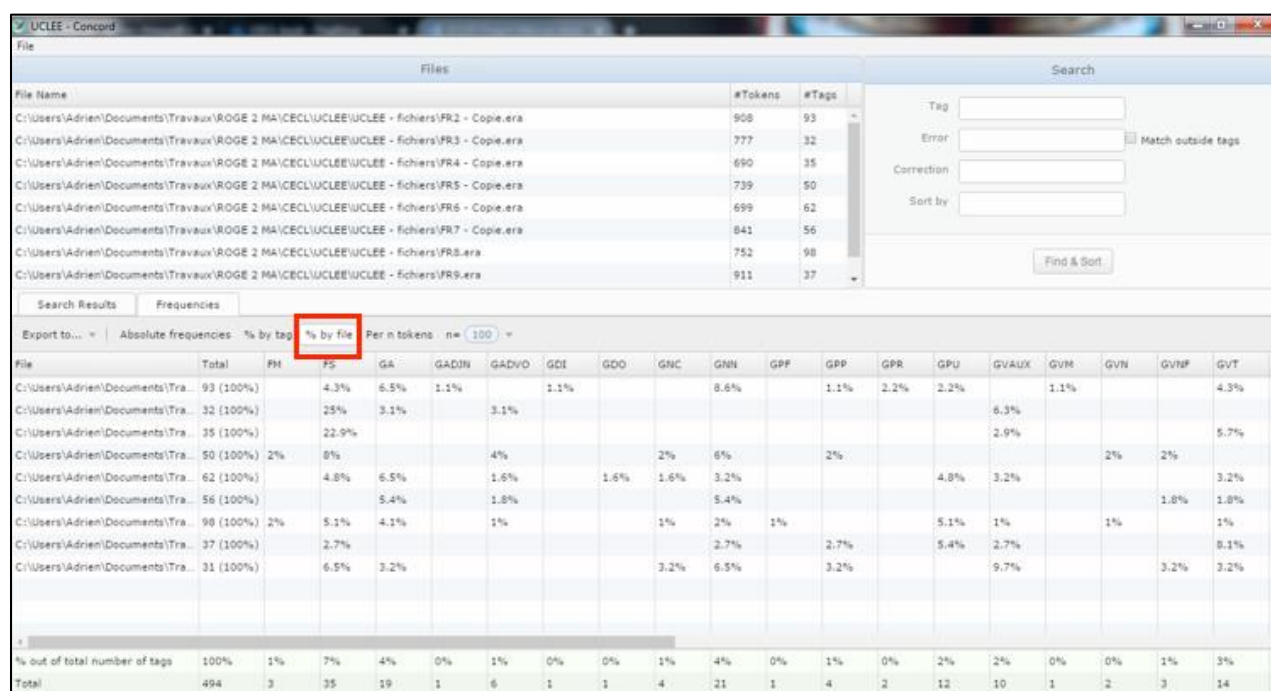
Figure 23: Detailed percentages of errors sorted by file

4. "Per n tokens" (relative frequencies)

This sub-header automatically calculates relative (or relative/normalized) frequencies to ensure comparability between texts of varying length ([Figure 24](#)). Following a 'rule of three', these results correspond to the division of absolute frequencies by the total number of words in the text(s), which are then multiplied by a normalisation base (e.g. relative frequency per **100** words, per **500** words, per **1,000** words).

Figure 24: Relative frequencies of error tags

It should also be noted that the user can at any time export the results displayed in the various sub-headers to CSV format for later use in other programs such as Microsoft Excel or statistical packages (see Figure 25).
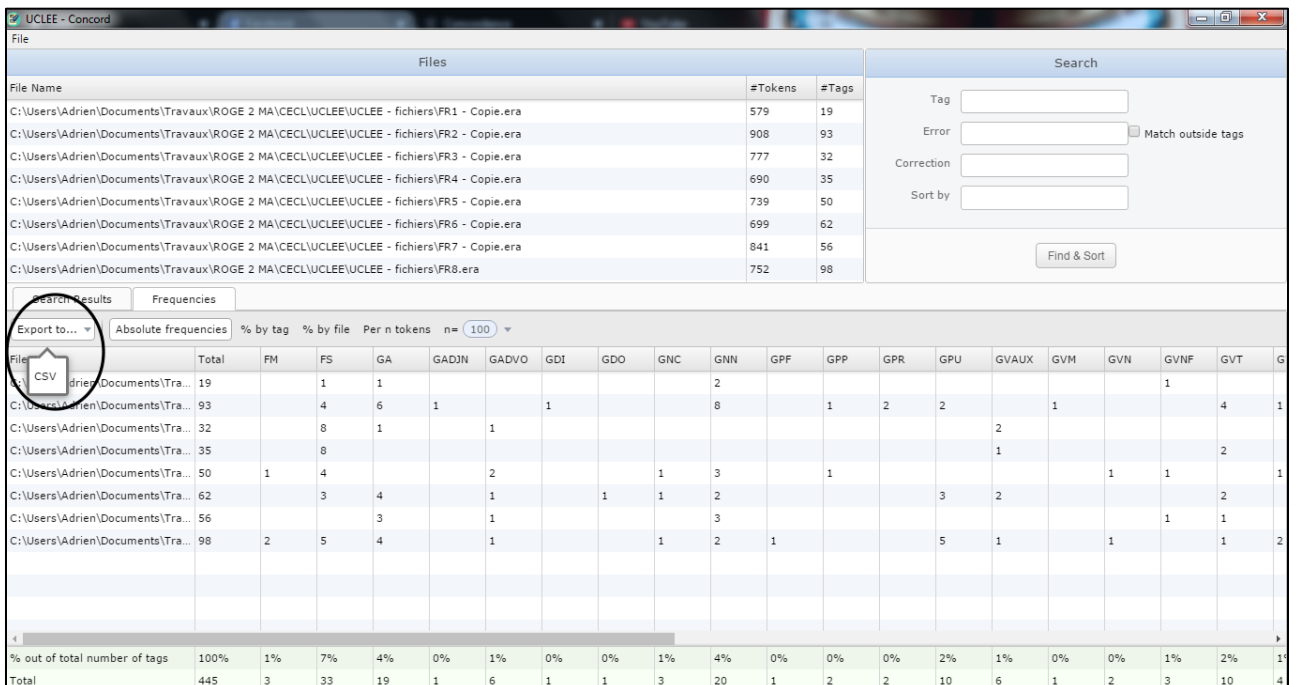
Figure 25: Exporting frequencies

