

ComPair

Compare documents, in any formats, in any languages

Quick start guide

contact@cubAlx.com

<http://cubAlx.com>

cub[AI]^x

Table of contents

1 Download and install ComPair.....	3
2 Document comparison.....	4
3 Alignment edition.....	11
4 HTML export.....	12
5 TMX export.....	14

1 Download and install ComPair

ComPair will enable you to compare 2 documents before/after a correction or before/after a translation. It produces comparison reports, that can be sent by email, to discuss your work with your customers, your translators, your partners, or your lawyers.

ComPair is a universal document comparer, to be installed on your computer:

- It compares two versions of a document,
- If needed in two different file formats,
- In the same language or in two different languages.
- All common office file formats are accepted, including PDF scanned documents.
- All languages are processed (languages can be added on demand in the selectors).

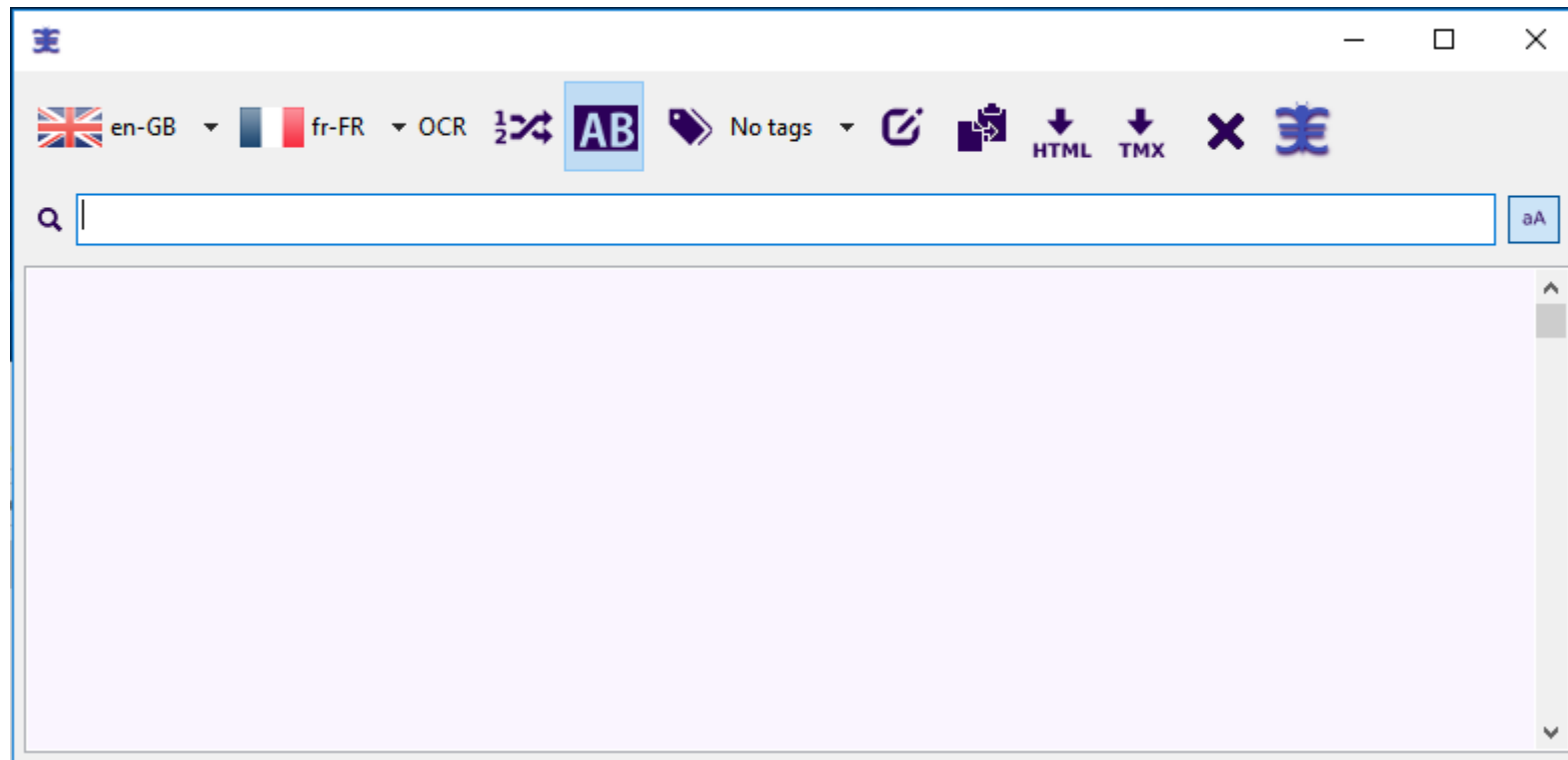
ComPair installer for Windows can be downloaded here:

