



Voice-text integrated system for interpreters

User Guide

Contenido

MODULE I	3
1. COMPARABLE CORPUS MANAGEMENT	3
1.1. <i>My corpora</i>	3
1.2. <i>Corpus query</i>	4
1.3. <i>Importing a corpus</i>	7
1.4. <i>Semi-automated corpus compilation</i>	8
2. PARALLEL CORPUS MANAGEMENT	9
2.1 <i>My corpora</i>	9
2.2 <i>Corpus query</i>	9
2.3 <i>Corpus importation</i>	10
3. DICTIONARY AND GLOSSARY MANAGEMENT	11
3.1. <i>My glossaries</i>	11
3.2. <i>Glossary management</i>	12
4. NAMED ENTITY RECOGNITION SYSTEM.....	13
4.1 <i>Named Entity Recognition in text</i>	13
4.2 <i>Named Entity Recognition in corpus</i>	14
5. TEXT SUMMARISATION SYSTEM.....	15
MODULE II	16
6. NOTE-TAKING	16
6.1 <i>Automatic note-taking</i>	16
7. MACHINE TRANSLATION.....	17
8. GLOSSARY QUERY	18
MODULE III	19
9. EXERCISES	19
9.1 <i>Anticipation</i>	19
9.2 <i>Numbers</i>	20
9.3 <i>Sight Translation</i>	21
9.4 <i>Glossaries</i>	22
10. SYMBOLS	23
APPENDIX: MICROPHONE CONFIGURATION	24
APPENDIX: VERTICALIZED TEXT (.VRT)	26



Module I

This module is designed to carry out the preparatory tasks for an interpretation, such as the creation and query of corpora, glossaries, etc. The different parts of this module are described below.

1. Comparable corpus management

This module offers different functionalities related to the corpora. You can create your own corpus, either using text files or from an Internet search. Different types of queries can also be performed on the created corpus.


1.1. My corpora

This section shows the basic information for the corpora that have been created: name, language, number of words and number of texts that comprise it. It also allows you to delete a corpus by clicking on the  icon, or go to the corpus query by clicking on the  icon.
























In addition, you can see the information for the preloaded corpora on the system.

Comparable corpus management

[My corpora](#)[Corpus query](#)[Import corpus](#)[Semi-automatic corpus compilation](#)

 Here you can see all the information regarding you own corpora as well as the system corpora.

My corpora (6/20)

Name:	Description:	Query:	Delete:
Cocina	 es  31,279 words  20 texts		
CoronaVirus	 en  211,405 words  59 texts		
coronavirus	 es  41,920 words  10 texts		
nat_disaster	 en  84,296 words  35 texts		
nesca_sub	 es		

1.2. Corpus query

In this section, you can perform different queries over the corpora that you have created or imported. You just need to select one or several corpora from the list, type the search term and choose the type of query. To select several corpora, first select one and then press Ctrl+click on the next corpus that you would like to include in the selection.

In addition, several options can be set for the term, such as “Case sensitive” (makes the search case sensitive), “Diacritics insensitive” (ignore diacritics), “Grammatical category” (only terms matching the selected grammatical category will be shown). You can also choose if you want to perform the search by word form (as it appears in the corpus) or by lemma (lemmatized form of terms). The different types of queries that can be performed are described below.

- Concordances

Concordances are shown in a small context containing the given term in the selected corpora. The results can be sorted alphabetically according to the first word from the left (L1), the first word from the right (R1), the second word from the left (L2), the second word from the right (R2), and so on. In addition, several ordering levels can be set.

The screenshot displays the 'Corpus query' interface. It includes a 'Corpus name' dropdown with 'CoronaVirus_en' selected, a 'Term' input field with 'disease', and 'Search type' options with 'Concordances' selected. The 'Order by' section shows 'R1' selected. There are checkboxes for 'Case sensitive', 'Diacritics insensitive', 'Word form', and 'Lemma' (selected). A 'Grammatical category...' dropdown is also present. At the bottom of the form are 'Search' and 'Clear' buttons.

Results for corpus: **CoronaVirus**

349 coincidences were found in 22 different texts (in 211,405 words [59 texts]; frequency: 1650.86 instances per million words).

Show/hide file names

...es/ Diseases/ Coronavirus disease 2019 Coronavirus	disease	(COVID-19) Pandemic Public Advice Country & technical guidance Latest updates
... Strategies and Plans Rolling updates on coronavirus	disease	(COVID-19) Donate Your questions answered Travel advice Situation reports Media
...anizations and communities Read More Coronavirus	disease	(COVID-19) outbreak situation View dashboard → Confirmed cases Confirmed deaths
...aging manifestations of the 2019 novel coronavirus	disease	(COVID-19): A multi - center study in Wenzhou city , Zhejiang ,

- N-grams

In this case you must assign a value to N (2 for bigrams, 3 for trigrams, etc.). The N-grams containing the given term in the selected corpora are shown for the chosen N value.

Corpus query

Corpus name:

cocina_es
CoronaVirus_en
medicina_es
micorpus_es

Term:

disease

☐ Case sensitive
☐ Diacritics insensitive

Grammatical category
Any

☒ Word form
☐ Lemma

Search type:

☐ Concordances
☒ N-grams 3
☐ Patterns
☐ Frequency list

Search

Clear

Add selected terms to:

Select glossary...

Add

Results for corpus: CoronaVirus

☐ Mark all

Nº	Resultados	Nº appearances	Percent
<input type="checkbox"/> 1	Coronavirus disease 2019	91	13.48% (total: 0.043%)
<input type="checkbox"/> 2	for Disease Control	26	3.85% (total: 0.0123%)
<input type="checkbox"/> 3	Disease Control and	24	3.56% (total: 0.0114%)

- Patterns

The terms or collocations that match the chosen pattern and contain the given term in the selected corpus are displayed. Custom patterns can also be created. If the “Ignore term” option is checked, the term will not be considered.

