

Manual for the *Lexical Toolbox* of *Lexogoth*
A practical bridge between language teaching and corpus linguistics

Welcome to the Lexical Toolbox of Lexogoth

Table of Contents

1. Introduction and starting the Lexical Toolbox Lexogoth module

2. Overview of the different modules

2.1 Module 1: Database functions

2.1.1 Creating a new database

2.1.2 Editing an existing database

2.1.3 Practicing with databases

2.1.4 Some key editing functions:

- (1) importing entries from Lexogoth
- (2) batch deletion
- (3) removing duplicates
- (4) sorting items by difficulty
- (5) shuffling entries randomly
- (6) adding/showing notes from the corpus
- (7) advanced database search
- (8) n-gram analysis
- (9) searching for word combinations (concordancer)

2.1.5 Merging databases

2.1.6 Exporting database

2.2 Module 2: Importing / Exporting

2.2.1 Importing entries from Lexogoth

2.2.2 Exporting database

2.2.3 Exporting test

2.2.4 Exporting test results

2.2.5 Exporting context module

2.3 Module 3: Testing

2.3.1 Creating tests and the different types of questions

2.3.2 Editing tests and importing data from a database or the Lexogoth corpus

2.3.3 Taking a test

2.3.4 Loading results

2.3.5 Merging tests

2.4 Module 4: Context Modules

2.4.1 Creating and editing context modules

2.4.2 Practicing context modules

2.4.3 Merging context modules

2.4.4 Exporting context modules

2.5 Module 5: Collocations and Syntax

2.5.1 Purpose of the module

2.5.2 Four types of searches: collocations, co-occurrences, syntactic patterns, and Metadata & Notes

2.5.3 Selecting items and reading notes

2.5.4 Exporting results to a text file, a database (lexdb), or a context module

2.6 Module 6: Lexgoth Workbench (text analysis)

2.6.1 What you can do with the Workbench: a concise overview

2.6.2 The main screen and the two side panels: overview of functions

2.6.3 The menu bar

2.6.3.1 The menu bar: file and formatting

2.6.3.2 The menu bar: Workbench

2.6.3.3 The menu bar: the Analyze button

2.6.3.4 The menu bar: the Statistics button

2.6.3.5 The menu bar: the Search in text button

2.6.4 The context menu (mouse click in the text field)

2.6.5 Left panel & detachable side panels

2.6.5.1 The Search zone

2.6.5.2 The Show windows zone

2.6.5.3 The Basket

2.6.5.4 The Export zone

2.6.5.5 Mode: reading or editing

2.6.6 Using grammatical information and footnotes while reading

2.6.7 For which didactic or linguistic purposes can the Workbench be used?

2.6.8 Technical explanation of the POS-tagger

3. Tips for smooth usage

4. Technical data

1. Introduction and starting the Lexical Toolbox Lexogoth module

Welcome to the Lexogoth Lexical Toolbox! As an integral part of Lexogoth, this module allows both students and teachers of French to work with Lexogoth in a targeted and differentiated manner. With the Toolbox, our aim is to provide teachers and students with concrete, practical tools to practice and teach vocabulary and lexical patterns by utilizing authentic corpus data.

With the six modules available in the Toolbox, users can:

- Create, edit, practice, and study lexical databases based on examples from Lexogoth.
- Study, edit, and create contextual learning blocks based on the databases.
- Convert created databases into text files (RTF format) for easy editing with standard word processing programs, and merge different databases.
- Create, edit, and take tests.
- Export created tests and test results to well-organized text files for later consultation and/or editing.
- Analyze, read, and study arbitrary French texts, automatically assign POS tags (parts of speech) to all words in a text, automatically recognize grammatical tenses, and perform advanced searches in text with an integrated search function for the Lexogoth corpus.

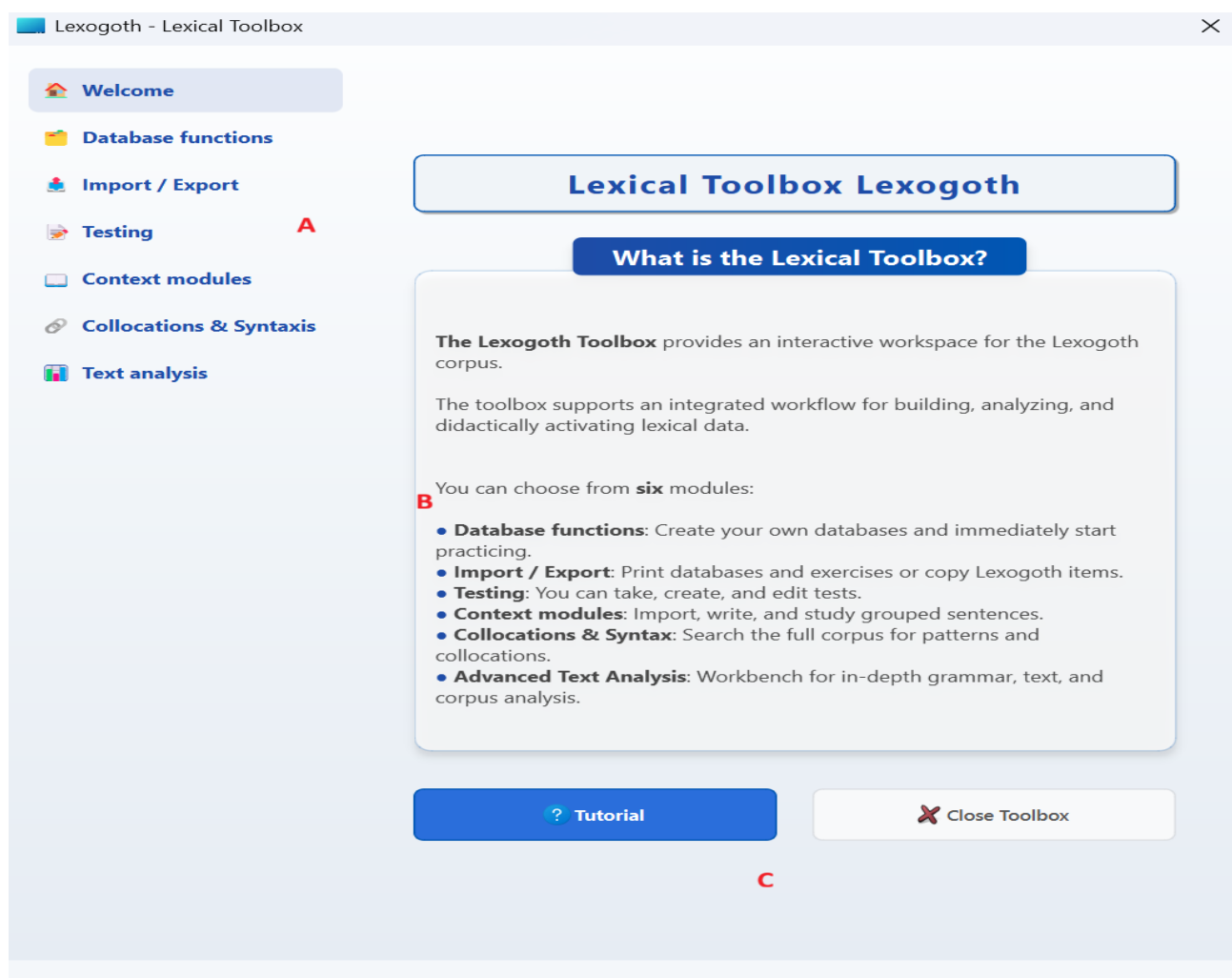
In Chapter 2, we provide a detailed overview of the various modules and demonstrate to users how they can be applied in studying or teaching. Chapter 3 contains a series of tips and tricks to work smoothly with the Toolbox.

Finally, please note that you can launch the Toolbox from within Lexogoth by clicking on **Open Lexogoth Database Editor** in the **File** menu, or by clicking on the leftmost icon in the toolbar.



Good luck !

2. Overview of the different modules



In the photo above, you can see the welcome page of the Toolbox. In part A of the screen, you will find the index, from which the user can click on one of the modules to activate it. These modules are discussed in detail in 2.1, 2.2, and 2.3. On the center screen, the user will find a brief explanation of the capabilities of each module. In part C of the welcome screen, you will find the buttons **Tutorial** and **Exit Program**.

The **Tutorial** button opens the current PDF file so that the user can read in more detail how the Toolbox should be used, while **Exit Program** simply closes the Toolbox.