2.7. Search

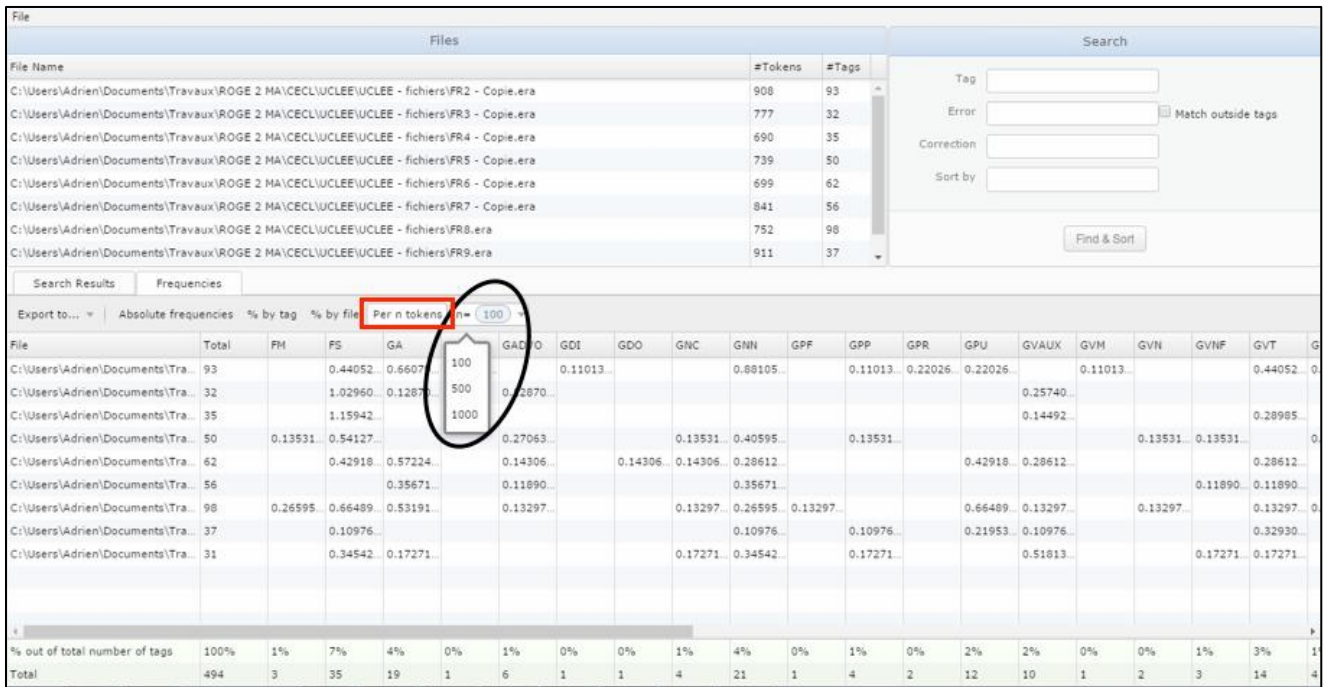The Search tab enables you to extract and highlight specific features by using Corpus Query Language[4]. For example, by entering /of|from/, all instances of the prepositions "of" and "from" are highlighted in the text (see Figure 26). The highlighted items can then be tagged using the Tags function.
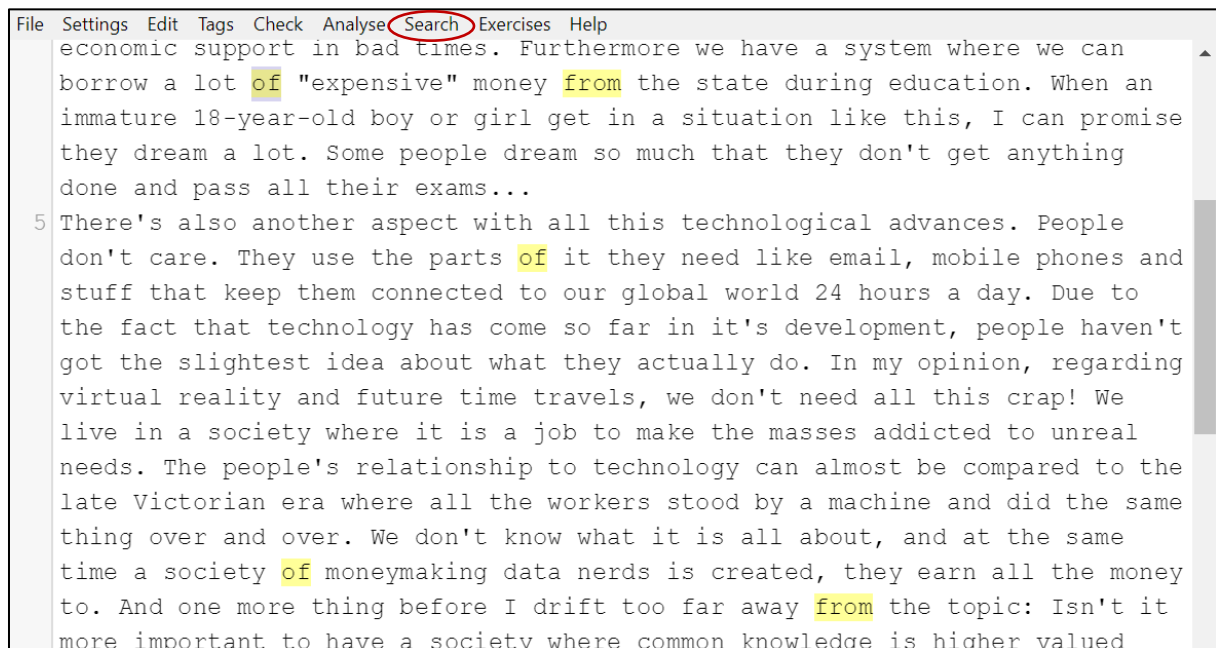
Figure 26: Search tab