<http://cubAIx.com/>

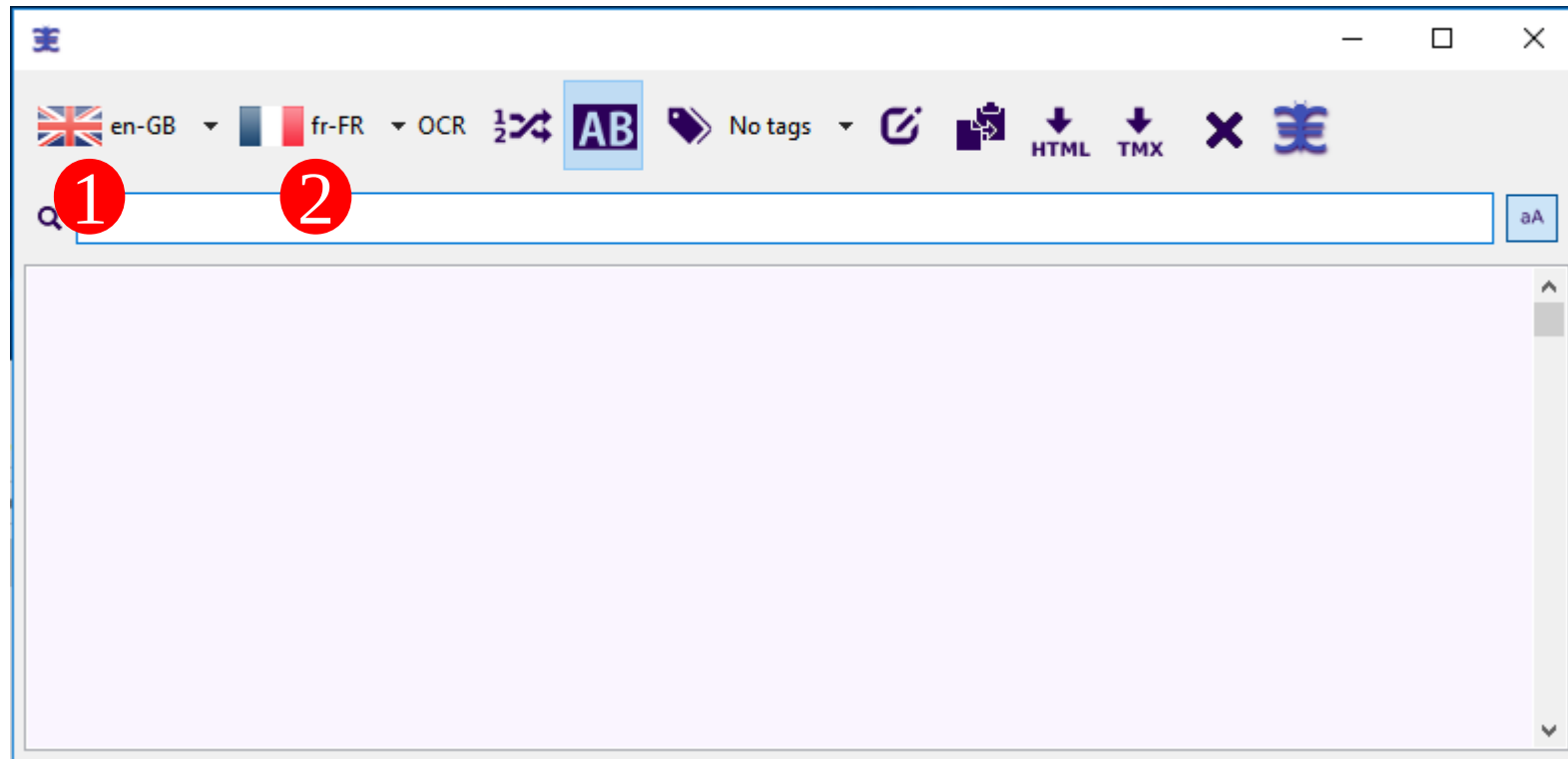
2 Document comparison

Remark: to compare PDF documents, you will obtain better results when converting them as DOCX before comparison. For this, the best is certainly to use a dedicated software, including an OCR (for example, Abbyy or Nuance products). MS Office, in its latest versions, is a quite good PDF to DOCX converter (without OCR).

Launch ComPair, you should get this blank window:

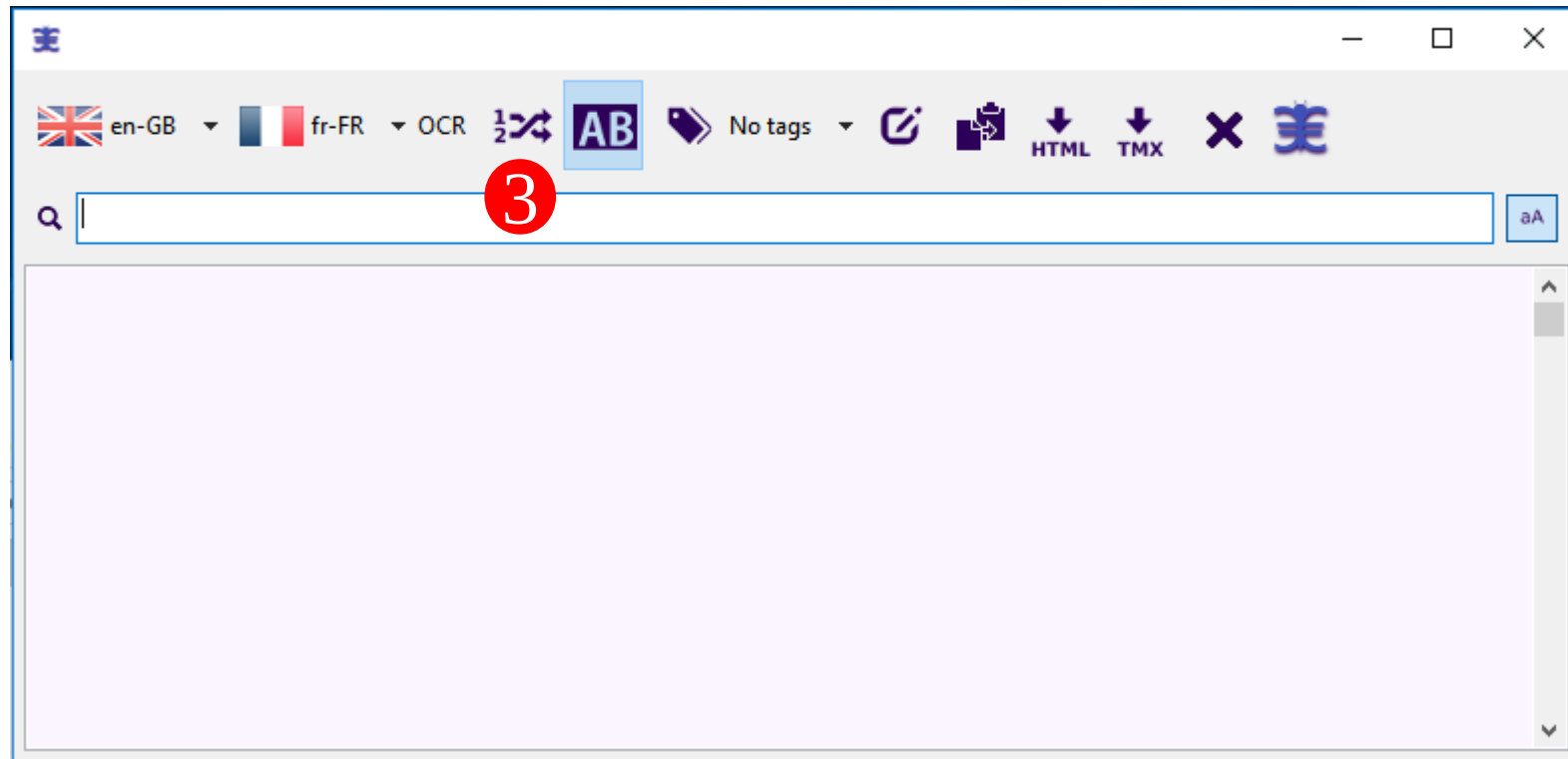


Click on the first two buttons to select the language for each of your documents to be compared:

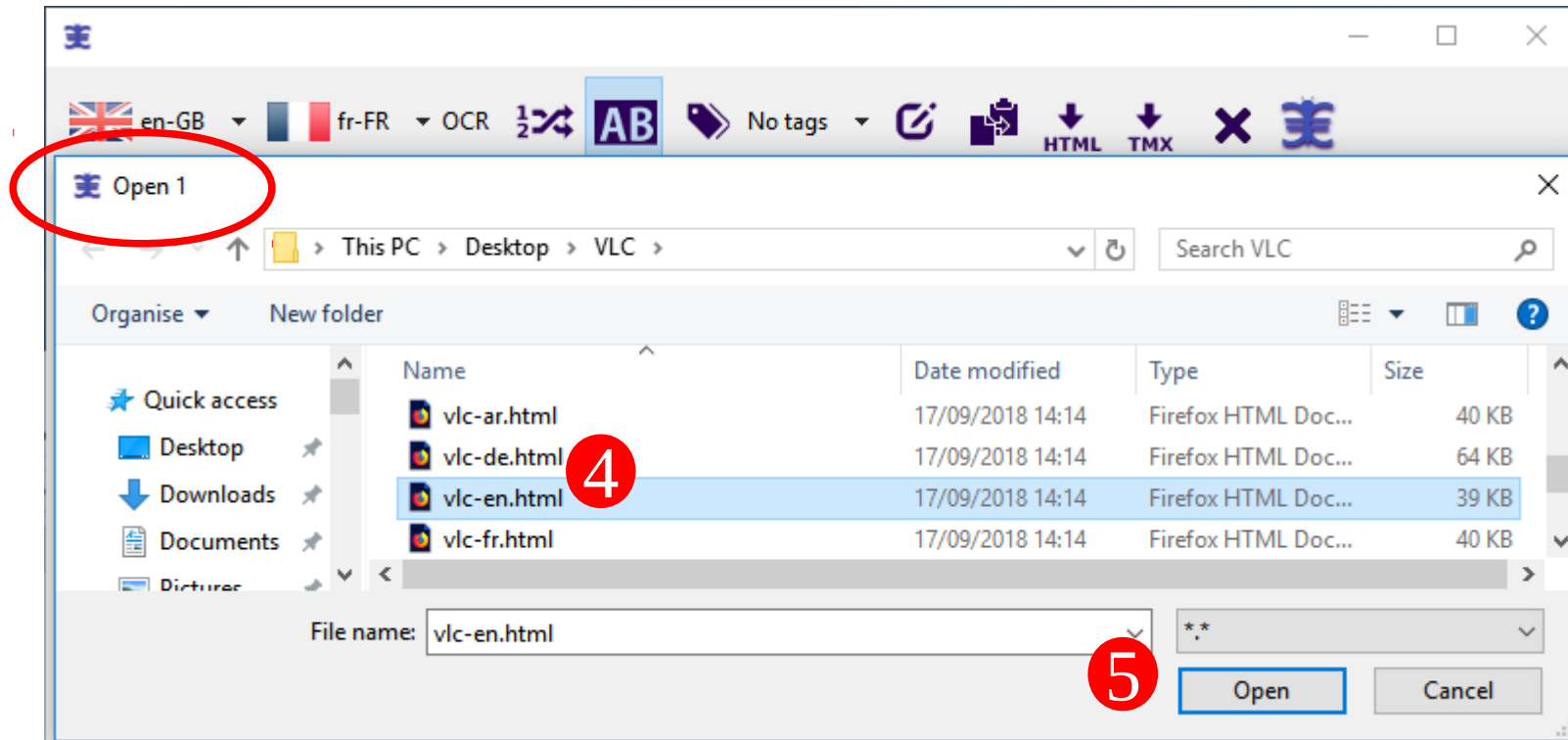


Click on the “OCR” button if you want to enable the text extraction from images in scanned