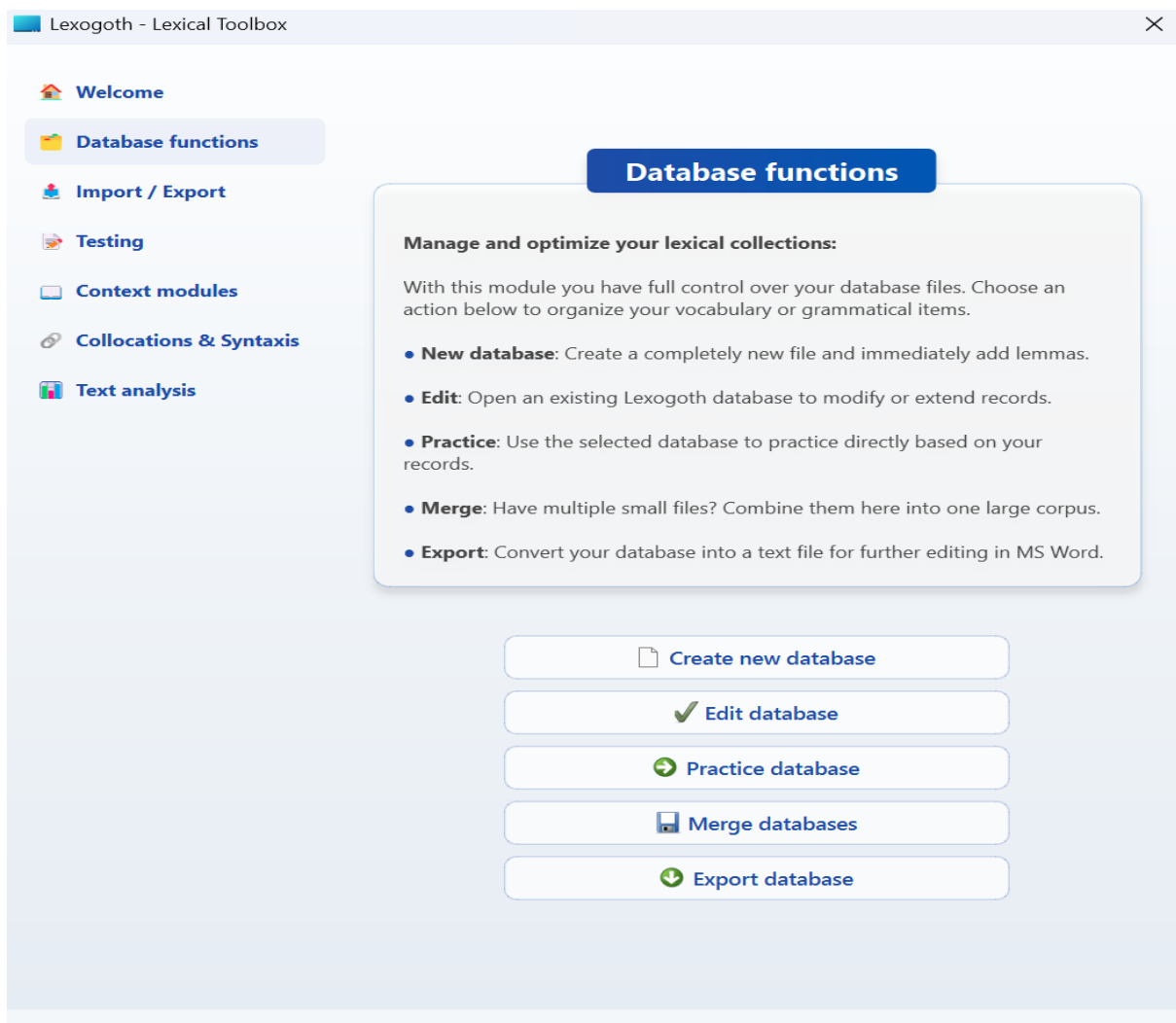
Now a brief explanation of the different modules.

2.1 Module 1: Database functions

In the module *database functions* users can:

- create a new database
- edit an existing database
- practice with a database

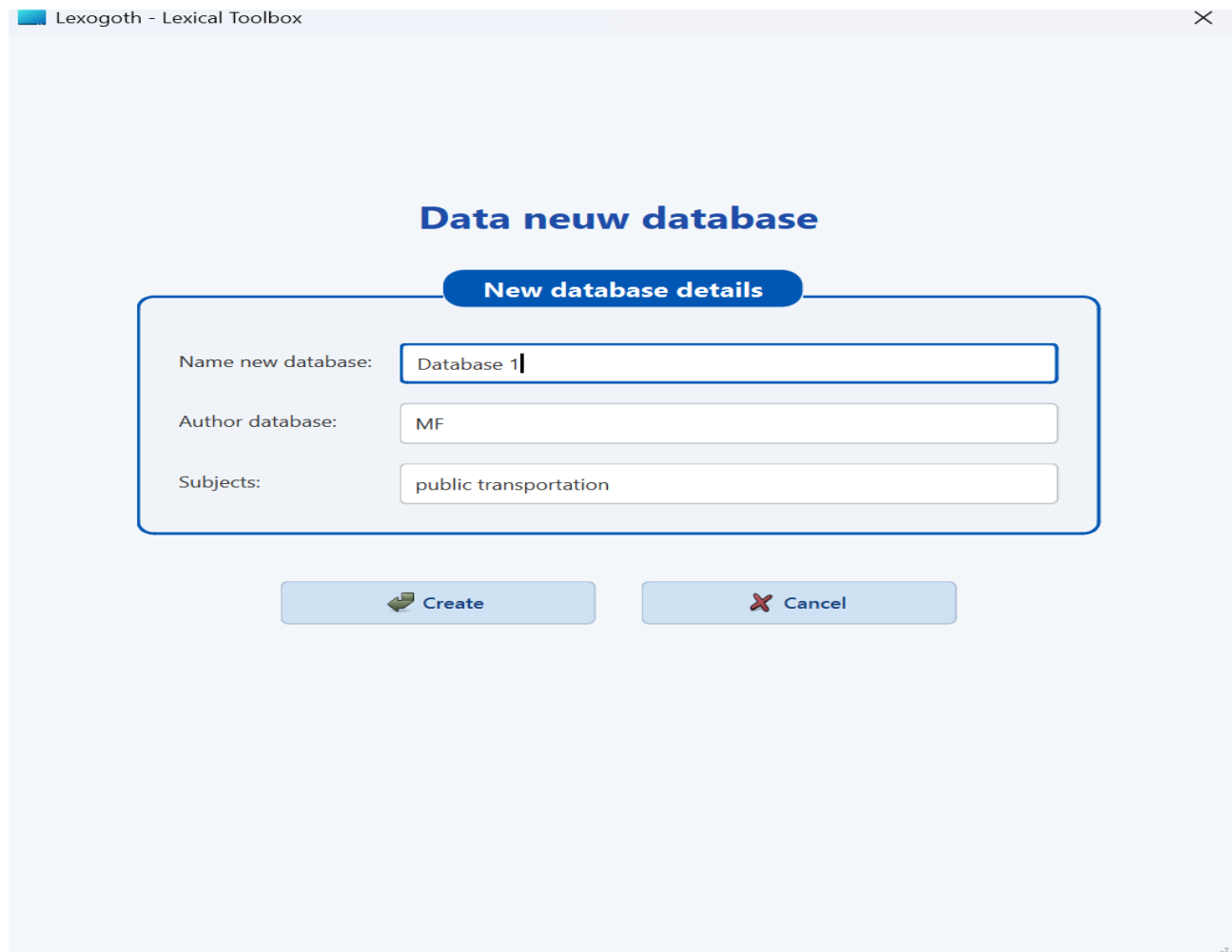
Each of these modules is activated by pressing the corresponding button (see red arrow on the photo below):



2.1.1 Create new database

To create a new database, the user must click “*Create new database*” after which a screen is opened where the user must enter the following data:

- Name of the new database
- Author of the database
- Subjects (i.e. what is the database about: *using the subjunctive after verbs of wishing, vocabulary concerning ‘electric driving’...*)



Lexogoth - Lexical Toolbox

Data neww database

New database details

Name new database: Database 1

Author database: MF

Subjects: public transportation

Create Cancel

After the user has filled in all the data, he presses the “*Create*” button and a Windows-window is opened in which the user must select a folder in which his database will be saved. Once this has happened, the screen appears on which the user can start entering entries.

The input screen consists of 3 main parts (named A, B, C in the illustration and description below):

The screenshot shows the 'Lexogoth - Lexical Toolbox' application window. The title bar says 'Lexogoth - Lexical Toolbox'. The main window has a title 'Database bewerken en lezen' and a close button. Below the title is a 'Database info' tab. This tab contains the following fields: 'Name database: Database 1', 'Author: MF', 'Subjects: public transportation', 'Current item: 1 of 0', and 'Date creation: 23/05/2026'. To the right of these fields is a grid of icons. Below the 'Database info' tab is an 'Entry' tab. This tab contains two large text input fields: 'French (lemma, collocation, sentence):' and 'English translation:'. Below these fields is a 'Comments:' field with a small icon. At the bottom of the window are five buttons: 'Previous item', 'Next item', 'Save database', 'Delete entry', and 'Close database'.

Database info (A):

- Name database: Database 1
- Author: MF
- Subjects: public transportation
- Current item: 1 of 0
- Date creation: 23/05/2026

Entry (B):

- French (lemma, collocation, sentence):
- English translation:
- Comments:

Action buttons (C):

- Previous item
- Next item
- Save database
- Delete entry
- Close database

- **Part A, Database Info**, contains data such as the name of the database, the author, the subjects, the position of the current item, the date of creation, and the total number of items present in the database. The icons on the right will be discussed in detail later.
- **Part B, Entry**, contains the input fields: French (lemma, collocation, sentence), English translation, Additional contexts and collocations, and Notes. The first two fields, French and English translation, are mandatory, while the other fields are optional.
- **Part C, Action buttons**: The "Previous item" and "Next item" buttons are used to navigate through the database, the "Delete entry" button is used to remove an added entry from the database, the "Save database" button saves the database in its current form, and finally, the "Close database" button saves and closes the database. Please note that after adding an

entry and pressing the "Previous item" / "Next item" buttons, the database file is automatically updated and saved.

Part A contains metadata about the database and cannot be edited. In the French field, the user enters a French sentence, collocation, or word, followed by the English translation in the field below. The "Additional contexts" and "Notes" boxes do not need to be filled in, but serve to add extra examples or chunks, as well as comments regarding usage, language register, etc. To create a valid entry, the French and English translation fields must always be filled in.

Once the user has finished entering an entry, they click on "Next Item," and the typed item is automatically saved to the file so that no information can be lost. You can also browse back through the database using the "Previous item" button. Once the database is complete, the user can press "Save database" or directly click "Close database" to save everything and exit. Note that "Close database" automatically saves everything one last time.