2.8. Exercises

This tab can be used to generate tailor-made exercises for specific users or user groups by clicking on the "Create exercise" option (see Figure 27).

---

[4] More information on CQL searches can be found on the following websites http://fedora.clarin-d.uni-saarland.de/teaching/Corpus_Linguistics/Tutorial_RegExp.html or https://www.sketchengine.eu/guide/regular-expressions/
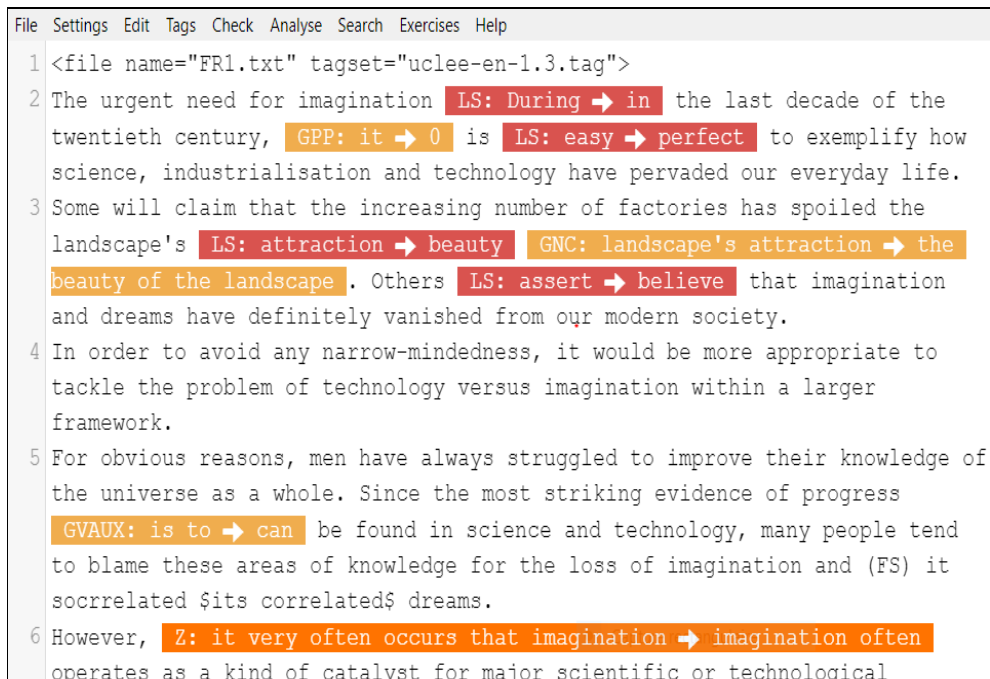
Figure 27: "Exercises" tab