Corpus query

Corpus name:
☒ CoronaVirus_en
☐ coronavirus_es
☐ nat_disaster_en

Term:

Search type: ☐ Concordances ☐ N-grams ☒ Patterns ☐ Frequency list

Patterns: Custom

☐ Ignore term

Grammatical category...
☒ Word form
☐ Lemma

Add selected terms to:

Results for corpus: CoronaVirus

☐ Select all

N°	Results	N° appearances	Percentage
<input type="checkbox"/> 1	infectious disease	16	47.06% (total: 0.0076%)
<input type="checkbox"/> 2	Respiratory disease	5	14.71% (total: 0.0024%)
<input type="checkbox"/> 3	Cardiovascular disease	3	8.82% (total: 0.0014%)

• Frequency list

This functionality offers a list of the most frequent words that could be candidates for terms or phraseological units in the selected corpora. You do not need to enter any term, only the grammatical category you do not want to be shown ("Noun", "Adjective", "Verb", "Preposition", "Conjunction", "Adverb".)

Corpus query

Corpus name:
☒ CoronaVirus_en
☐ medicina_es
☐ micorpus_es

Term:

Search type: ☐ Concordances ☐ N-grams ☐ Patterns ☒ Frequency list

Stop categories: ☒ Preposition ☒ Verb ☐ Adjective ☐ Noun

Grammatical category: Any
☒ Word form
☐ Lemma

Add selected terms to:

Results for corpus: CoronaVirus

☐ Mark all

N°	Resultados	N° appearances	Percent
<input type="checkbox"/> 1	New	3231	6.81% (total: 1.5283%)
<input type="checkbox"/> 2	Cases	2203	4.64% (total: 1.0421%)
<input type="checkbox"/> 3	source	2161	4.55% (total: 1.0222%)
<input type="checkbox"/> 4	Deaths	1054	2.22% (total: 0.4986%)
<input type="checkbox"/> 5	COVID-19	770	1.62% (total: 0.3642%)

After performing the query, you can select (by checking the checkbox) the resulting terms (except for *Concordances*) and add them to a glossary that you can later consult and edit (see section *Glossary query*). To do this, select a glossary from the drop-down menu or choose a name to create a new one and press the "Add" button. In order to quickly select several terms, you have two options:

- 1) Check "Select all" to select all terms.
- 2) Select a range. To do this, check an initial checkbox and, while pressing the SHIFT button, check an ending checkbox. All checkboxes in between will be automatically checked.

Search Clear

Add selected terms to: Select glossary... Add

Results for corpus: C

☐ Select all

N°	Results	N° appearances	Percentage
<input type="checkbox"/> 1	infect	16	47.06% (total: 0.0076%)
<input type="checkbox"/> 2	Respiratory disease	5	14.71% (total: 0.0024%)
<input type="checkbox"/> 3	Cardiovascular disease	3	8.82% (total: 0.0014%)

1.3. Importing a corpus

In this section you can create your own corpus from a set of texts. To do so, choose a name for the corpus and select the text files from your computer. These files can be in TXT, PDF or VRT¹ (VeRTicalized Text) format. Select the language of the texts. After pressing the "Import" button, the corpus will be created, and you will be able to consult it in the section *Corpus query*.

Comparable corpus management

My corpora Corpus query Import corpus Semi-automatic corpus compilation

Create a corpus from your own text files. Type a name, upload your files and select the language.

Import corpus

Corpus name:

Files: ? Add

Select language: ☒ Spanish ☐ English

Import

¹ The columns order must be: token, part of speech, lemma. See *Appendix: Verticalized text*.

1.4. Semi-automated corpus compilation



In this section you will be able to create a corpus using texts from the Internet. You just need to type a search term or phrase (you can use the usual search techniques such as putting the phrase in quotation marks to search for an exact match or typing the "-" Boolean operator before a word to exclude it from the search), select a level (each one represents a different search engine) and select a language. After pressing the "Search" button, a list of related websites will be displayed, from which you will be able to choose as many as you consider appropriate to add to your corpus. Then you only have to choose a name for the corpus and select one of the three following options: "Import and download .txt" (the corpus will be downloaded in plain text format), "Import and download .vrt" (the corpus will be downloaded in vertical text format with morphological labelling, see *Appendix: Verticalized text*) or "Import only". In any case, a corpus which can be consulted in the section *Corpus query* will be created. You can also choose whether to combine all the obtained texts into a single file (by selecting "Create a file with all the information") or create a text file for each chosen web resource in addition to a metadata file (by selecting "Create a file for each web").

The screenshot shows a web interface for creating a corpus. At the top, there is a search bar with the text "lockdown". To its right are two sections: "Select search engine:" with radio buttons for Level 1 (selected), Level 2, Level 3, Level 4, and Level 5; and "Select language:" with radio buttons for Spanish and English (selected). A green "Search" button is below these options. Below the search bar, it says "100 results". There are buttons for "Select all", "Remove all", and a "Select" button with a dropdown menu showing "n". Below this, there are three search results, each with a "Go to website" button. The first result is "National lockdown: Stay at Home - GOV.UK" with a summary dated 04/01/2021. The second result is "Prime Minister announces national lockdown - GOV.UK" with a summary dated 04/01/2021. The third result is "lockdown - latest news, breaking stories and comment - The ...". At the bottom of the results, there is a "Covid: What are the UK's lockdown rules and when will they ..." result. On the right side, there is a section titled "Added URLs (2)" showing two URLs from gov.uk. Below this is a "Corpus name" field with the placeholder "Name". At the bottom right, there is a section titled "What do you want to do?" with three radio buttons: "Import and download .txt", "Import and download .vrt", and "Import only" (selected). Below this is a section titled "How?" with two radio buttons: "Create one file with all information" and "Create one file per web" (selected). A green "Process" button is at the bottom right.

2. Parallel corpus management


In this module you can create your own parallel corpus using text files. Different types of queries can also be performed on the created corpus.

2.1 My corpora









This section shows the basic information for the parallel corpora that have been created: name, language, number of words and number of texts that comprise it. It also allows you to delete a corpus by clicking on the  icon, or go to the corpus query by clicking on the  icon.

Parallel corpus management

[My corpora](#)[Corpus query](#)[Import corpus](#)

 Here you can see all the information regarding your own corpora.