2.1.2 Editing a database

You can of course also edit existing **databases**. If you press the button *Edit database* within the *Database functions* screen, you will get a Windows popup asking to open a database file. Please note, the Toolbox only opens database files that were created with the Toolbox (International version of Lexogoth) and that have the correct header (information package at the top of a file); the file extension is ".lexdb". Once the database is opened, the user can go through it with the same buttons as above and make changes or add new entries:

The screenshot shows the 'Lexogoth - Lexical Toolbox' application window. The title bar reads 'Lexogoth - Lexical Toolbox'. The main window has a header 'Database bewerken en lezen' and a sub-header 'Database info'. Below this, there is a form with the following fields:

- Name database:** Database 1
- Author:** MF
- Subjects:** Onderwerpen: public transportation
- Current item:** 461 of 493 (with a red arrow pointing to the number 461)
- Date creation:** 23/05/2026

To the right of these fields is a vertical toolbar with icons for adding, deleting, and editing entries, as well as a search icon.

Below the 'Database info' section is the 'Entry' section. It contains two main text areas:

- French (lemma, collocation, sentence):** Vous pouvez enlever les restes de l'autocollant et les traces de celui-ci avec de l'huile végétale ou de l'acétone.
- English translation:** You can remove the remains of the sticker and the traces of it with vegetable oil or acetone.

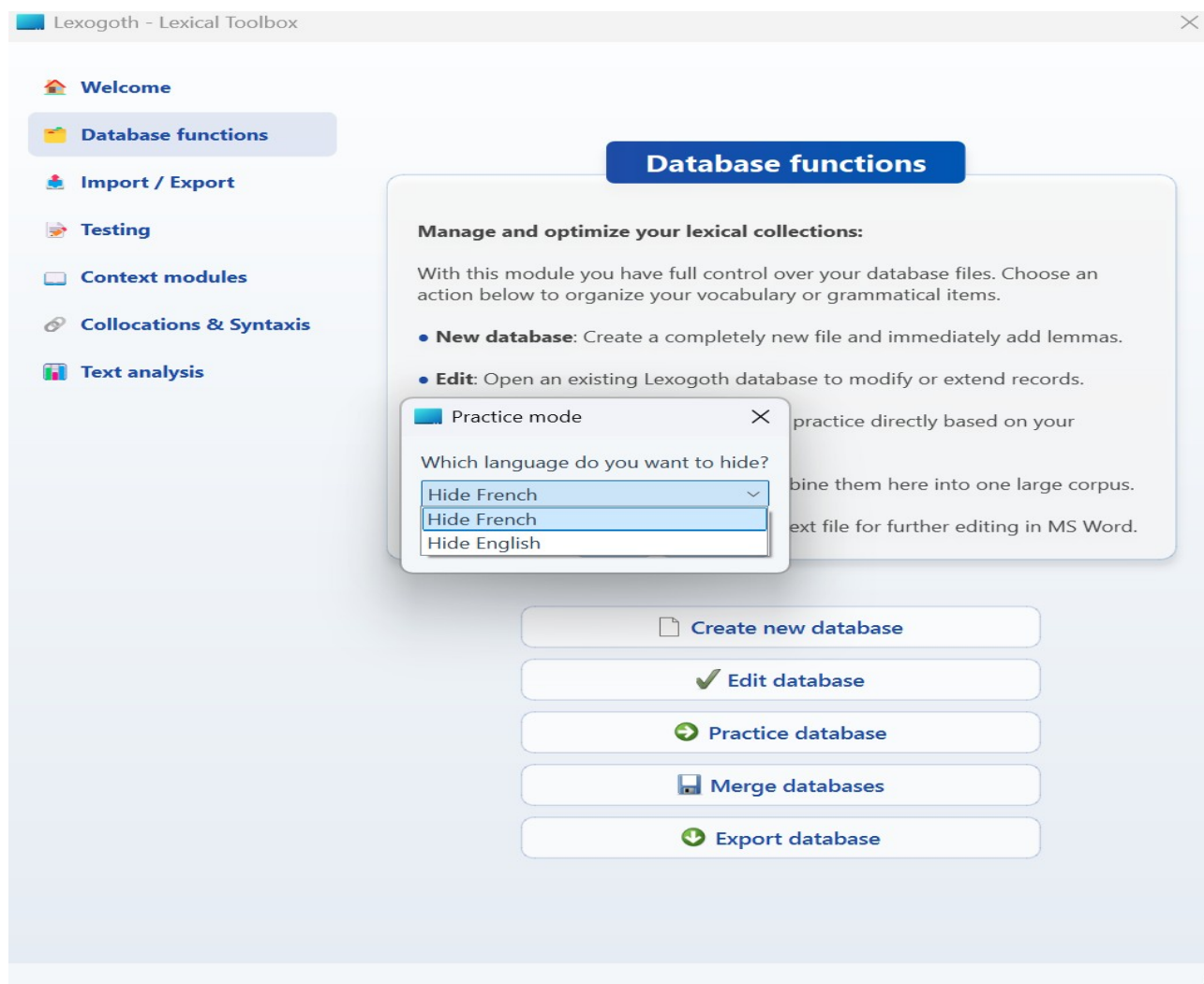
Below these text areas is a 'Comments' section with a text input field labeled 'Comments...'.

At the bottom of the window is a navigation bar with five buttons: 'Previous item', 'Next item', 'Save database', 'Delete entry', and 'Close database'.

To make navigating through large databases easier, you can also enter a (valid) index number via the “*Current item*” field to go directly to a desired item; this way you do not always have to click on *Next* / *Previous*.

2.1.3 Practicing with a database

Users who have created their databases can also practice them by choosing the option Practice database from the Database functions panel. As soon as this option is chosen, the user gets a small popup asking which language he wants to hide (English or French). This means that a database is loaded but either the English translation or the French translation is omitted so that the user can practice:



Suppose we choose to hide English, then the database starts on the first item:

Lexogoth - Lexical Toolbox

Oefenmodus – vertaal het ontbrekende deel

Database info

Name: Database 1 **Author:** MF **Subjects:** Onderwerpen: public transportation
Items: 4 / 493 **Score:** 0




Exercise

French (lemma, collocation, sentence):

English translation:

On dénombre 34 000 accidents de la circulation faisant au total 42 000 victimes, dont 38 000 blessés légers, 3000 blessés graves et 500 morts.

<Type your English answer here...>

Correct answer:

Comments:

Click here to show the correct answer...

Previous item Next item Correct Incorrect

Close exercise Restart

As you can see, the chosen language (English in this case) is unchecked and in this practice mode, the intention is now for the user **to practice in the following sequence (as seen in the screenshot above, numbers 1-4):**

- (1) The user begins by visualizing the translation of the French sentence. Depending on their preferred learning style, they can either type the answer or practice it aloud
- (2) Next, the student hovers over the "Correct answer" field and clicks it to reveal the correct translation.
- (3) Subsequently, the user can either select '*Correctly answered*' or '*Incorrectly answered*' to track their score at the top of the screen.
- (4) finally the user presses *Next item* to do a next exercise.
- (5) At the top, you will find two icons: the first allows you to shuffle the items randomly, which is very useful to prevent memorizing items "in order." The second icon allows you to consult the Lexogoth notes associated with a sentence in a well-organized popup.

Note that this is not a test mode but simply a practice mode. The button *Restart exercise* resets all counters and brings the user back to the first item.

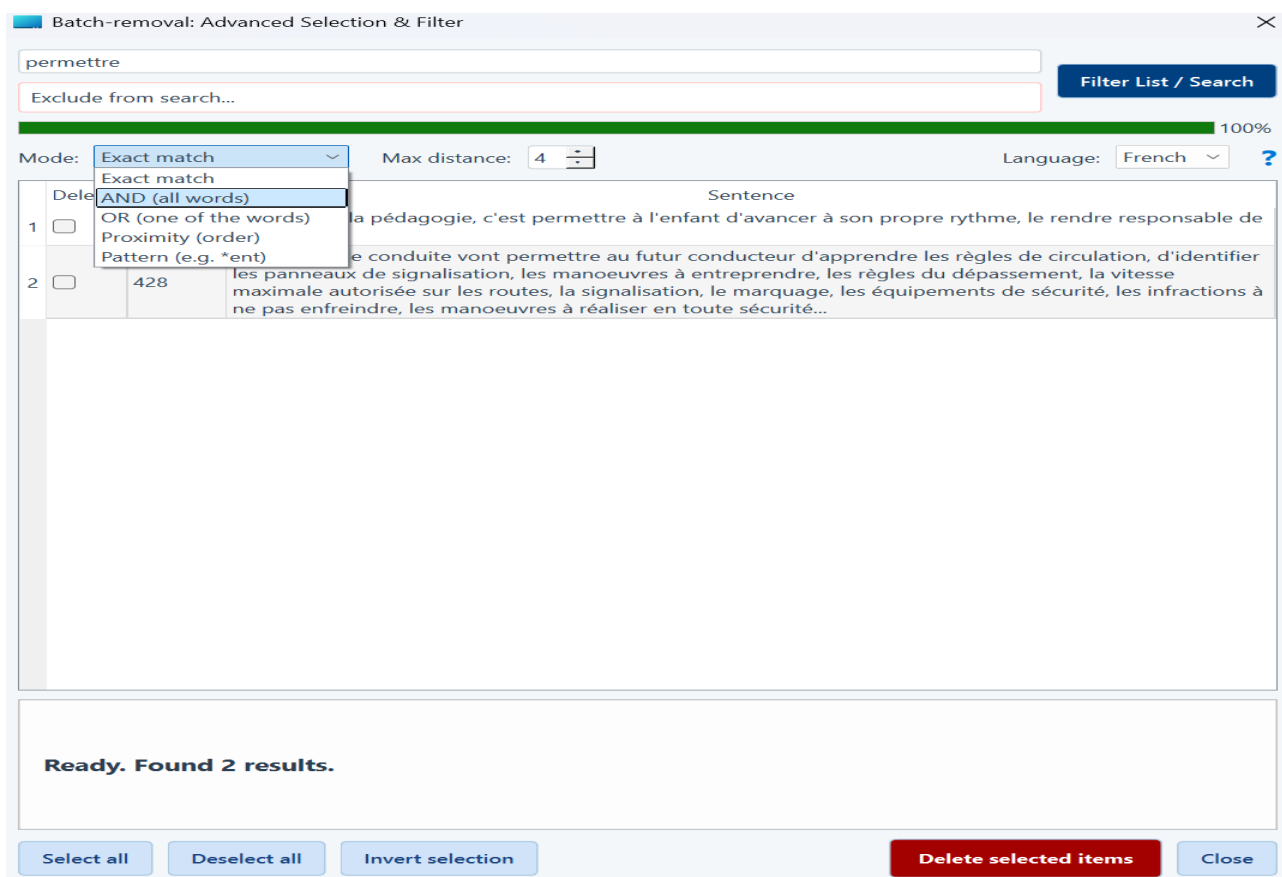
2.1.4 The nine buttons in the information field for the options *Edit database* and *Create new database*.