You can create exercises focused on particular error tags by selecting these in the scroll-down list. Multiple tags can be selected at the same time by holding the <CTRL> (or <cmd ⌘> in Mac) key down while (de)selecting several tags (Figure 28).
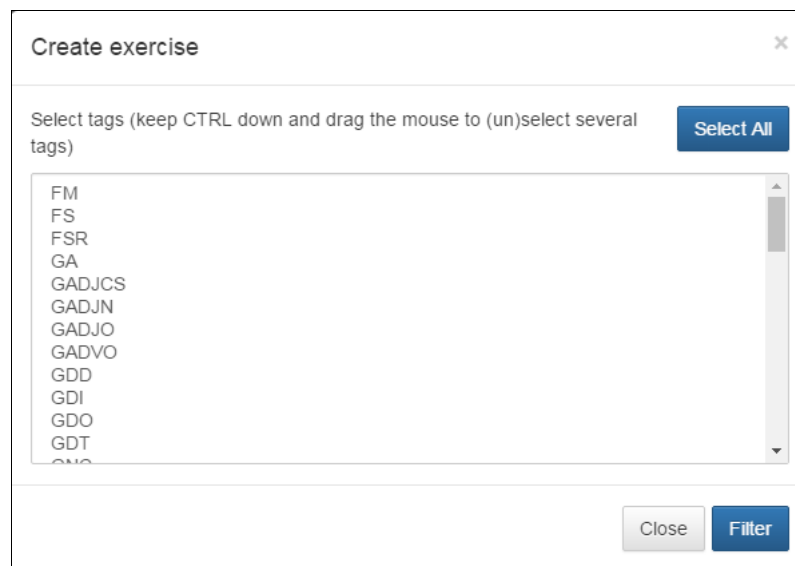

Figure 28: Creating exercises based on multiple tags

When one or several tags have been selected, the software opens a new window which displays all occurrences of the selected tags in the file. The default state is that all the tags are selected; the user can then choose to keep all examples or remove unwanted instances

by unticking the box situated next to the number of the concordance line (for example, to delete instances that are too challenging or too easy for particular user groups) (Figure 29).

The Louvain Error Tagging Editor allows the user to choose between three types of gap-filling exercise. These can be saved and exported as raw text to the clipboard, as a .txt file or an html webpage. To do this, select the desired format in the scroll-down menu and specify the information to be displayed (detailed tags, simplified tags or no tags), using the buttons at bottom left, above "Format" (Figure 29).
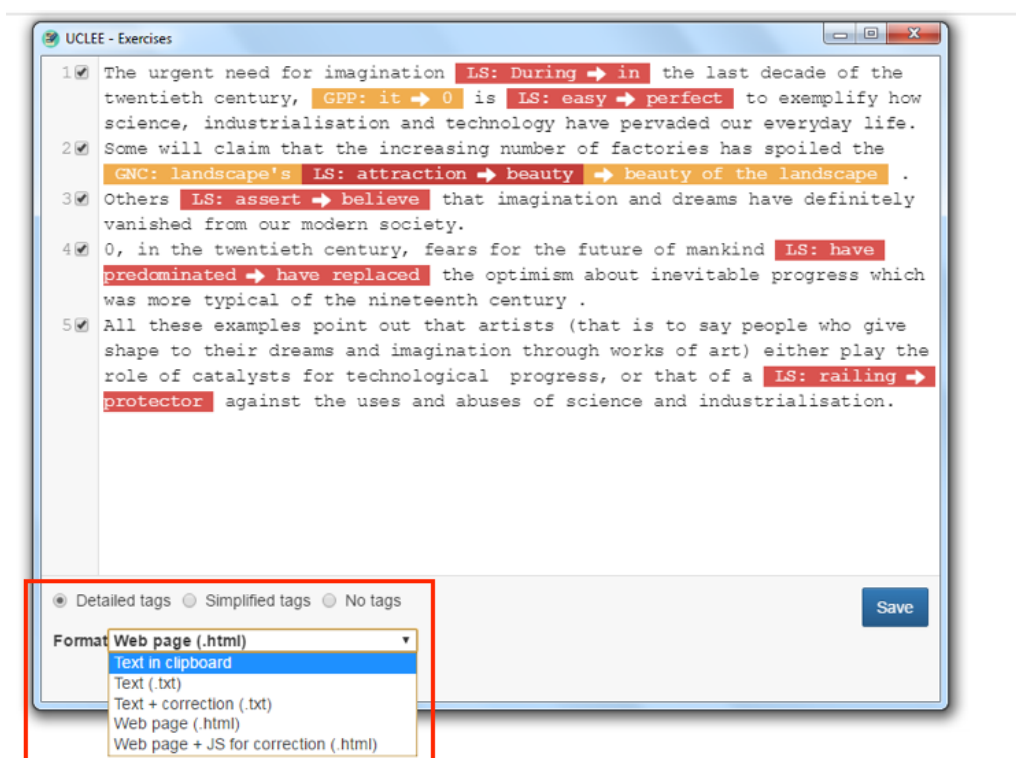
Figure 29: Exporting tailor-made exercises

As exemplified in Figure 30, the "Detailed tags" option makes use of the full error tagset to provide detailed information about the type of error made. "Simplified tags" is designed to guide learners in the right direction but only uses the main error categories (e.g. L instead of LS). "No tags" is intended to point out the error to the learner without adding any additional information; this option allows for a variety of possible answers.