PDF documents (if the free Tesseract OCR is installed).

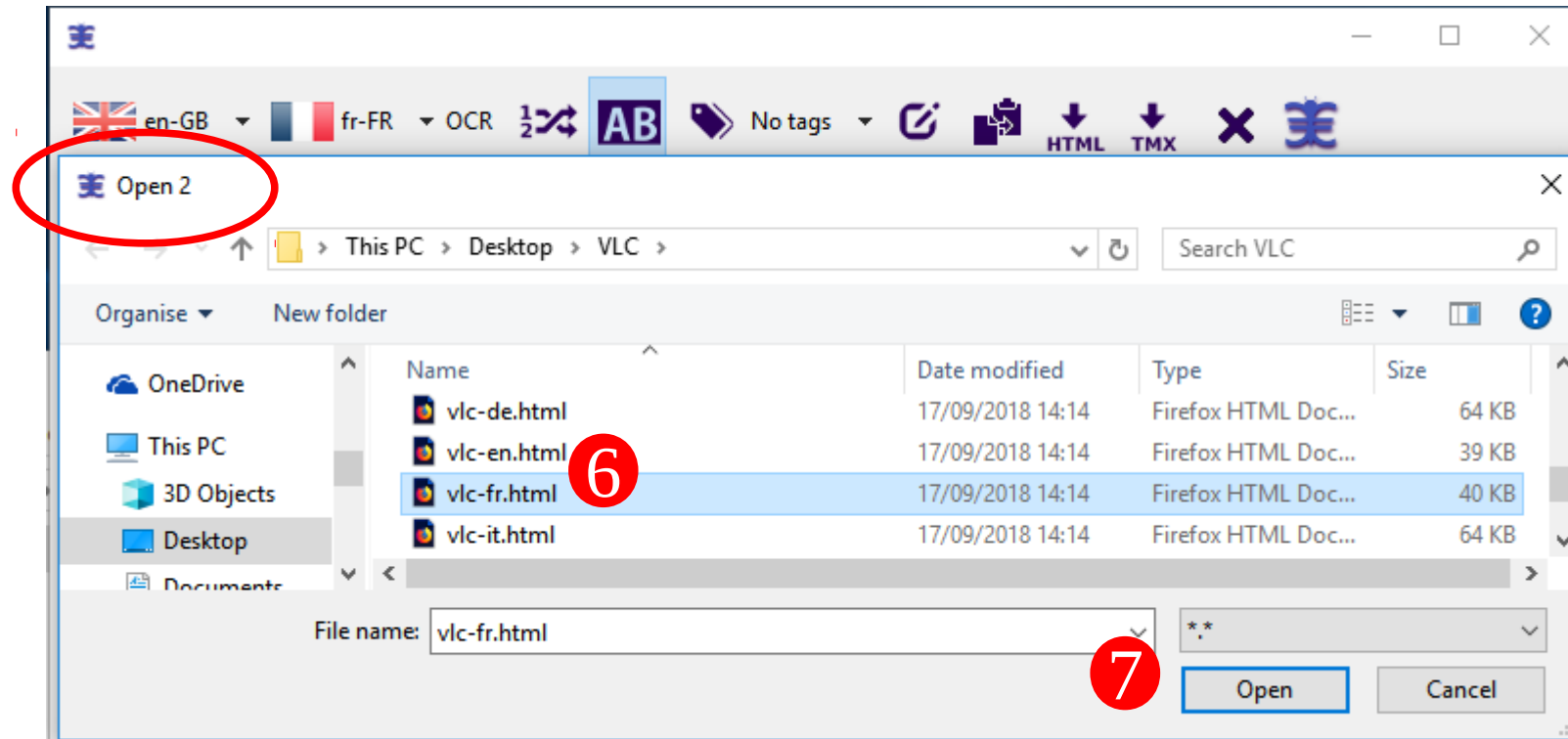
Click on the fourth “1/2” button:



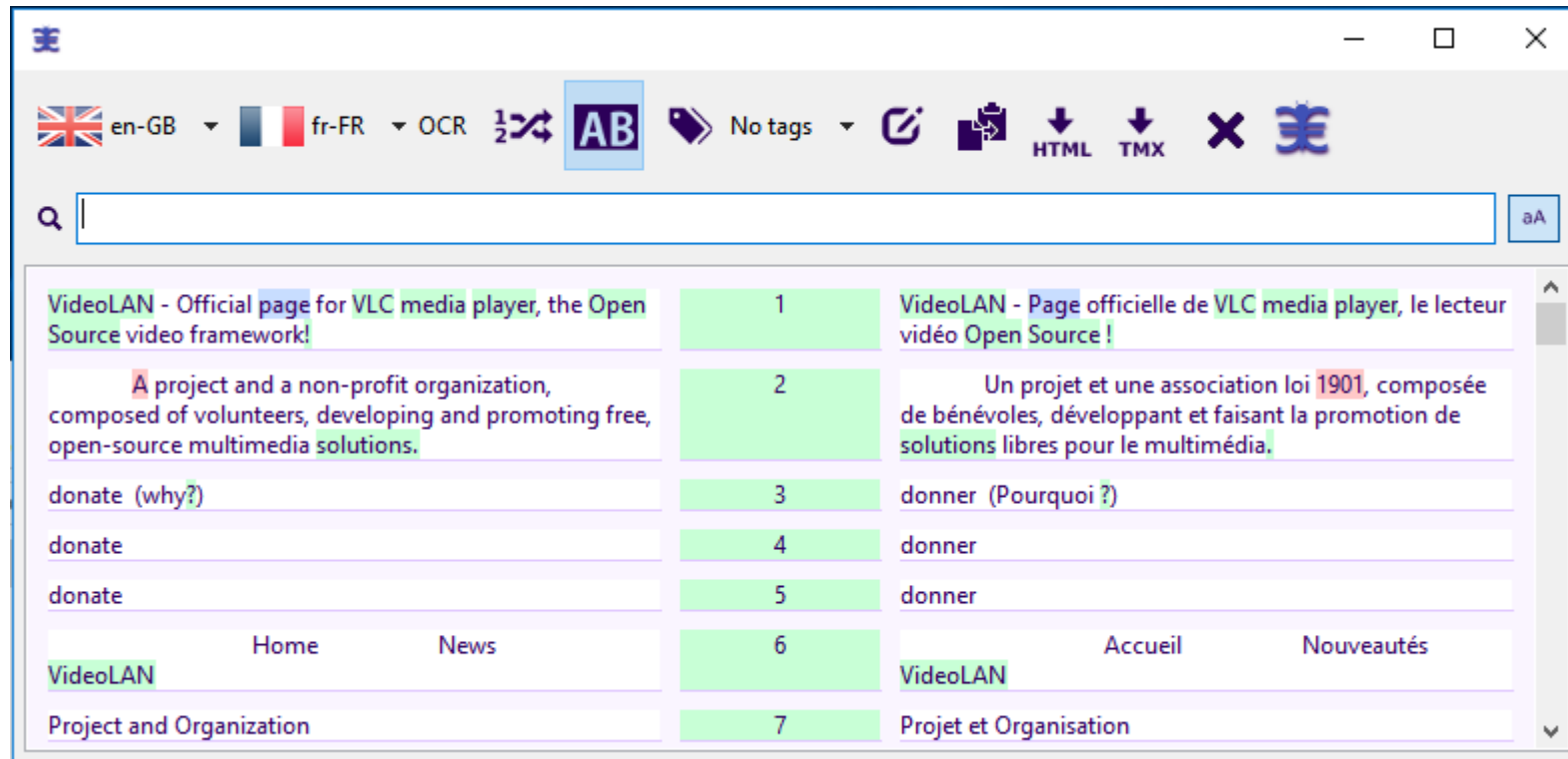
It should open a file selector. Select the first file to be compared, and click the “Open” button:



It should open a second file selector. Select the second file to be compared, and click the “Open” button:



After a short calculation time, the comparison result should appear with color highlighting:



Colors are showing elements you should certainly look at with more attention. They don't highlight something especially good or bad, it depends on your texts:

- **GREEN**: identical text
- **BLUE**: identical text with case differences
- **RED**: special text objects (capitalized, punctuation, number, ..) without a found corresponding pair

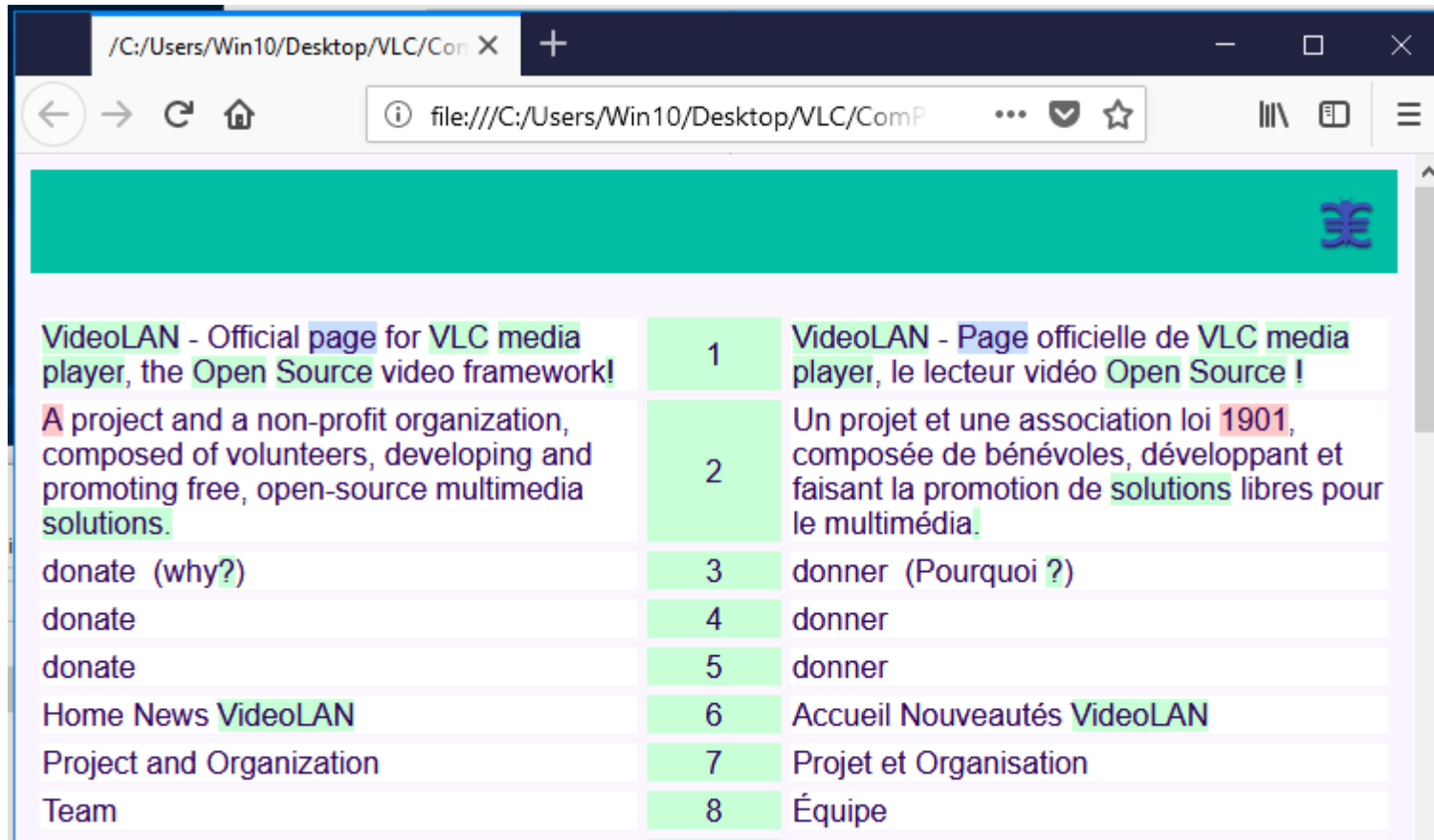
You can enter some text in the search bar to limit the comparison table on lines containing this text, with or without being case sensitive.