In the photo below you can see the buttons we have labeled A - I.



- Button A (**Import from Lexogoth**) is the most important: using this button, you can add entries directly from Lexogoth to your current database. We describe this feature in detail under *Importing / Exporting > Importing entries from Lexogoth* later in Chapter 2.2.

- Button B (**Batch deletion**): With this button, you can select large numbers of examples using advanced search functions and remove them all at once.



Tip: If you want to keep only very specific examples in a certain database, you can select them after the search query and then use the "Invert selection" button. This way, you can customize your database more quickly.

- **Button C** allows you to remove duplicates from the database. When merging databases regularly, it is likely that the same entry will be copied multiple times. This button allows you to remove them easily.

- **Button D** is used to sort all entries from smallest to largest, providing a very rudimentary ranking from easy to difficult entries. However, keep in mind that the length of a sentence or chunk does not always reflect its actual complexity.

- **Button E**: This button is for shuffling the database items randomly. This stimulates alertness and retention when students practice the database in exercise mode.

- **Button F**: When importing entries from the Lexogoth corpus, the notes from Lexogoth are copied in the background. When the user clicks this button while viewing an entry, a popup appears showing an overview of the block from which the sentence originated. When the user double-clicks one of the notes, it is copied directly into the entry's "Notes" field. This allows the user to easily and quickly add the appropriate usage notes.

Lexogoth - Lexical Toolbox

Read and edit database

Database info

Name database: Database 1

Author: MF

Subjects: Onderwerpen: public transportation

Current item: 8 of 493

Date creation: 23/05/2026

Entry

French (lemma, collocation, sentence): La non-utilisation du clignotant est le comportement le plus irritant en

English translation: Failure to use the turn signal is the most irritating behavior in urban areas: it is cited by

Lexogoth Intelligence: Corpus Insights

Corpus Analysis

Double-click a segment to add it to your notes.

1. The primary meanings of "carrefour" according to Robert: (a) [literal] intersection/crossroads, i.e., a place where several roads meet and cross each other; note that "carrefour" refers to the entire infrastructure of the intersection (sidewalk, signs, lights, ...). In everyday language, you can also use the following words as synonyms: "une intersection" or "un croisement". "Un croisement" (crossing, intersection) is a point where two or more roads cross each other or the act of crossing itself. "Une intersection" (intersection) is a much more general term and refers to a place where two or more lines intersect; in a traffic context, this is the area that roads with different driving directions use together when crossing. Strictly speaking, "une intersection" is therefore only one part of "un carrefour". However, in everyday language, the three words are used interchangeably, with "carrefour" being the most obvious choice. (b) [figurative] crossroads, i.e., a new situation (in someone's life, career, ...) in which they must choose from different paths; here you can also use "un tournant" (turning point). "Carrefour" is also used in the sense of a "place where a confrontation of ideas occurs or where different views, cultures, ... meet each other" (synonym: "une rencontre") and "a meeting, where a discussion takes place; a symposium".
2. "trouver chaussure à son pied" is an expression meaning "to find what one is looking for or needs".

Ready

- Button G (Advanced search): This allows you to search the entire database.

Lexogoth - Lexical Toolbox

Read and edit database

Database info

Name database: Database 1

Author: MF

Subjects: Onderwerpen: public transportation

Current item: 10 of 493

Date creation: 23/05/2026

Advanced search in Database

sans

Keywords...

Zoeken

Typ hier woorden die moeten uitgesloten worden van de zoekopdracht...

Mode: Exact match ? Maximum words between: 4 Language: French

15 resultaten gevonden (in 0.03 sec. uit 493 items)

10	16	32	50	115	186
233	393	398	402	410	413
424	438	442			

Selecteer item **Overview (0)** **Save in database** **Save to context block** **Close**

There are six different search options:

(1) **Exact match:** This searches for the literal search string. For example, “aller” will only yield “aller” as a result, not the derived forms like “vais, ira,...” If multiple terms are entered, it will search for that exact string, which is very useful for finding conjunctions and other immutable structures such as “pour que / au fur et à mesure que.” Note that if a compound term ends in “que, de, si,” it will also include “qu, d, s” respectively in its search query.

(2) **AND search:** This searches for all forms of a word. For example, “aller” is broken down into all known forms of “aller,” such as “vais, va, ira,...” and a search is then performed in the corpus using those forms. In the case of multiple search terms, it will search for all sentences in which both terms appear.

(3) **OR search:** For single terms, this behaves the same as the AND search (searching for all forms of a word). However, for multiple terms, it will search for all sentences in which at least one of the requested terms appears.

(4) **Proximity search:** This allows you to enter multiple words, and the pattern is then searched for in all sentences in the database. The box with a numerical value determines the maximum number of words allowed between the search terms.

If the maximum distance is set to 1, Lexogoth searches for all search terms next to each other, without any words in between. For instance, the search term “même si” with distance = 1 will yield all results for “même si,” such as:

- *Votre article m' irrite profondément même s'il s'y pointent quelques accents de vérité.*
- *Même si tu ne crois pas à l'existence d'aliens, il n'en est pas moins vrai que [...]*

When the distance is set to 2, Lexogoth searches for all combinations in the form of “si” [another word] “que.” Thus, the search query “si que” with distance = 2 (meaning 1 word in between) yields, among others, the following:

- *Cet énième report est la preuve que le dossier n'est pas si simple que ça.*

(5) **Pattern search:** This option allows you to search for patterns within words themselves and is primarily used for studying morphology. Searching is done using a wildcard, specifically the asterisk: *. You can place this wildcard in the following positions:

- **Before a string** (connected to it), such as “*amment”: this will search for all words ending in “amment.”
- **After a string** (connected to it), such as “ant*”: this will search for all words beginning with the specified prefix “ant.”
- **Inside a word**, such as “al*en”: this will search for words that begin with “al” and end with “en.”

It is important to note that the wildcard (*) must always be attached to or within a word, as this feature focuses on patterns **within** words, not syntactic patterns.

You can also combine this search option with a 'complete' lexical item. For example, the query “en *ant” will search for all *gérondifs* in the corpus.

Finally, there is a field below the search bar (the **NOT field**): here you can type words that should be excluded from your search query. This allows users to filter their search results even further.

(6) **Syntactic pattern**: This search option is the most complex but offers teachers and students the ability to look up complex patterns in the corpus based on sentence structure.

The user can perform a search here with or without the wildcard ‘*’, and it is also possible to combine a search query with fixed keywords. The fixed keywords are as follows:

Keyword	Betekenis
NOM	Zelfstandig naamwoord
PAPR	<i>Participe présent (onvoltooid deelwoord)</i>
PAPA	<i>Participe passé (voltooid deelwoord)</i>
VER	<i>Verbe (werkwoord)</i>
ADJ	<i>Adjectif (bijvoeglijk naamwoord)</i>
ADV	<i>Adverbe (bijwoord)</i>
PREP	<i>Préposition (voorzetsel)</i>
INF	<i>Infinitif (infinitief)</i>
IIMP	<i>Indicatif imparfait</i>
IFUS	<i>Indicatif futur simple</i>
CPRE	<i>Conditionnel présent</i>
SPRE	<i>Subjonctif présent</i>
IPRE	<i>Indicatif présent</i>

These fixed keywords are much more precise in de Tooblox-module.

→ search without wildcard

When using a keyword, you must always precede it with a "+" or simply use the "**Keywords...**" button on the right to select the appropriate label. Note that if you set the maximum distance to 1, Lexogoth will search for words that are exactly next to each other; distance 2 means searching for words with one or no words in between, and so on.

Below are some valid search queries:

Search query	Which pattern, construction...
<i>en +PAPR</i>	gérondif
<i>avoir être +PAPA</i>	Le passif
<i>venir de +INF</i>	Le Passé récent
<i>tellement +VER que</i>	<i>Tellement + verb + que</i>
<i>autrefois +IIMP</i>	Adverb <i>autrefois</i> used with <i>imparfait</i>

→ search with wildcard

You can also search using a wildcard (*) that represents exactly one word. You can select a distance for this search query as well, but you need to be careful here. The distance is not applied to the word immediately preceding the wildcard. In other words, in the search query “si * que” with distance = 3, Lexogoth will always look for “si” immediately followed by a random word, after which the “que” may be at a maximum distance of three from the wildcard. It is important to realize that in this search mode, you write the wildcard separately from the word, unlike the Pattern search where it is attached.