For example, in the sentence *My house has been evaluated by an expert*, the erroneous string of words displayed is *has been evaluated*. In the "Detailed tags" option the learner is provided with the LSVF tag, which indicates that the error lies in the use of a lexical verb which is a false friend. The learner therefore knows that they do not have to think

26

about tense, aspect, etc. Their attention is clearly directed towards a specific issue. In the "Simplified tags" option the learner is provided with the LS tag, which indicates that they need to focus on a lexical error, but is not given the "false friend" hint. Finally, in the "No tags" option the learner is presented with only the erroneous string of words, so that they have to check whether the tense, aspect, number, choice of verb, etc. are correct.

1. <u>Detailed tags</u>

The urgent need for imagination [LS: During] _____ the last decade of the twentieth century, [GPP: it] _____ is [LS: easy] _____ to exemplify how science, industrialisation and technology have pervaded our everyday life.
Some will claim that the increasing number of factories has spoiled the [GNC: landscape's [LS: attraction] _____] _____ .
Others [LS: assert] _____ that imagination and dreams have definitely vanished from our modern society.
0, in the twentieth century, fears for the future of mankind [LS: have predominated] _____ the optimism about inevitable progress which was more typical of the nineteenth century .
All these examples point out that artists (that is to say people who give shape to their dreams and imagination through works of art) either play the role of catalysts for technological progress, or that of a [LS: railing] _____ against the uses and abuses of science and industrialisation.

1. Simplified tags

The urgent need for imagination [L: During] _____ the last decade of the twentieth century, [G: it] _____ is [L: easy] _____ to exemplify how science, industrialisation and technology have pervaded our everyday life.
Some will claim that the increasing number of factories has spoiled the [G: landscape's [L: attraction] _____] _____ .
Others [L: assert] _____ that imagination and dreams have definitely vanished from our modern society.
0, in the twentieth century, fears for the future of mankind [L: have predominated] _____ the optimism about inevitable progress which was more typical of the nineteenth century .
All these examples point out that artists (that is to say people who give shape to their dreams and imagination through works of art) either play the role of catalysts for technological progress, or that of a [L: railing] _____ against the uses and abuses of science and industrialisation.

2. No tags

The urgent need for imagination [$During] _____ the last decade of the twentieth century, [$it] _____ is [$easy] _____ to exemplify how science, industrialisation and technology have pervaded our everyday life.
Some will claim that the increasing number of factories has spoiled the [$landscape's [$attraction] _____] _____ .
Others [$assert] _____ that imagination and dreams have definitely vanished from our modern society.
0, in the twentieth century, fears for the future of mankind [$have predominated] _____ the optimism about inevitable progress which was more typical of the nineteenth century .
All these examples point out that artists (that is to say people who give shape to their dreams and imagination through works of art) either play the role of catalysts for technological progress, or that of a [$railing] _____ against the uses and abuses of science and industrialisation.

Figure 30: Illustration of exercise formats

2.9   Help

2.9.1. *The Louvain Error Tagging Manual (2.0)*

This submenu opens the Louvain Error Tagging Manual (version 2.0).

2.9.2. *"UCLEE Helpfile"*

This submenu opens the UCLEE user guide.

Appendix: Working with texts previously tagged with uclee-en-1.3

This manual relates exclusively to the most recent version of UCLEE, version 2.0. However, some users may wish to continue working with texts already tagged using the previous version (UCLEE 1.3) and to tag further texts within that system. This appendix describes the procedures to follow if you are in that situation.

I. To convert previously tagged texts from version 1.3 to version 2.0 display

1. Open UCLEE version 1.3 as follows:

The "Tagset selection" window that opens when you launch UCLEE provides several tagset options (Figure A 1):
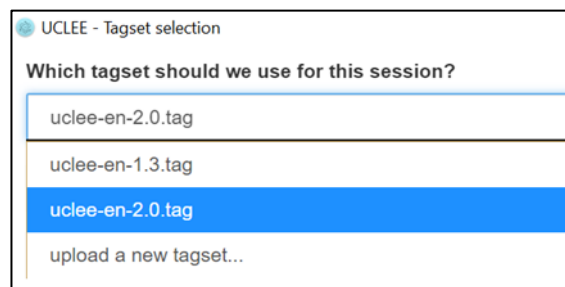

Figure A 1: Tagset selection

Select uclee-en-1.3.tag.