My corpora (2/20)

Name:	Description:	Query:	Delete:
legal_corpus	en  1 texts (153277 words) es  1 texts (160798 words)		
para_corpus	en  2 texts (405 words) es  2 texts (433 words)		

2.2 Corpus query

In this section, you can perform different queries over the parallel corpora that you have created or imported. You just need to select one or several corpora from the list and type the search term. To select several corpora, first select one and then press Ctrl+click on the next corpus that you would like to include in the selection.

In addition, several options can be set for the term, such as “Case sensitive” (makes the search case sensitive), “Diacritics insensitive” (ignore diacritics), “Grammatical category” (only terms matching the selected grammatical category will be shown). You can also choose if you want to perform the search by word form (as it appears in the corpus) or by lemma (lemmatized form of terms).

After clicking “Search” button, fragments matching the typed text in selected corpus will be shown next to the equivalent in the other language corpus fragments.

Corpus query

Corpus name:

legal_corpus [en][es]
para_corpus [en][es]

Term:

budget

☐ Case sensitive
☐ Diacritics insensitive
Grammatical category...

☒ Word form
☐ Lemma

Language:

☐ ES
☒ EN

Search Clear

legal_corpus: 69 results


EN	ES
Article 17 The budget of the Union for the financial year 2004 shall be adapted to take into account the accession of the new Member States through an amending budget taking effect on 1 May 2004.2 .	Artículo 17 El Presupuesto de la Unión para el ejercicio presupuestario 2004 será adaptado , a fin de tener en cuenta la adhesión de los nuevos Estados miembros , mediante un presupuesto rectificativo que entrará en vigor el 1 de mayo de 2004.2 .
Article 17 The budget of the Union for the financial year 2004 shall be adapted to take into account the accession of the new Member States through an amending budget taking effect on 1 May 2004.2 .	Artículo 17 El Presupuesto de la Unión para el ejercicio presupuestario 2004 será adaptado , a fin de tener en cuenta la adhesión de los nuevos Estados miembros , mediante un presupuesto rectificativo que entrará en

2.3 Corpus importation

In this section, you can create your own parallel corpus from a set of text files. The name of these files must be in a specific format: the name of the files which are equivalent in each language must be the same except for the final part of the name, which must include the language code preceded by a low dash (_xx, being "xx" the language code. For example, the Spanish file filename1_es.txt corresponds to the English file filename1_en.txt, the Spanish file file_name2_es.txt corresponds to the English file file_name2_en.txt, etc.). Then, choose a name for the corpus and select the text files from your computer. These files can be in TXT or VRT (VeRticalized Text) format. After pressing the "Import" button, the corpus will be created, and you will be able to consult it in the section Corpus query.


Parallel corpus management

My corpora Corpus query Import corpus

 Create a corpus from your own text files. Type a name and upload your files.

Import corpus

Corpus name:

Files: 






The name of the files which are equivalent in each language must be the same except for the final part of the name, which must include the language code preceded by a low dash (_xx, being "xx" the language code). Example:

Files of the Spanish corpus: `filename1_es.txt; file_name2_es.txt`
Files of the English corpus: `filename1_en.txt; file_name2_en.txt`

3. Dictionary and glossary management

In this section, you can see and manage the glossaries you have created.

3.1. My glossaries

Here you can see basic information about the glossaries you have created (name, description, number of terms) as well as create and delete glossaries. In order to create a new glossary, you just need to type a name and press the "Create" button. In doing so, an empty glossary will be created. You can change the description of any glossary by pressing on the edit icon  which appears in the description when you hover your mouse over it. Then you can either confirm the change  or discard it . To delete a glossary, press on the corresponding delete icon: . In order to add terms to a certain glossary you must go to *Glossary management* section (instructions are in the next section). If you click on the  icon, you will be automatically redirected to the management of the selected glossary.

Glossary management

My glossaries
Glossary management






Here you will be able to see the information regarding your glossaries. You can also create a new one or delete an existing one.

My glossaries (9/20)

+ Create

Name:	Description:	Manage:	Delete:
Pesca submarina	21 terms Glosario de pesca		
Natural disasters	24 terms Glossary about natural disasters terms		
nat_disasters	6 terms Natural disasters glossary		
Covid-19	104 terms		

3.2. Glossary management

In this section you can edit the created glossaries (modify, add, etc.). Just select a glossary from the list and click on "Show". To edit any term in the glossary, press on it or on the edit icon  which appears when you hover the mouse cursor over it. Then you can either confirm the change (pressing  or ENTER) or discard it (pressing  or ESC). If you click on the external resources search icon , a drop-down list with different resources (Google, Wikipedia, Linguee, etc.) will be displayed. When selecting any of these resources, a search for the selected term in that resource will be performed. You can also delete a glossary entry by pressing on the corresponding delete icon , or add a new blank one by pressing the "Add" button.


Moreover, the tool allows both importing terms into glossaries and exporting those glossaries. To import terms to a glossary, press the "Import" button and select the desired file (the allowed formats are XLSX, XLS, CSV and ODS). This file must include the Spanish terms in the first column and their English equivalent in the second column. This way all the terms included in your file will be copied to the current glossary. To export a glossary to an XLS file, just click the "Export" button and the file will be automatically downloaded.

Glossary management

Glossary name:

Glosario de prueba ▾

ShowClear

ES 

EN 

Prueba

trial



+ Add

Export | Import ▾

File formats: XLSX XLS CSV ODS

Seleccionar archivo Ningún archivo seleccionado

Import terms

4. Named Entity Recognition system

In this section you can extract the named entities from any text.

4.1 Named Entity Recognition in text

In this section you can paste a text into the corresponding field, upload a text file (.txt) or use the text extraction tools to get the text from an URL or a PDF file. These tools will show the text in the corresponding field and allow the user to download the text in .txt format. Then, you just need to press the "Extract entities" button and the different types of named entities in the text will be highlighted (PER/PERSON: people, LOC: places, ORG: companies, institutions, etc.). The full list of recognised entity types can be found under the "Label Descriptions" drop-down menu.

ORDINAL

"first", "second", etc.

CARDINAL

Numerals that do not fall under another type.