- Export results for a query:

Advanced search in Database

Keywords... Zoeken

Typ hier woorden die moeten uitgesloten worden van de zoekopdracht...

Mode: Exact match ? Maximum words between: 4 Language: French

15 resultaten gevonden (in 0.03 sec. uit 493 items)

10	16	32	50	115	186
233	393	398	402	410	413
424	438	442			

Geselecteerd Overview (3) Save in database Save to context block Close

When viewing the search results (photo above), it is possible to double-click on the results (the numbered buttons) and then export them to a database file or a context block file (examples grouped into a single block) by choosing the buttons **Save to context block** and **Save to database**. With the **Save to database** option, all associated comments are automatically copied along with the entries, while with the **Save to context block** option, the user is prompted via a popup to choose which comments to export and which to exclude.

Subsequently, the user is given the option to add the examples to a new database or context block file.

These export options are useful for bundling sentences that belong together as teaching material centered around specific constructions or themes.

You can also choose to add everything to an existing database file or context block file.

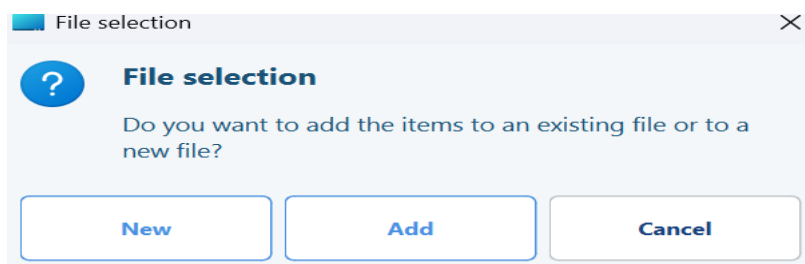
If it concerns a database file, the procedure is simple: the user just selects the `.lexdb` file in the popup, and the records are added to it.

If it concerns a context block file, things are slightly more complicated:

First, the user chooses from a popup which comments they wish to copy along.



Next, the user selects the "OK" option, after which they select a valid `.lexcb` file:



- Then, a popup opens with two columns. The first column contains the context blocks of the opened file, while the second column contains the items that need to be added to these blocks:

Lexogoth: Add sentences

Context blocks (Target)

Plus d'un Belge sur deux veut changer le système démocratique actuel. / Seule une minorité de Belges estiment que le régime démocratique doit être

(1) Batibouw s'ouvre aux nouvelles formes d'habitations et au tiny houses mais dans l'ensemble, on reste quand même dans des formes

le sacrifice des intérêts nationaux (SYN sacrifie: abandon)

L'heure est grave! A cause de l'absence prolongée de la bourgmestre, les partis de l'opposition craignent de voir un laisser-aller s'installer au sein

contrer, prévenir, éviter le décrochage scolaire / Le taux d'abandon scolaire précoce tourne autour de

+ New empty block

All in one new block

Sentences (Source)

Vidéo: Une conductrice de scooter traverse tranquillement un carrefour sans remarquer qu'un camion force droit sur elle.

Si vous arrivez sur un croisement difficile, sans panneau de signalisation, observez bien la circulation en sens inverse et commencez à ralentir pour céder le

Son refus serait en rapport avec son absence ? Non, cela est sans rapport. (SYN se rapporter à: avoir trait à, concerner)

Save

Cancel

- The user can now add the individual sentences from the right to the blocks on the left via drag-and-drop. As soon as a sentence has been successfully added, the block will turn light red to remind the user that the sentence has already been added. Nevertheless, a single sentence can be added to multiple blocks.

- Once finished, press **Save** to store all results. Alternatively, the user can choose to add all sentences from the right column in one go by selecting **All in one block**. The user can also create one or more new blocks manually and fill them with various items using the **+New empty block** option.

- Note that to drag, you simply need to hold down the mouse button on a sentence and drag it to the desired block.

- Button H (**N-gram analysis**): N-gram analysis examines which sequences of 2 to 5 consecutive words occur most frequently within the corpus. The benefit of this is that you can immediately map out patterns, set expressions, and commonly used sentence structures in the source language. This helps in creating representative exercises that align with the actual language use found in the database.

Database-analysis: Pattern & N-Gram Explorer

Pattern & N-Gram Explorer

Language for analysis: French

Pattern	Nr.	Sentence (with context)
> Patterns of 2 words		
> Patterns of 3 words		
sur l'autoroute	17	
de la route	11	
la bande de	10	
sur la voie	8	
sur la bande	8	
la voie de	8	
n'est pas	7	
de la circulation	6	
code de la	6	
kilomètres à	5	
un accident de	5	
se délester de	5	
ce n'est	5	
accident de la	5	
il y a	5	
la priorité à	5	
à l'heure	5	
n'empêche que	4	
à un carrefour	4	
la police a	4	
de l'autoroute	4	
il est interdit	4	
bandes de circulation	4	
est interdit de	4	
faire marche arrière	4	
et la circulation	4	
se dérouler dans	4	

Nr.	Sentence (with context)
1 25	la bande de circulation (SYN: la voie de circulation)
2 43	Dès que le conducteur est sur la voie de décélération, il doit ralentir pour adapter sa vitesse à la signalisation en place.
3 44	Si un véhicule circule sur la voie de droite au moment où vous vous insérez, et s'il n'a pas la possibilité de se décaler sur la voie centrale, il est d'usage d'attendre et
4 45	En voyant la voie de décélération au niveau de la sortie d'autoroute, je freine, j'essaie du moins. Mais la Ferrari ne répond pas.
5 57	Lorsqu'un usager circule sur l'autoroute, il lui est conseillé de tenir sa droite tout au long de son trajet. En revanche, en cas de ralentissements, voire
6 157	rouler dans la voie de droite quelle que soit l'allure du véhicule
7 398	Sur l'autoroute, il est interdit de circuler sans raison sur la voie du milieu ou sur la voie de gauche.
	Rouler sur la voie de gauche ou sur la voie centrale

Analysis completed.

Search selected item in database

Close

Using the button at the bottom labeled “Search selected item in the database,” you can look up the item currently selected in the database.

- Button I (**Word combinations**): This function analyzes the specific combinations in which a chosen word is used within the database. It provides the user with immediate insight into the contextual coherence of words, which is essential for understanding idioms and natural language usage.

French Concordance

Concordance Search

sans

before: 3

after: 4

Analyze Database

1.	tranquillement un carrefour	sans	remarquer qu'un camion
2.	croisement difficile,	sans	panneau de signalisation,
3.	, cela est	sans	rapport
4.	pouvoir s'intégrer	sans	trop de problèmes
5.	priorité de droite	sans	regarder à gauche
6.	acceptation intégrale et	sans	réserve des conditions d'
7.	le gouvernement va	sans	doute relever la taxe
8.	Mélangez	sans	cesse jusqu'à l'
9.	interdit de circuler	sans	raison sur la voie
10.	bande de gauche	sans	raison est considéré comme
11.	plein virage.	sans	aucune visibilité

Volledige context:

Pour compenser cette baisse des recettes, le gouvernement va sans doute relever la taxe de mise en circulation, augmenter l'imposition de l'électricité ou encore introduire une taxe kilométrique.

Export to RTF

Use in Search

Close

The search query for “sans” yields the results shown above. In this way, one can quickly obtain an overview of several common constructions for this verb.

For a detailed analysis, you can always click the button “Use in Search,” which will search for “sans” in the database and display all corresponding examples. Once you press that button, you arrive at this familiar field:

Advanced search in Database

sans

Keywords...

Zoeken

Typ hier woorden die moeten uitgesloten worden van de zoekopdracht...

Mode: AND (all words)

Maximum words between: 4

Language: French

15 resultaten gevonden (in 0.03 sec. uit 493 items)

10

16

32

50

115

186

233

393

398

402

410

413

424

438

442

2.1.5 Merging & exporting databases

With this button, the user can merge different databases (with the **.lexdb** extension). As soon as you press the button, a window opens in which you can select multiple files.

The screenshot shows a window titled "Lexogoth Database Merge Center" with a close button (X) in the top right corner. The window is divided into two main sections by a horizontal line.

Section 1: 1. Select source files (indicated by a red header bar)

- A large empty rectangular box labeled **A** in red text, intended for displaying selected files.
- Below the box, there are two blue buttons: **Add file(s)** and **Remove**.
- To the right of these buttons, there are two small blue buttons with upward and downward arrows, labeled **B** in red text.

Section 2: 2. New Database Metadata (indicated by a red header bar)

- Database name:** A text input field containing "E.g.: Merged_Data_2026".
- Author:** A text input field containing "Author name", labeled **C** in red text.
- Subjects:** A text input field containing "Subjects (comma-separated)".