Click on the "Open UCLEE" button at bottom right of the window. This will take you to the software's main interface (Figure A 2).
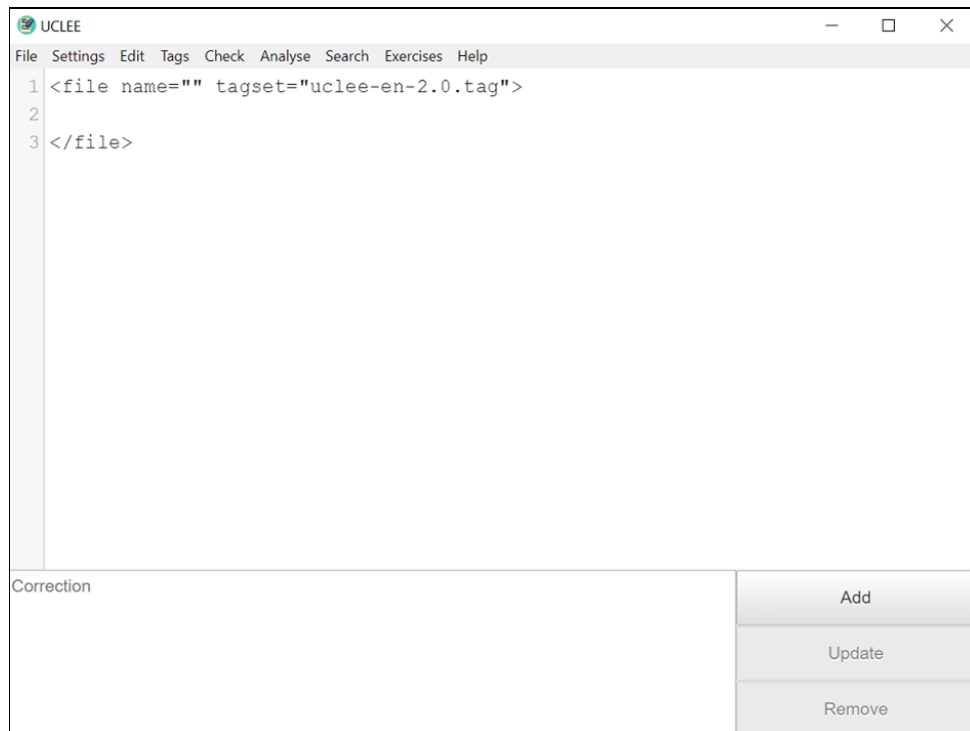
Figure A 2: Main interface

2. Convert the file(s) from version 1.3 to version 2.0 display

Click on the "Convert files from 1.3 tagset with double dollars" option in the "File" menu ([Figure A 3](#)).
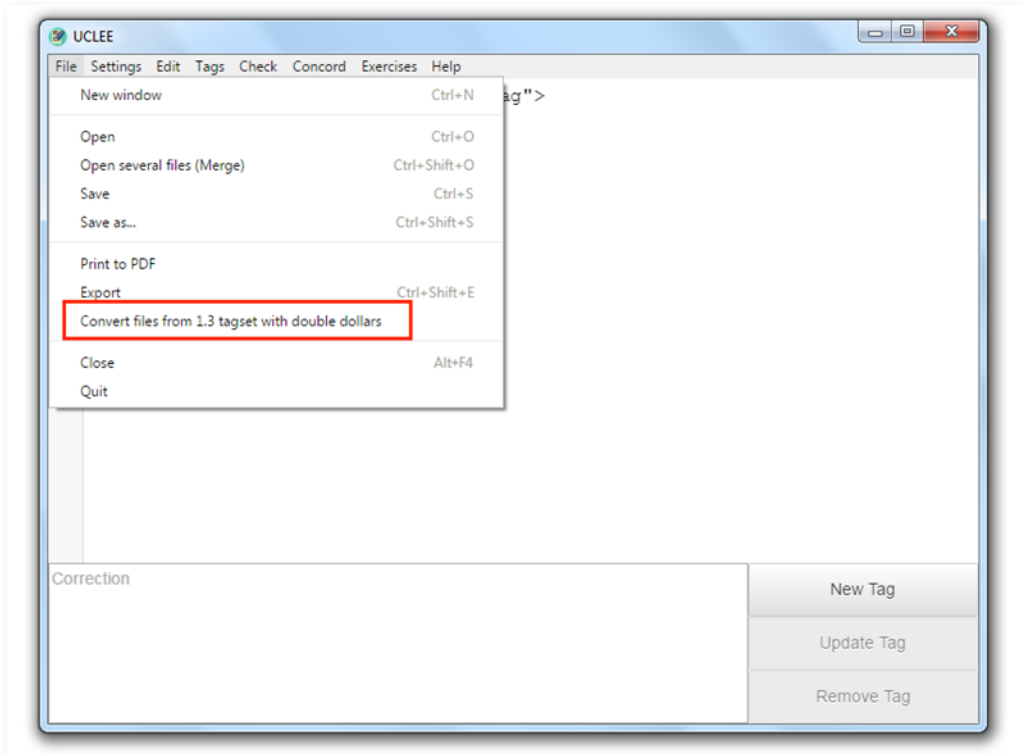
Figure A 3: Convert files from 1.3. tagset with double dollars

The window shown in <u>Figure A 4</u> will open.

Click on 'Select files to convert' and browse to the file(s) you wish to convert (<u>Figure A 4</u>).