MISC

Miscellaneous entities.

TXT converter

Named Entity Recognition System

Upload txt

Seleccionar archivo

Ningún archivo seleccionado

Text:

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak was first identified in Wuhan, China, in December 2019. The World Health Organization declared the outbreak a public health emergency of international concern on 30 January, and a pandemic on 11 March. As of 28 May 2020, more than 5.8 million cases of COVID-19 have been reported in more than 188 countries and territories, resulting in more than 359,000 deaths, more than 2.39 million people have recovered.

The virus is primarily spread between people during close contact, most often via small droplets produced by coughing, sneezing, and talking. The droplets usually fall to the ground or onto surfaces rather than travelling through air over long distances. Less commonly, people may become infected by touching a contaminated surface and then touching their face. It is most contagious during the first three days after the onset of symptoms, although spread is possible before symptoms appear, and from people who do not show symptoms.

Common symptoms include fever, cough, fatigue, shortness of breath, and loss of sense of smell. Complications may include pneumonia and acute respiratory distress syndrome. The first few symptoms to onset of symptoms in Wuhan, around December 2019, but may mean dates as late as January 2020. There is no known

Extract entities

Clear

Add selected terms to:

Select glossary...

Add

Text

Table

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of coronavirus disease 2019 DATE (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 CARDINAL (SARS-CoV-2). The outbreak was first ORDINAL identified in Wuhan GPE, China GPE, in December 2019 DATE. The World Health Organization ORG declared the outbreak a public health emergency of international concern on 30 January DATE, and a pandemic on 11 March DATE. As of 28 CARDINAL May 2020 DATE, more than 5.8 million CARDINAL cases of COVID-19 have been reported in more than 188 CARDINAL countries and territories, resulting in more than 359,000 CARDINAL deaths; more than 2.39 million CARDINAL people have recovered.

Terms to be added

- The World Health Organization
- 11 March
- China

After processing the text, you can select the highlighted entities and add them to a glossary, together with a preliminary machine translation, that you can later consult and edit (see Glossary query section). To do this, click on any found entity or select it from the "Table" tab, where entities are shown in an ordered list. Then, select a glossary or type a name to create a new one and click on the "Add" button.

In order to quickly select several terms, you have two options:

- 1) Check "Select all" to select all terms.
- 2) Select a range. To do this, check an initial checkbox and, while pressing the SHIFT button, check an ending checkbox. All checkboxes in between will be automatically checked.

4.2 Named Entity Recognition in corpus

In this section you can upload several text files (.txt) by clicking "Add" button. Then, you just need to press the "Extract entities" button and a table of named entities in the text grouped by type that can be filtered will be shown (PER/PERSON: people, LOC: places, ORG: companies, institutions, etc.). The full list of recognised entity types can be found under the "Label Descriptions" drop-down menu.

14

Named Entity Recognition System

Upload corpus in TXT format Select language:

File name	Add
TEXT001	-
TEXT002	-
TEXT003	-

☐ Spanish
☒ English

Extract entities Clear Export entities

Add selected terms to: Select glossary... Add

Filter labels

ORG LOC

<input type="checkbox"/> Select all	Label
<input type="checkbox"/> China	LOC
<input type="checkbox"/> Wuhan	LOC

After processing the text, you can select the desired entities and add them to a glossary, together with a preliminary machine translation, that you can later consult and edit (see Glossary query section). To do this, click on any found entity, select a glossary (or type a name to create a new one) and click on the “Add” button.

In order to quickly select several terms, you have two options:

- 1) Check “Select all” to select all terms.
- 2) Select a range. To do this, check an initial checkbox and, while pressing the SHIFT button, check an ending checkbox. All checkboxes in between will be automatically checked.

5. Text Summarisation system

In this section you can summarize a text. For this purpose, you can either load one or several plain text files (.txt), paste the text directly into the corresponding field, or use the text extraction tools to get the text from an URL or a PDF file. These tools will show the text in the corresponding field and allow the user to download this text in .txt format. Then, choose the type of summary by selecting the number of words or the percentage of the original text that it will take up. After pressing the "Summarize" button, the result will be shown next to the original text.

Text summarisation system

Advanced summarization ▼

Text:

United Kingdom of Great Britain and Northern Ireland for Cooperation on the uses of Atomic Energy for Mutual Defense Purposes. It allows the United States and the UK to exchange nuclear materials, technology and information. While the US has nuclear co-operation agreements with other countries, including France and other NATO countries, this agreement is by far the most comprehensive. Because of the agreement's strategic value to Britain, Harold Macmillan (the Prime Minister who presided over the United Kingdom's entry into the agreement) called it "the Great Prize".[1]

The treaty was signed on 3 July 1958, after the Soviet Union shocked the American public with the launch of Sputnik on 4 October 1957, and the British hydrogen bomb programme successfully tested a thermonuclear device in the Operation Grapple test on 8 November. The Anglo-American Special Relationship proved mutually beneficial, although it was never one of equals after the World Wars; the United States was (and is) far larger than Britain both militarily and economically. Britain soon became dependent on the United States for its nuclear weapons, as it lacked the resources to produce a range of designs. The treaty allowed American nuclear weapons to be supplied to Britain through Project E, for use by the Royal Air Force and British Army of the Rhine.

Summarize

Clear

The 1958 US–UK Mutual Defense Agreement, or UK–US Mutual Defence Agreement, is a bilateral treaty between the United States and the United Kingdom on nuclear weapons co-operation. The treaty's full name is Agreement between the Government of the United States of America and the Government of the United Kingdom of Great Britain and Northern Ireland for Cooperation on the uses of Atomic Energy for Mutual Defense Purposes. It allows the United States and the UK to exchange nuclear materials, technology and information. While the US has nuclear co-operation agreements with other countries, including France and other NATO countries, this agreement is by far the most comprehensive. Because of the agreement's strategic value to Britain, Harold Macmillan (the Prime Minister who presided over the United Kingdom's entry into the agreement) called it "the Great Prize".[1] The treaty was signed on 3 July 1958, after the Soviet Union shocked the American public with the launch of Sputnik on 4 October 1957, and the British hydrogen bomb programme successfully tested a thermonuclear device in the Operation Grapple test on 8 November. The Anglo-American Special Relationship proved mutually beneficial, although it was never one of equals after the World Wars; the United States was (and is) far larger than Britain both militarily and economically. Britain soon became dependent on the United States for its nuclear weapons, as it lacked the resources to produce a range of designs. The treaty allowed American nuclear weapons to be supplied to Britain through Project E, for use by the Royal Air Force and British Army of the Rhine.