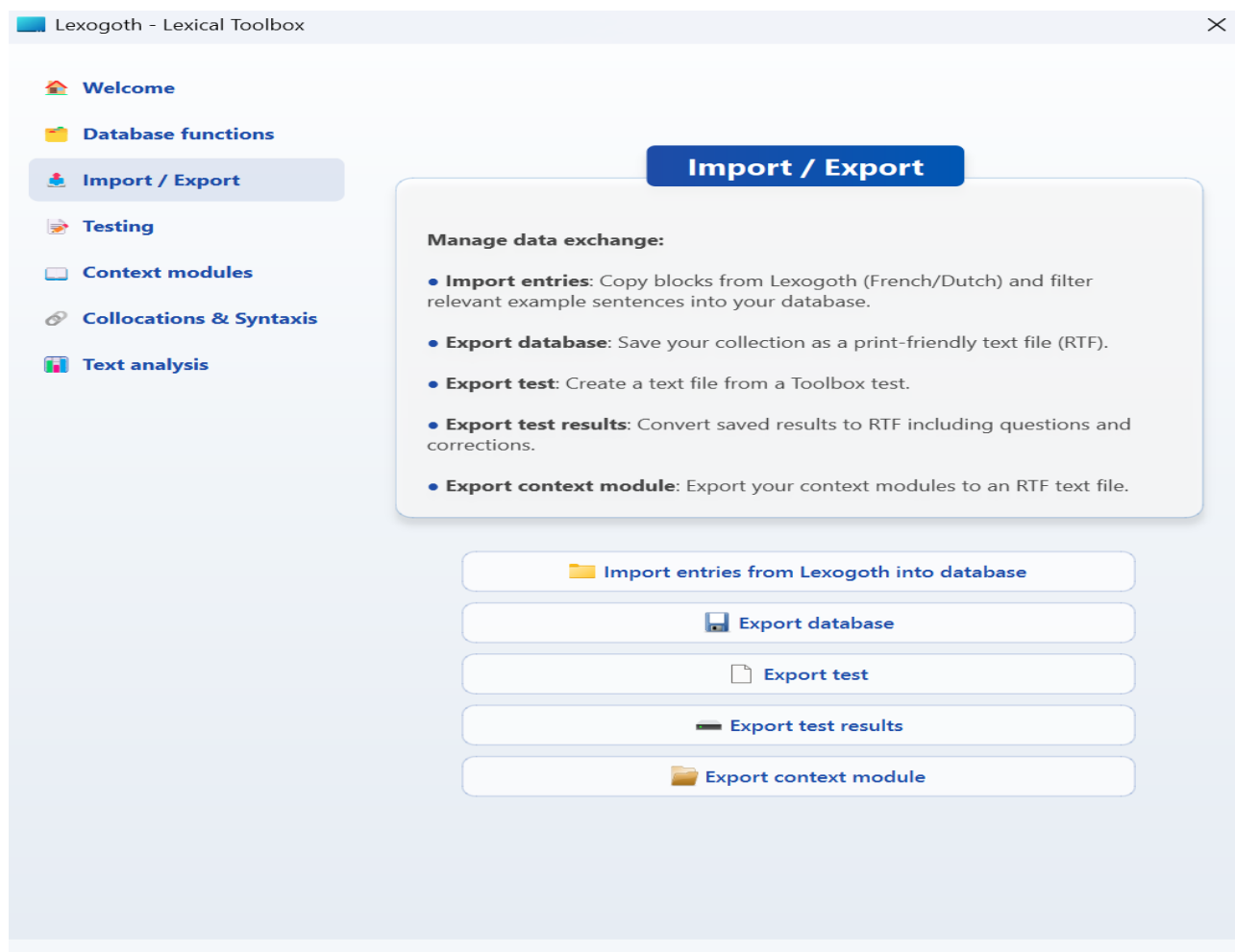
At the bottom right of the window, there are two buttons: **Start merge** (disabled, grey) and **Cancel** (blue). A red label **D** is positioned below the "Start merge" button.

In the window mentioned above, the user can add the databases to be merged via field B by using Add files; if a mistake is made, the user can remove the database in question. The arrows are used to determine the order of addition. In field C, the user can choose the name of the database, the author, and the subjects. Once all fields have been correctly filled in, the user can start the merging process by clicking *Start merge*.

In order to export a database to a Word-file, you choose *Export Database* from the main menu "Database Functions", select a file (.lexdb) to export, enter a valid filename and then click on *Export to RTF*.

2.2 Module 2: Import / export

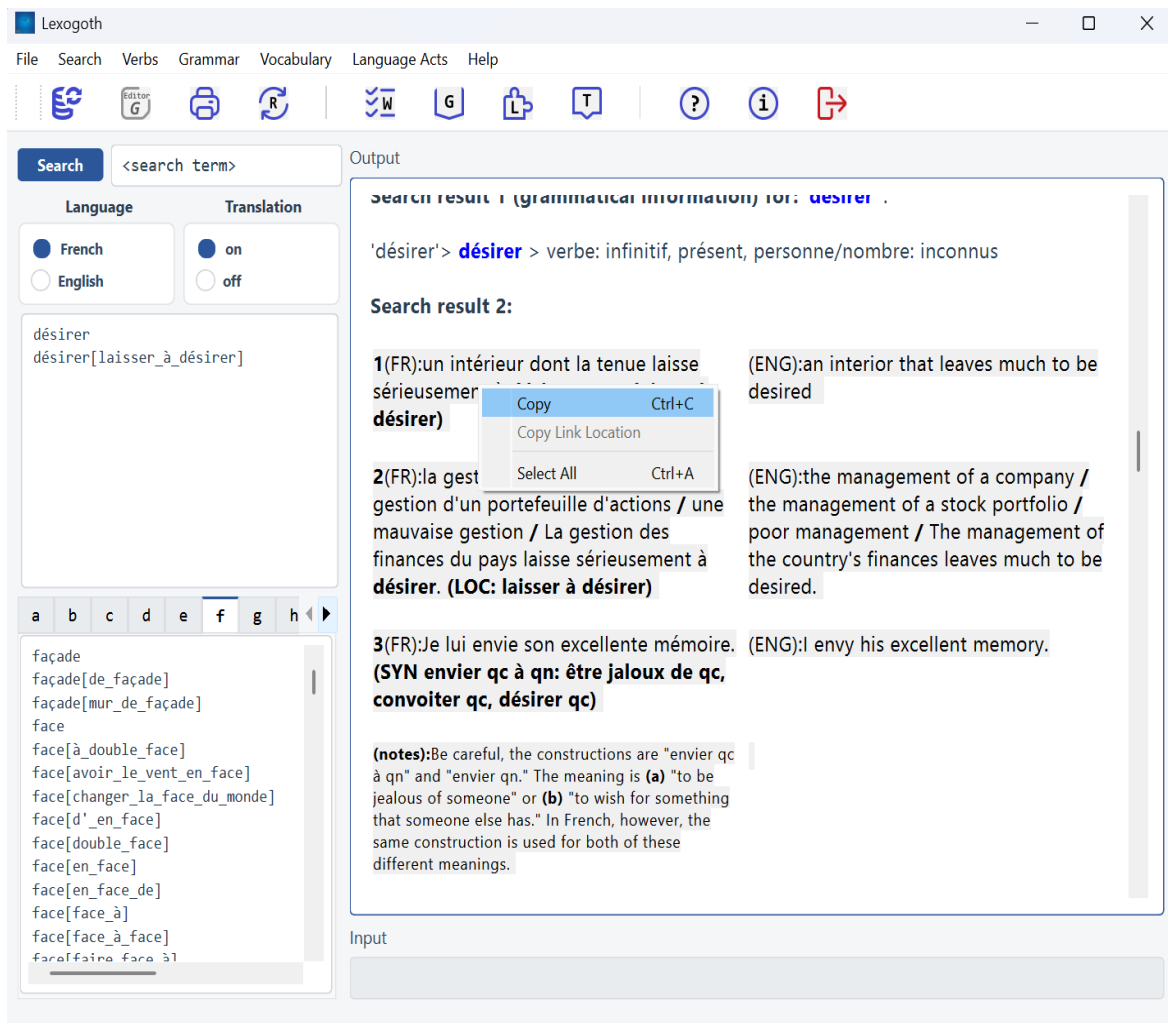
The action buttons in this field are *Import entries from Lexogoth*, *Export database*, *Export test*, *Export test results* and *Export Context Module*:



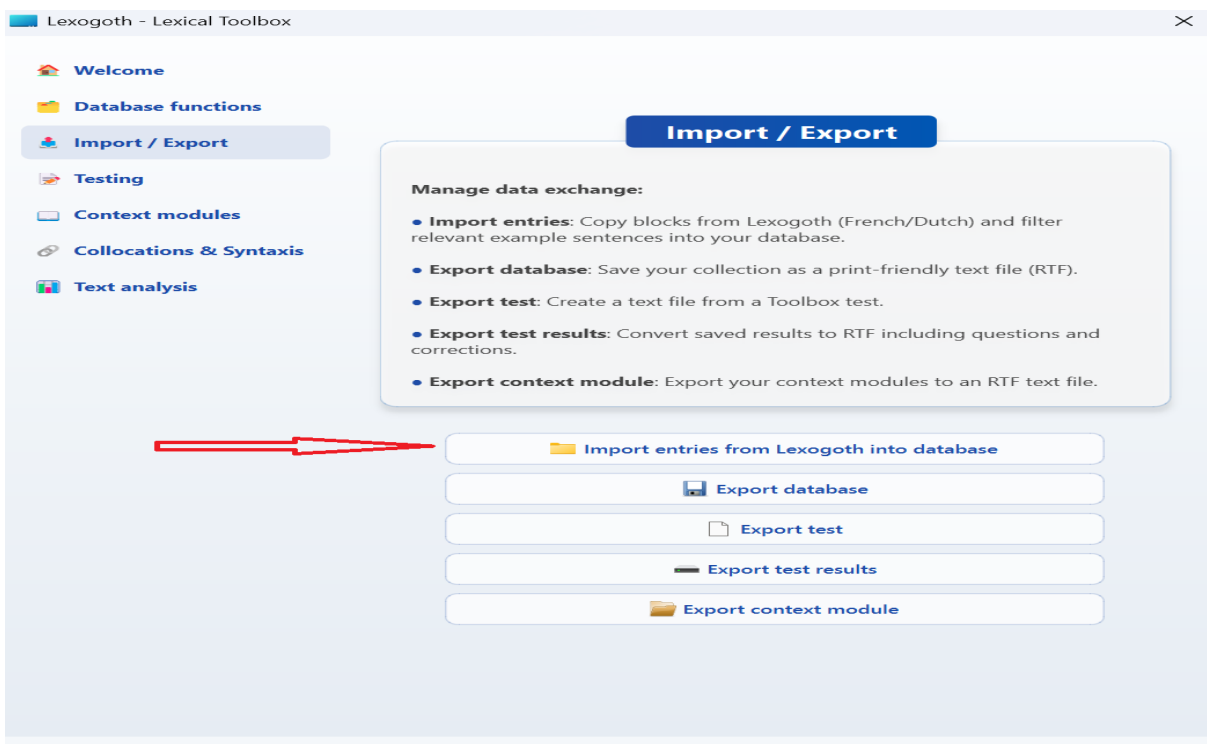
2.2.1 Import entries from Lexogoth

The button *Import entries from Lexogoth* is the most important functionality from the *Toolbox* and allows users to copy blocks of French text with their English translations (from Lexogoth) and place them into a database. Through this command, a field is created for each entry (sentence or chunk) and its corresponding English translation. Note that the comments from the Lexogoth entries are not copied along. We show in detail how this works:

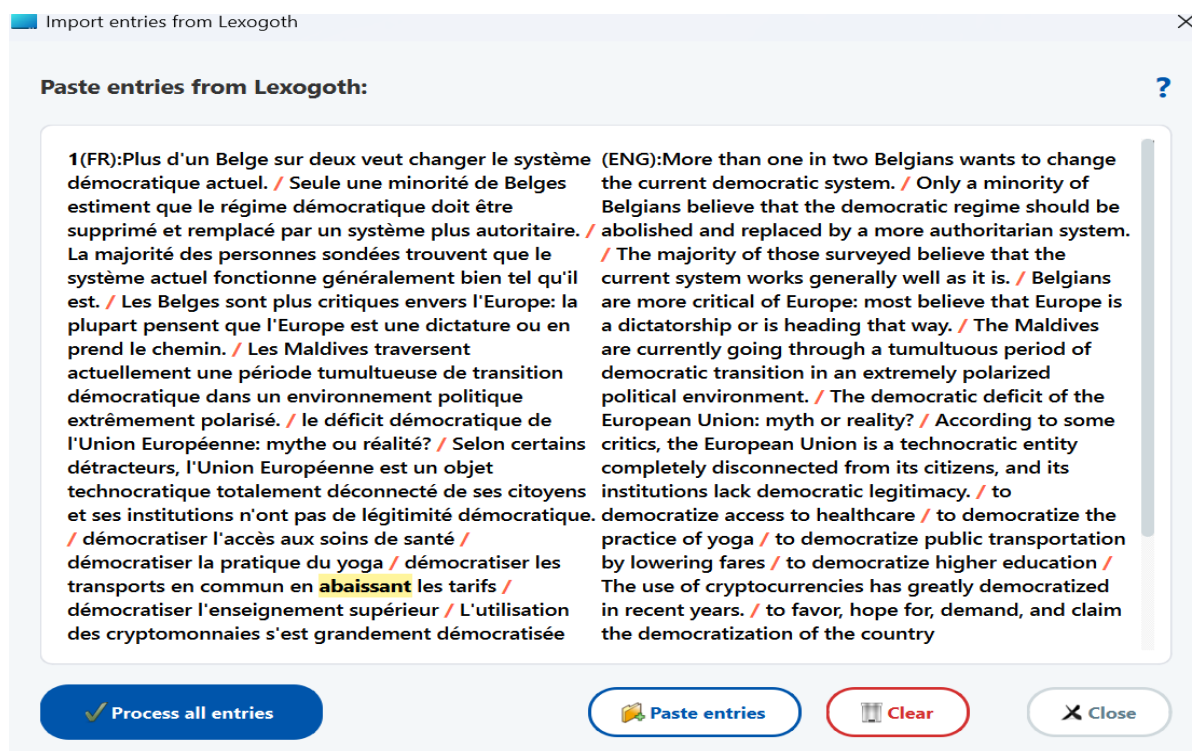
- **step 1:** select a block (or multiple blocks) of text from Lexogoth and copy (via mouse or CTRL-C); the search language must be French and not English. **Do not forget that you must copy both the French block and the corresponding translation:**



- **step 2:** Then in the *Toolbox* under *Import / export* you choose the button *Import entries from Lexogoth*:



- **step 3:** Then a popup screen appears where you have the options *Process all entries* and *Paste entries from Lexogoth*. You choose *Paste entries from Lexogoth* and then the text you selected in Lexogoth appears in the editor field:



- **Step 4:** Next, click on **Process all entries**, and a new popup divided into two screens will appear. At the top, you can see all sentences or chunks with their corresponding English translation (already selected). The bottom part of the popup is initially empty.

In the search field in the middle of the screen, you can enter a French word, after which the Toolbox will search through all the individual chunks and sentences that were just imported. This is very useful when copying large pieces of text from Lexogoth, as not all examples are necessarily relevant to the user. By typing a search term (French) here, all examples that match your query will appear in the bottom screen.

Next to the search field, you have an important option to refine the search query:

- **Search exact match**
- **Search all forms of all terms (AND)**
- **Search all forms of at least one search term (OR)**
- **Find terms in correct order (maximum distance of 3 words)**

This option allows the user to execute a search query within the imported sentences, enabling highly targeted searching:

Imported entries

	FRANS	NEDERLANDS
9	☒ démocratiser la pratique du yoga	to democratize the practice of yoga
10	☒ démocratiser les transports en commun en abaissant les tarifs	to democratize public transportation by lowering fares
11	☒ démocratiser l'enseignement supérieur	to democratize higher education
12	☒ L'utilisation des cryptomonnaies s'est grandement démocratisée ces dernières années.	The use of cryptocurrencies has greatly democratized in recent years.
13	☒ favoriser, espérer, réclamer, revendiquer la démocratisation du pays (EXP: prendre le chemin de, un déficit démocratique)(LOC: plus d'un, t...	to favor, hope for, demand, and claim the democratization of the country

[Alles selecteren](#) | [Alles uitvinken](#)

SEARCH IN CORPUS [?]

démocratiser

All forms (AND)

French

Search

GESELECTEERDE

démocratiser les transports en commun en abaissant les tarifs

démocratiser l'enseignement supérieur

L'utilisation des cryptomonnaies s'est grandement démocratisée ces dernières années.

favoriser, espérer, réclamer, revendiquer la démocratisation du pays (EXP: prendre le chemin de, un déficit démocratique)(LOC: plus d'un, tel que, se démocratiser, les transports en commun)

✓ Add selected examples to...

✗ Cancel

Number of results found: 6

What do the search options really mean ?

(a) **Search all forms of all terms (AND):** Here the user can type in one or more search terms and the toolbox will then search the sentences and only retain those sentences in which EVERY word of the search term appears in any possible form. Suppose you type in “*acquérir*”, then all sentences are retained in which a possible form of the verb (e.g. *acquiers*, *acquerrai*, *acquis*...) appears. For multiple search terms, the Toolbox does the same: searching all sentences and determining whether a form of EVERY search term appears in the sentence. This means that when searching for the multiple search term “*film cinéma*” with this option, the Toolbox will search for the sentences in which a form of *cinéma* and a form of *film* must appear. Because every word (or form thereof) must appear in the sentence, we called this option **AND**, a reference to the logical operator.

(b) **Search all forms of at least one search term (OR):** Here the user can also type in one or more search terms and the toolbox will then only retain the sentences in which AT LEAST one word (or form thereof) appears in the sentence. This is useful if you have used multiple search queries from Lexogoth in one processing task. Suppose you type in the search term *pollution voiture réchauffement* here, then the Toolbox will return every sentence in which at least one of the search terms (and their derivative forms) appears. This search query is ideal for compiling thematic databases. For example, to study certain connectors, you can request all examples in which *cependant par contre* or *au contraire* appear.

(c) **Search exact match:** This search option will consider the given input in the search field as one string. In other words, if you type in *pour que*, it will search for the sentences in which this exact sequence (the conjunction) occurs. Note that compounds ending with *que*, *si*, *de* are also combined with *qu*, *s*, *d*

respectively, which means that the *Toolbox* will not only search for *pour que* but also for *pour qu*.

(d) The option **Order (max. 5 distance)** will retrieve all sentences where the terms appear in the correct order, in all forms, and within a maximum distance of 5 words.

Some examples:

- Suppose that you search for “bien que” (conjunction), then you type in the search term “bien que” and you choose -because this word combination has no variation- the option “Search exact match” and then the *Toolbox* will only retain the sentences in which “bien que” or “bien qu” occur as a whole.

- If, however, you search for the combination “faire face à”, you will find results more difficultly if you treat this as a single search term using the option “Search exact string match”. After all, this expression can appear in many forms such as “fait face aux / ferait face à / ne fait pas face à ...”. And to overcome this problem, you can type “faire face” in the search field and then select the option “Search all forms of the search terms (AND)”. Then the *Toolbox* searches for all sentences in which BOTH all forms of *faire* AND all forms of *face* appear and displays these in blue so that the user can certainly find all examples of the expression.