After conversion is complete, open the corresponding .era file(s), the path to which is indicated in the 'conversion completed' message that appears if the conversion was successful (<u>Figure A 5</u>).
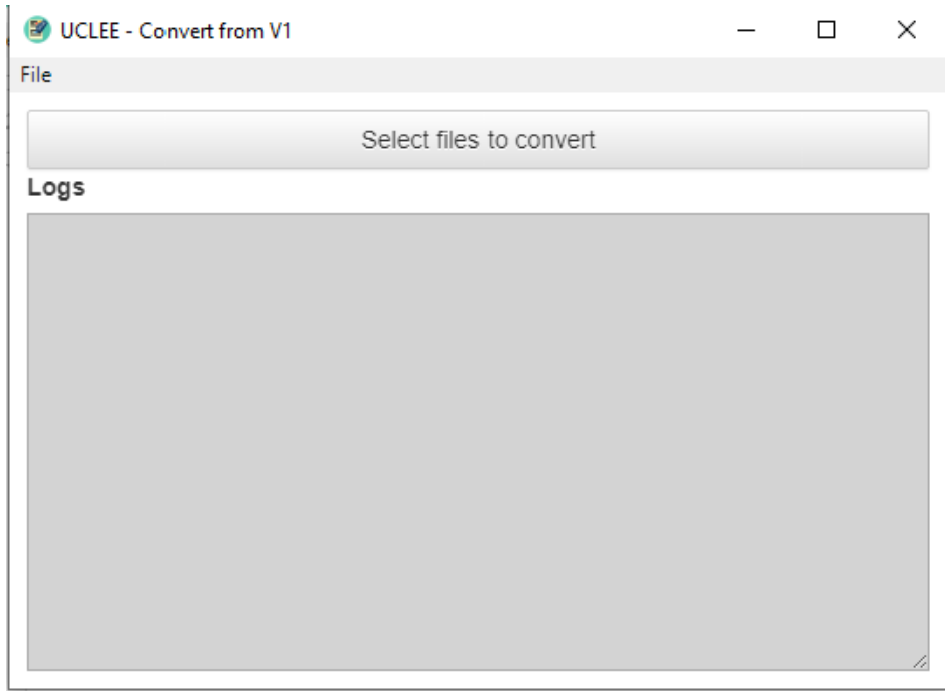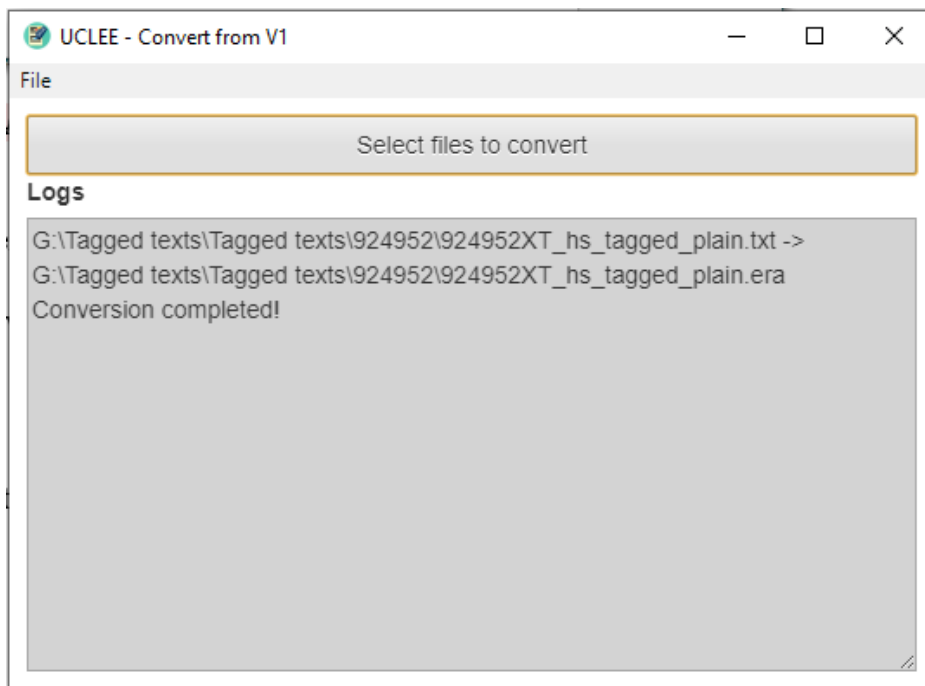
Figure A 4: Select files to convert