Module II

This module is intended to be used during the interpretation process. The different sections are detailed below.

6. Note-taking

This section includes tools which can be useful during the note-taking process.

6.1 Automatic note-taking



Before using the microphone, kindly grant the browser the necessary permission. You will see a crossed-out camera icon on the right-hand side of the address bar. Please click on the icon and follow the instructions. See *Appendix: Microphone Configuration*.

This section provides automatic note-taking. The system will generate a text from an oral speech and will detect automatically the named entities and numbers appearing in the speech. To do

so, you only have to select the language of the speech and the output format. If you want to get a list of named entities and detected numbers, select the 'Entities Only' option. If you prefer to get the full text highlighting the entities and numbers, select the "Marked text" option. By pressing the "Start" button, the system will start transcribing everything that is said through the microphone until the "Stop" button is pressed.

The numbers are not shown in isolation. They appear next to the noun to which they refer. In this way, you will be able to have important information, such as the physical magnitude (e.g. 20 tons), but also the common nouns (e.g. 10 rockets), as can be seen in the following capture:

Automatic note taking

Language: ☐ ES ☒ EN ☐ EN

Output format: ☒ Just entities ☐ Marked text

Named entities:

- 11 [CARDINAL]
- 20 tons [CARDINAL]
- 520000 [CARDINAL]
- NASA [ORG]
- first [ORDINAL]
- 10 rockets [CARDINAL]
- Earth [LOC]
- NASA [ORG]

Text:

The NASA has a new and revolutionary plan.

The first step is to launch up to 10 rockets to the Earth orbit.

The NASA knows them as Ares.

7. Machine translation

In this section you can make a translation from Spanish to English or from English to Spanish automatically, so it could be used for any of the stages of an interpretation. You just need to enter the text to be translated in the "Text" field, choose the language (ES->EN or EN->ES), the translation engine (*Apertium*² or the machine translation integrated VIP tool) and press the "Translate" button. The result will be shown in the "Translation" field. If you are using the machine translation integrated VIP tool, any word that cannot be translated will appear in red.


² <https://apertium.org/>

If you press on any of the words highlighted in red, a context menu will be shown and you will be able to choose either searching the term in an external searcher (*Linguee*, *DeepL*, *Google translate*) or confirming the term as correct.

If you detect that the translation is not completely correct, you can modify the result and send it to the system by pressing the "Send" button (this button appears when modifying the resulting text); in this way you can help us to improve the machine translation integrated VIP tool.

The screenshot shows a web interface titled "Machine translation". It has two sections: "Language:" and "Translation engine:". Under "Language:", there are two options: "ES → EN" (Spanish to English) and "EN → ES" (English to Spanish). Under "Translation engine:", there are two options: "Apertium" and "VIP translator (experimental)". Below these is a "Text:" input field containing a paragraph about the COVID-19 pandemic. Below the text field are "Translate" and "Clear" buttons. Below these is a "Translation:" output field showing the translated text. A context menu is open over the word "coronavirus" in the translated text, showing options: "Search in", "Google translate", "Linguee", "DeepL", and "Accept". At the bottom of the interface, there is a "Send" button and a note: "For helping to improve this system you can correct the translation by pressing this button: Send".

8. Glossary query

In this section you can use the glossaries previously created in Module I. You just have to select the desired glossaries, press on "Load Glossaries" and a search field will appear. When typing a term, the results will start appearing automatically in a quick and convenient way. If you check the "Fuzzy match" option, results with partial coincidences will be shown too. In addition, you can perform the search using your voice, by clicking the  icon or pressing the "s" key.

Pesca submarina
21 terms
Glosario de pesca

Natural disasters
24 terms
Glossary about natural disasters terms

nat_disasters
6 terms
Natural disasters glossary

Covid-19
104 terms
Coronavirus glossary

teletrabajo
11 terms
Glosario sobre el trabajo a distancia

coronavirus
10 terms
No description

Glossary WORKSHOP
78 terms
No description

ener_glo
10 terms
No description

miglosario
8 terms
No description

Load glossaries

✓ Covid-19 (104) ✓ coronavirus (10)

Search: d| ☐ Fuzzy match

6 results.

ES	EN
días	days
enfermedad	disease
disponible	available
diario	daily
diferente	different
murió	died

Module III

This module aims to take advantage of the information used in the previous modules to create several types of exercises in order to practice and improve the different skills of the interpreter. Moreover, it can also be used to prepare an interpreting assignment.

9. Exercises



Before using the microphone, kindly grant the browser the necessary permission. You will see a crossed-out camera icon on the right-hand side of the address bar. Please click on the icon and follow the instructions. See *Appendix: Microphone Configuration*.

This system offers four different types of exercises. To generate the exercises automatically, you only have to choose the type of exercise, the number of exercises and the configuration, which is different for each type of exercise:

9.1 Anticipation

In this type of exercise, you choose the corpus you would like to use and the gap content, where you can choose either “Anything” (random terms from the selected corpus), “Only terms from my glossaries” (terms from the selected corpus that can be found in any of your glossaries if

possible) or use a custom pattern (Adj+N, V+N, etc.). By clicking on the “Generate” button, you will obtain a list of small texts, which have been extracted from the selected corpus, in which one of the terms has been omitted according to the selected option in “Gap content”. If you choose the “Pattern” option, the omitted term will be the one indicated by the number in “Gap position”. For instance, if you create a pattern such as N + Prep + N and set 2 for the “Gap position”, the omitted term will be a preposition. To complete the exercise, you must find out which term is missing and type it in the corresponding field.