You may choose how to view the tags, especially for HTML/XML file comparison:

The screenshot displays the ComPair application window. The toolbar at the top shows language settings for en-GB and fr-FR, an OCR icon, a bidirectional arrow icon, and a highlighted 'AB' icon. To the right of the 'AB' icon is a dropdown menu labeled 'Reduced tags'. Further right are icons for 'HTML', 'TMX', and a close button. Below the toolbar is a search bar with a magnifying glass icon and an 'aA' button. The main content area is a table with six rows of comparison results. The first row shows a comparison between an English version and a French version of the VideoLAN website. The second row shows a comparison of the '<a> ' tag, which is circled in red. A red arrow points from the 'Reduced tags' dropdown menu to this circled tag. The table has columns for the source text, a green numerical score, and the target text. The source text for the circled tag is '<a> project and a <a> non-profit organization , composed of volunteers, developing and promoting free, open-source multimedia solutions.' The target text is '<a> Un projet et une <a> association loi 1901 , composée de bénévoles, développant et faisant la promotion de solutions libres pour le multimédia.'

VideoLAN - Official page for VLC media player, the Open Source video framework!	1	VideoLAN - Page officielle de VLC media player, le lecteur vidéo Open Source !
<a> project and a <a> non-profit organization , composed of volunteers, developing and promoting free, open-source multimedia solutions.	2	<a> Un projet et une <a> association loi 1901 , composée de bénévoles, développant et faisant la promotion de solutions libres pour le multimédia.
donate <a> (why?)	3	donner <a> (Pourquoi ?)
donate	4	donner
donate	5	donner
<a> <a> <a> <a> Home <a> News <a> <a> <a> <a> <a> <a> Accueil <a> Nouveautés <a> <a> VideoLAN 	6	<a> <a> <a> <a> Accueil <a> Nouveautés <a> <a> VideoLAN

This exported file can be opened with a simple web browser (Internet Explorer, Firefox, Chrome, ..):



5 TMX export

The TMX export is useful for translators and translation agencies, to enrich their translation memories.