Figure A 5: Conversion completed

The file(s) originally tagged in 1.3 (Figure A 6) will now be displayed as shown in Figure A 7.
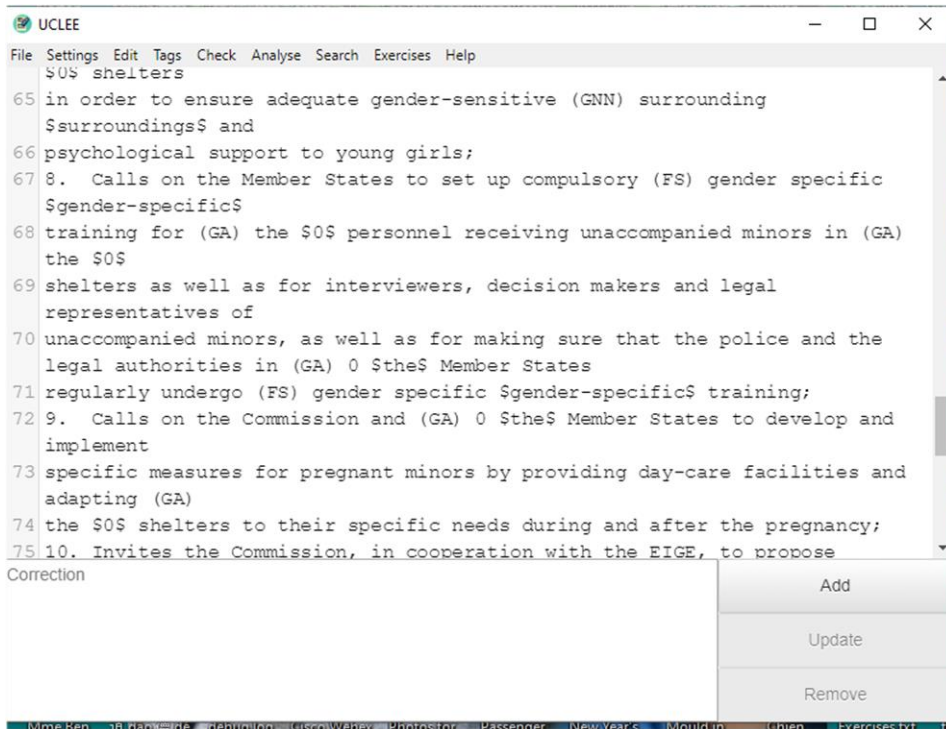
Figure A 6: File in 1.3 tagset with double dollars before conversion
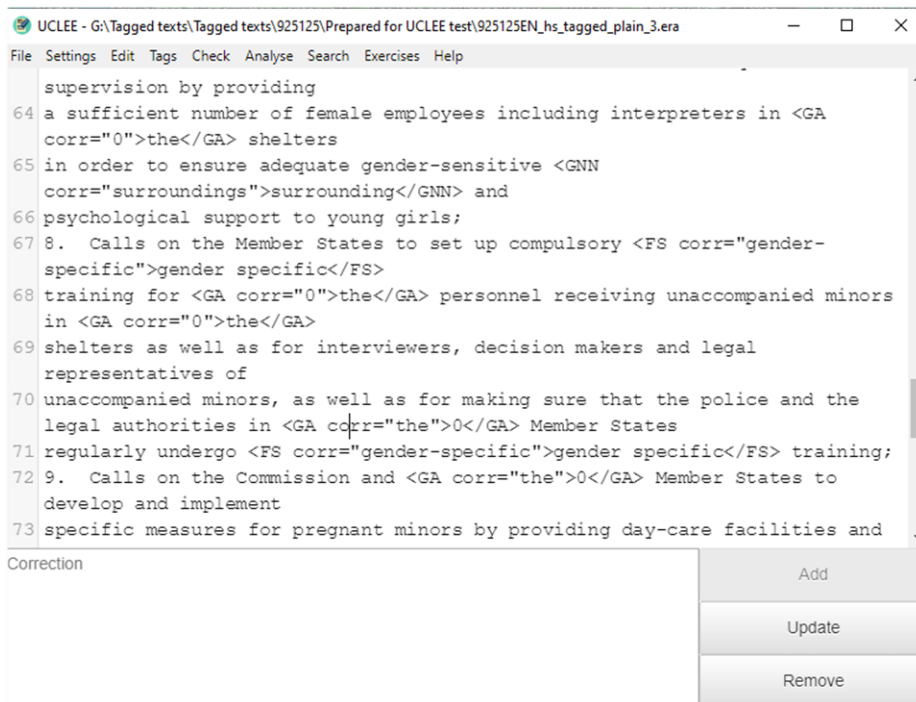

Figure A 7: Dialog box after conversion

II.   UCLEE 1.3 and 2.0 tagsets

The UCLEE 2.0 tagset is not identical to that of UCLEE 1.3: some tags have been added and others changed, so that they overlap only partially with earlier tags. For example, the

LS (Lexical Single) category, which in UCLEE 1.3 comprised only two sub-categories (LS and LSF), in UCLEE 2.0 comprises ten (LSADJ, LSADV etc.). It is very important that you should be aware that the conversion process just described is a technical conversion only, which changes the way that your previous error tags are formatted and displayed, but **does not change the tag labels into those of the UCLEE 2.0 tagset**.

III.   Tagging new texts with UCLEE 1.3

You may also wish to tag new texts in order to add them to those previously tagged in UCLEE 1.3 to form a single corpus. Please be aware that **the two tagsets cannot be mixed**. You must continue to work in UCLEE 1.3 until your corpus is complete. If you were to work in the two versions and then attempt to open your whole corpus for analysis, you would be unable to open it in either version.