The screenshot shows the 'Training module' interface. It has four main sections: 'Exercise type', 'Number of exercises', 'Corpus name', and 'Configuration'.
 - 'Exercise type' has radio buttons for 'Anticipation' (selected), 'Numbers', 'Sight Translation', and 'Glossaries'.
 - 'Number of exercises' has a text input field with the value '10'.
 - 'Corpus name' has a dropdown menu with options: 'CoronaVirus_en', 'medicina_es', 'micorpus_es', and 'Natural_Disasters_en' (selected).
 - 'Configuration' has a 'Gap content' section with radio buttons for 'Anything', 'Only terms from my glossaries', and 'Adj + N' (selected). Below this is a pattern builder with buttons for 'V', 'N', 'Adj', and 'Prep'. A 'Delete' button is also present. At the bottom of this section is a 'Gap position' input field with the value '1'.
 At the bottom of the configuration area are 'Generate' and 'Clear' buttons.
 Below the configuration area, there are three numbered text examples:
 1. ...from natural processes of the Earth A blizzard in Maryland in 2009 A rope tornado in its dissipating stage , Tecumseh , Oklahoma . A [] wildfire in California . 1755 copper engraving depicting Lisbon in ruins and in flames after the 1755 Lisbon earthquake . A tsunami overwhelms the ships in...
 2. ...event will not rise to the level of a disaster if it occurs in an area without vulnerable population . [3][4] In a [] area , however , such as Nepal during the 2015 earthquake , an earthquake can have disastrous consequences and leave lasting damage , which can require...
 3. ...generally meaning the cooled ash , may form a cloud , and settle thickly in nearby locations . When mixed with water this forms a



9.2 Numbers

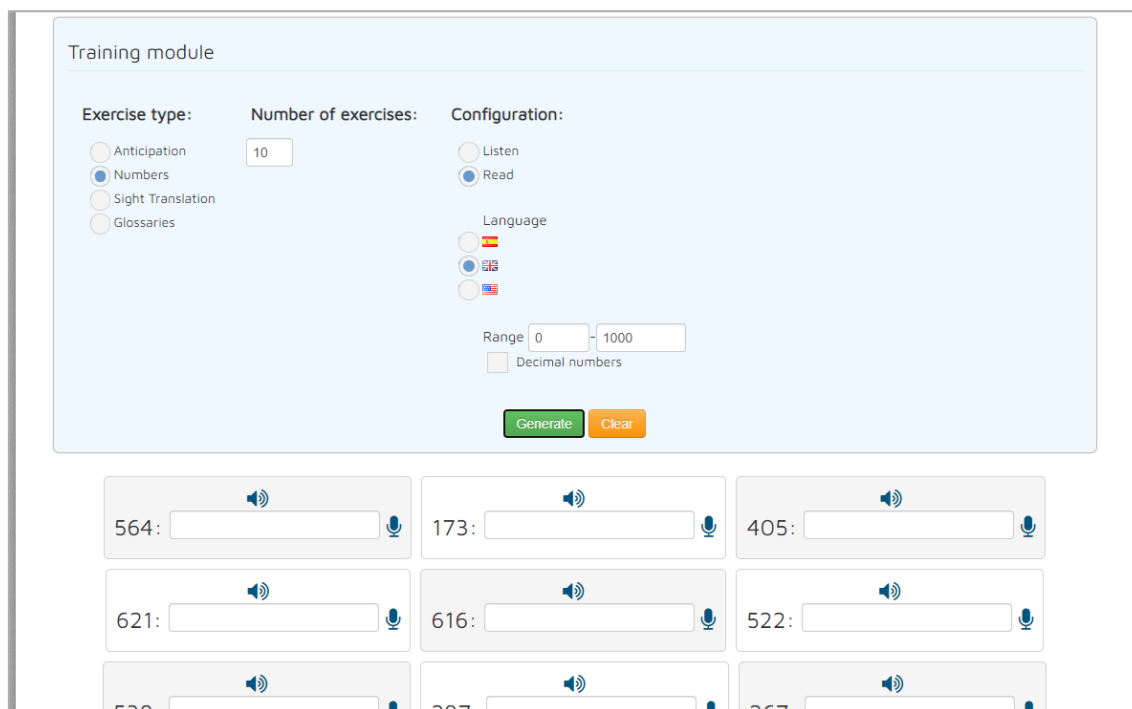
For this type of exercise, you need to choose the language³, the range of numbers and whether you would like to "Listen" to or "Read" the numbers. By pressing the "Generate" button, random numbers will be created and used for the exercises.

In the "Listen" mode, you must press the 🔊 icon and enter the number that you hear into the text fields (the number is reproduced in the selected language). You can also say the number aloud after clicking the 🗣️ icon.

3



There are two variants for the English language: British English and American English.

In the "Read" mode, written numbers will appear next to text fields that must be filled in by reading the number (in the selected language) aloud after pressing the  icon. If you would like to listen how the number is pronounced in the selected language, you can press the  icon which is above the text field.



The screenshot shows a "Training module" interface. At the top, there are three sections: "Exercise type:", "Number of exercises:", and "Configuration:". Under "Exercise type:", there are radio buttons for "Anticipation", "Numbers" (selected), "Sight Translation", and "Glossaries". Under "Number of exercises:", there is a text input field containing "10". Under "Configuration:", there are radio buttons for "Listen" and "Read" (selected). Below these are language selection options with flags for Spanish, British English (selected), and American English. There is also a "Range" input field set to "0 - 1000" and a checkbox for "Decimal numbers". At the bottom of the configuration section are "Generate" and "Clear" buttons. Below the configuration section is a grid of exercise boxes. Each box contains a number followed by a colon and a text input field. Above each input field is a speaker icon for audio playback, and to the right of each input field is a microphone icon for voice recording. The visible numbers in the grid are 564, 173, 405, 621, 616, 522, 529, 207, and 267.

9.3 Sight Translation

In this type of exercise, select the language⁴ and the parallel corpus that will be used to generate the exercises. By clicking on the "Generate" button, two fields will appear for each exercise. When the  icon is pressed, the text (in the source language) will appear in the first field and the voice detection will be activated. Everything you say will be recorded (in the target language) and will be written in the second field. When you have finished, press the  icon to end the voice detection. Press the "Check" button and you will obtain the parallel text together with a percentage that measures how similar the resulting text is to the output of the sight translation.

⁴ British English and American English are available for the English language.

Generate
Clear

EuroParl

I welcome, however, President Kibaki's recent move to revise this media law and his gesture to consider amendments to legislation proposed by members of the media. Freedom of expression is a fundamental human right, as stated in Article 19 of the Universal Declaration of Human Rights. I call on the Kenyan Government to initiate a stakeholder consultation in order to build consensus on how to better regulate the communications industry without interfering with press freedom and without infringing rights contained within the Universal Declaration. Finally, I would like to underline the need to adjust the culture of impunity in Kenya in order to bring those responsible for the post-election violence a year ago to justice. I call for the setting-up of an independent commission consisting of local and international legal experts who would carry out investigations and prosecutions into the violent events following the flawed elections in December 2007.


Universal de Derechos finalmente me gustaría subrayar la necesidad de luchar contra la cultura de impunidad que existe en Kenia para llevar ante los tribunales a los responsables de la violencia posterior a las elecciones organización de una comisión independiente concurso por el centro Leganés locales e internacionales que realicen investigaciones y presenten cargos por los hechos violentos que sucedieron a las elecciones fraudulentas diciembre 2007

Check

Accuracy: 92.94%

No obstante, acojo positivamente el paso dado por el Presidente Kibaki para la revisión de esta ley sobre los medios y su gesto de considerar enmiendas a la legislación propuestas por los miembros de los medios de comunicación. La libertad de expresión es un derecho humano fundamental, según establece el artículo 19 de la Declaración Universal de los Derechos Humanos. Pido al Gobierno de Kenia que inicie consultas con las partes interesadas para llegar a un consenso sobre la mejor forma de regular el sector de las comunicaciones sin interferir con la libertad de prensa y sin infringir los derechos contenidos en la Declaración Universal de Derechos. Finalmente, me gustaría subrayar la necesidad de luchar contra la cultura de impunidad que existe en Kenia para llevar ante los tribunales a los responsables de la violencia posterior a las elecciones. Pido la organización de una comisión independiente compuesta por expertos legales locales e internacionales que realicen investigaciones y presenten cargos por los hechos violentos que sucedieron a las elecciones fraudulentas de diciembre de 2007.

9.4 Glossaries

For this type of exercise, you need to choose a glossary (previously created in Module I) and the language (ES->EN or EN->ES). By pressing the "Generate" button, a list of pairs of terms will be shown where one of them is blank. It can be filled in by typing directly in the field or by pressing the  icon and saying the answer out loud.

Training module

Exercise type:

☐ Anticipation
 ☐ Numbers
 ☐ Sight Translation
 ☒ Glossaries

Number of exercises:

Glossary name:

Pesca submarina

Natural disasters

nat_disasters

Covid-19

Configuration:

☒ EN-ES
☐ ES-EN
☐ ES-EN

Generate
Clear

ES < EN

cangaroo

eruption

wildfire

mortality

new

eruptions

Impact

Heat

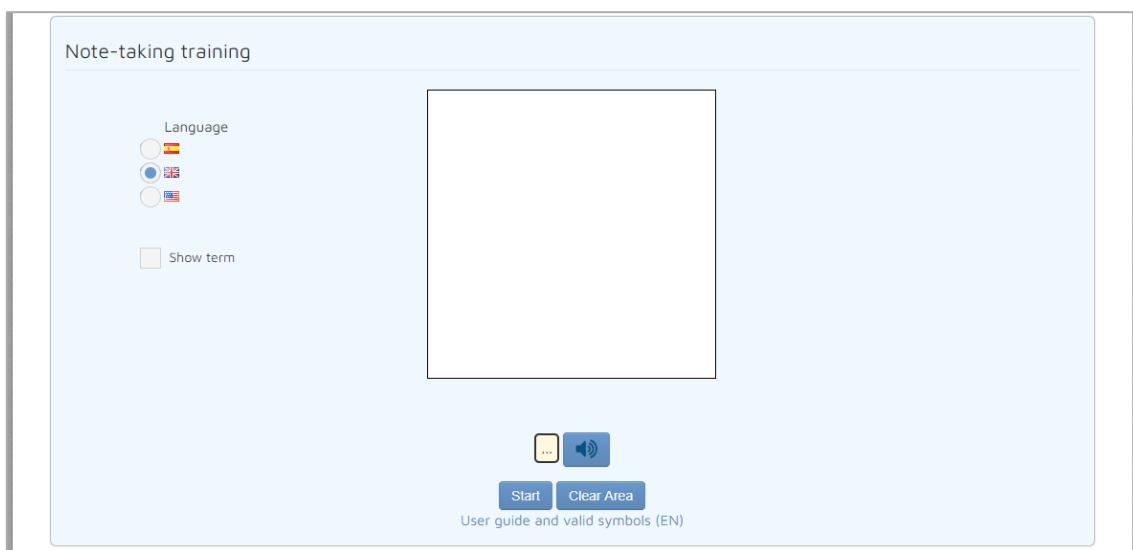
earthquake

blizzard

22

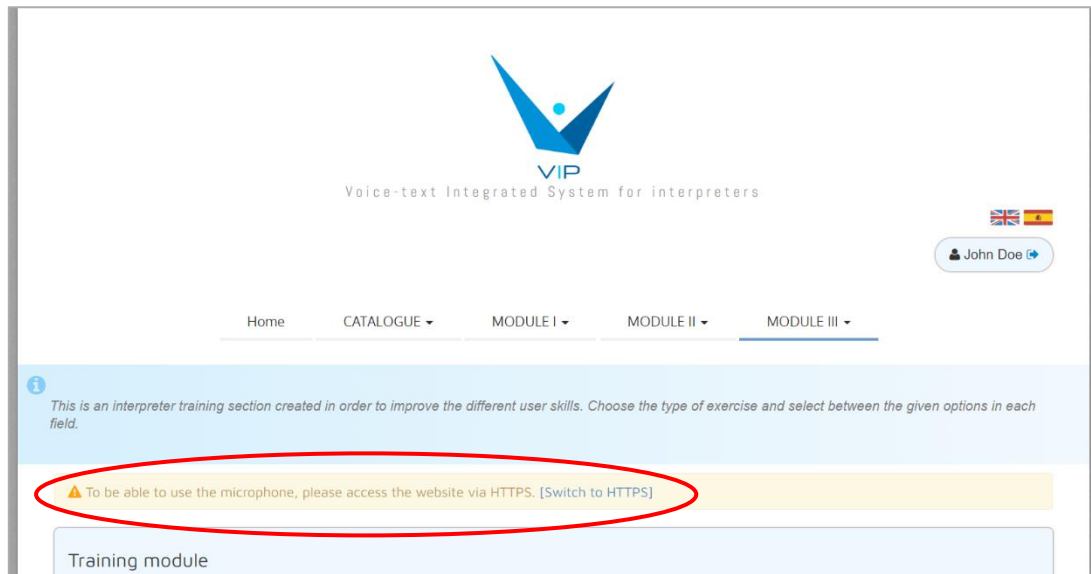
10. Symbols

In this section you can practise the symbols used during an interpretation. If you click on "List of symbols", a list of available concepts will be displayed. After clicking the "Start" button, you will listen to one of the concepts from the list. In addition, by checking the "Show term" checkbox, the written concept will appear. You must draw the corresponding symbol to the given concept and, in case you do it correctly, another concept will be offered. Under the drawing area, a bar indicating the accuracy degree of the drawn symbol will be shown (the longer, the more similar to the correct symbol). If you draw it incorrectly several times, the system will show you the solution.

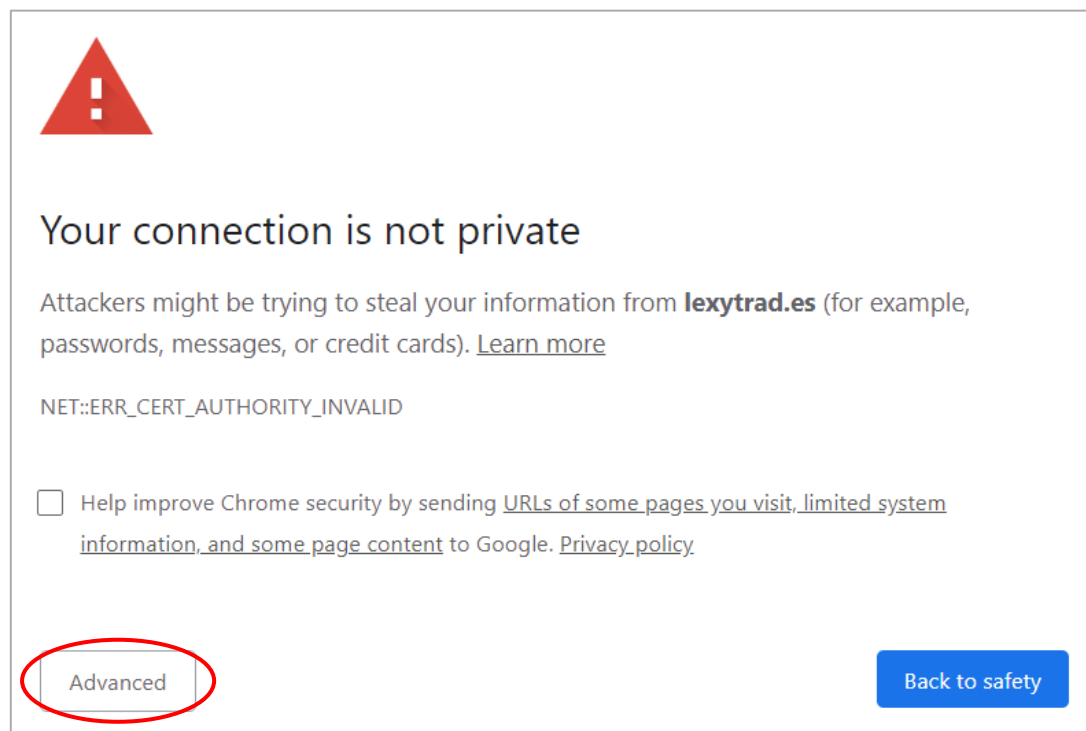


Appendix: Microphone configuration

In order to use the microphone in the system, you must access the platform via the secure HTTPS protocol. Otherwise, if you are accessing via HTTP, the following warning message will appear in the sections where the use of the microphone is required:



To access via HTTPS, click on "Switch to HTTPS" and the browser will display the following warning message:



To continue, press the "Advanced" button and the following message will be displayed:

This server could not prove that it is **lexytrad.es**; its security certificate is not trusted by your computer's operating system. This may be caused by a misconfiguration or an attacker intercepting your connection.

[Proceed to lexytrad.es \(unsafe\)](#)

Finally, click on "Access lexytrad.es (non-secure site)" and you will return to the system, where you will be able to use the microphone.

Appendix: Verticalized text (.vrt)

Verticalized text (.vrt), also known as one-word-per-line, is a format in which each token is in a different line. It can contain additional information related to each token, as the grammatical category or lemma (separated by tabulations). Structures such as sentences or paragraphs can be set too using XML tags.

If we take this text as example:

This is the first sentence. This is the second.

Verticalized text would be:

```
This
is
the
first
sentence
.
This
is
the
second
.
```

We can include additional information (such as grammatical category or lemma) next to each token, leaving one tabulation separation:

This	DET	this
is	VER	be
the	DET	the
first	ADJ	first
sentence	NOU	sentence
.	PUN	.
This	DET	this
is	VER	be
the	DET	the
second	ADJ	second
.	PUN	.

In addition, we can set structures such as sentences (<s>) or paragraphs (<p>) by using XML tags:

<p>		
<s>		
This	DET	this
is	VER	be
the	DET	the
first	ADJ	first
sentence	NOU	sentence
.	PUN	.
</s>		
<s>		
This	DET	this
is	VER	be
the	DET	the
second	ADJ	second
.	PUN	.
</s>		
</p>		