- One last example: suppose that you have searched in Lexogoth for all blocks in which “apprendre” occurs and that you have copied these to the *Toolbox*, then it is important that you search the sentences found with the option “Search for all forms of all search terms (AND)” or “Search for all forms of at least one search term (OR)” so that the sentences with the forms “appris, apprendras,...” also appear in the search results. If you use the option “Search exact string match”, then it will only search for the literal string “apprendre” and not for the derivative forms of “apprendre”.

After your search, clicking **Add selected examples to...** will trigger a popup asking you to add the examples to an **Existing database** or a **New database**. Once you have made your choice, the sentences or chunks are copied directly to the destination, where you can begin editing or adding annotations.

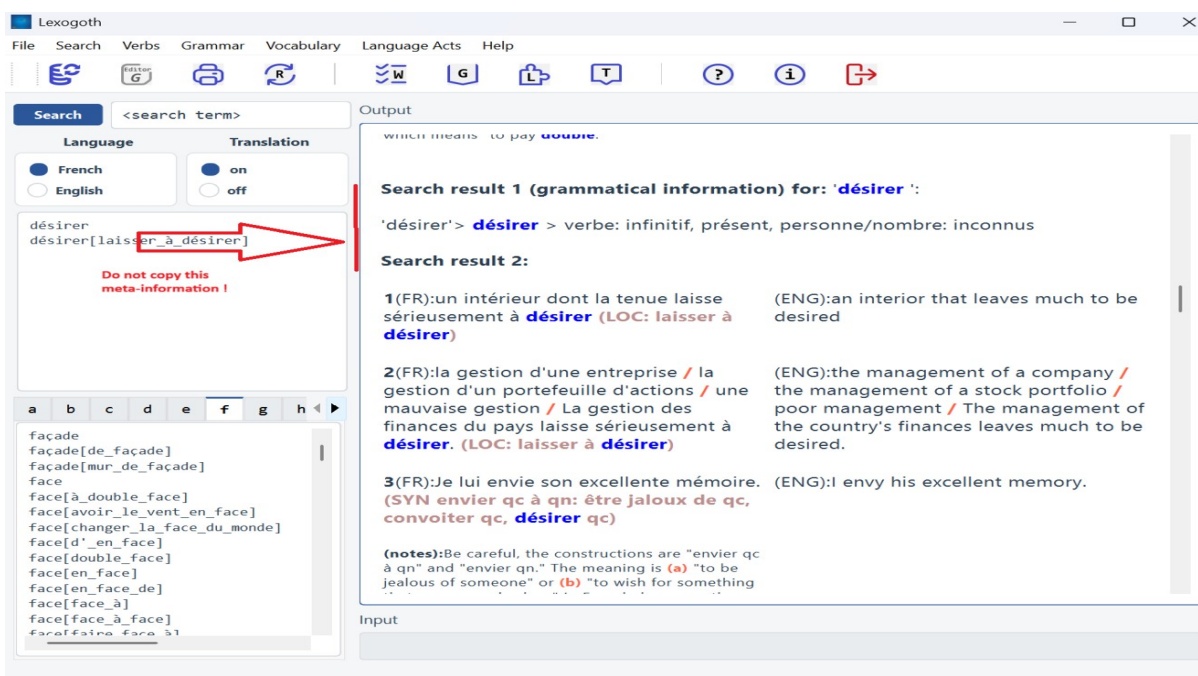
Important: If a search query using **AND** or **OR** yields no results, be sure to repeat your search one last time using **Exact match**.

Note that you can also import directly while you are already working within a database; simply press the **Batch Import** button (icon) to add items to your current database.

Some important tips for importing:

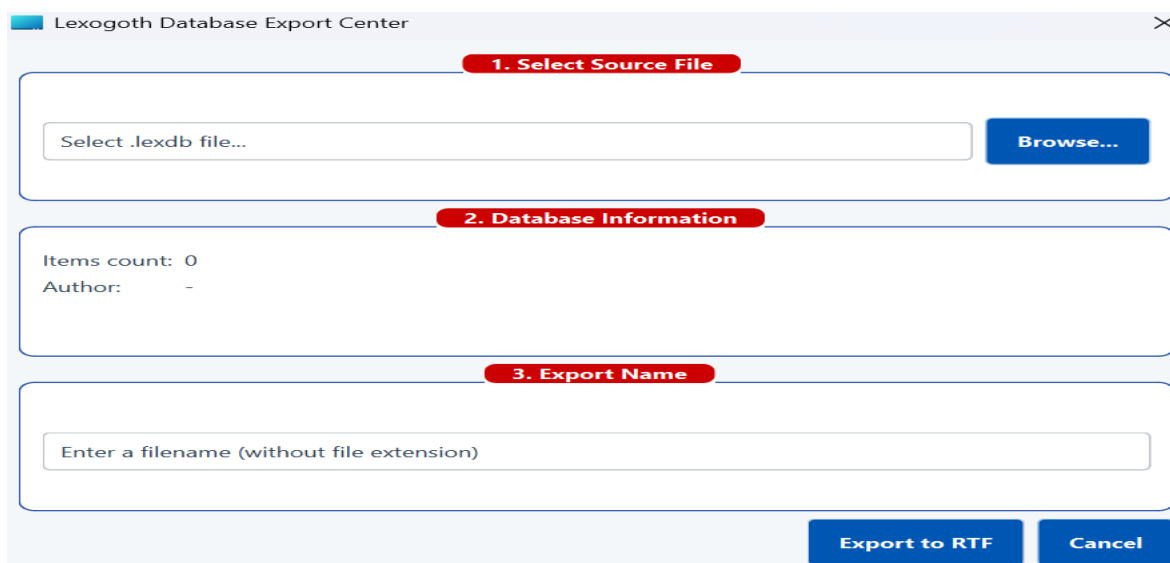
- You must always copy at least one complete block, which means a set of French sentences with the corresponding English translation. You may, but do not have to, copy the comments, as they will be ignored anyway.

- Do not perform copy operations when the search language in Lexogoth is English. If you do, the editor will consider the entries invalid and will not paste them. The direction is always French → English.
- When you copy, do not copy the meta-information (red arrow in photo below) that precedes each search query in Lexogoth; if you do, the Toolbox will reject the input as invalid:



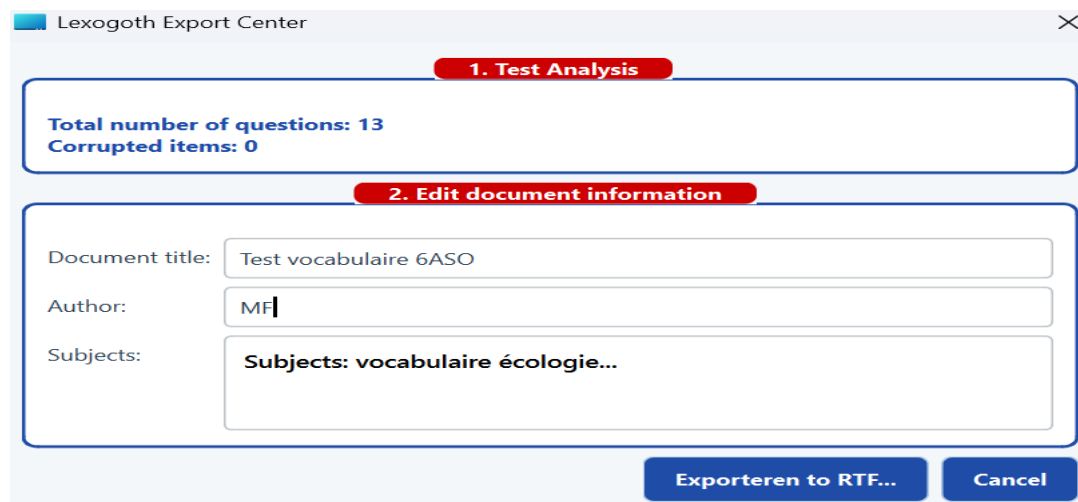
2.2.2 Export database

With the Export database button, the user can convert an entire database into a reader-friendly text file in *RTF* (rich text format) which he can freely edit with his favorite word processing program.



2.2.3 Export test

It is also possible to convert tests created with the Toolbox into an editable text file in RTF format. Note that the test files that have been created with the Toolbox, have the **file extension “.lextest”**.



The screenshot shows the 'Lexogoth Export Center' window. It has a title bar with a close button. The window is divided into two main sections. The first section, '1. Test Analysis', shows 'Total number of questions: 13' and 'Corrupted items: 0'. The second section, '2. Edit document information', contains three text input fields: 'Document title:' with the value 'Test vocabulaire 6ASO', 'Author:' with the value 'MF', and 'Subjects:' with the value 'Subjects: vocabulaire écologie...'. At the bottom right, there are two buttons: 'Exporter en RTF...' and 'Cancel'.

2.2.4 Merge databases

With this button the user can merge several databases (**file extension: .lexdb**). As soon as you press the button, a Windows window opens in which you can select multiple files. **Note, the selection of multiple files is done by holding down the CTRL key while selecting the file.** After the selection, the user is asked to provide the details for the new database.

2.2.5 Export test results

Here the user can reopen the saved result of a created test and export it to a text file. As soon as you press the button, you are asked to open a file with results (**file extension: *.lexresult**). Once selected, the file is opened with a table overview. At the bottom, you have the option to view the test in text view via the Show text view button. The latter presents the questions asked, points earned, and the given and correct answers in a clear manner.

By pressing the *Export RTF* button at the bottom right of the screen, you can also convert this reader-friendly file into an editable RTF file.

2.3 Module 3: Testing

2.3.1 Various types of questions

You can also set up tests with the *Toolbox*. The principle remains the same as for the databases: you click on the *Create Test* button and you are guided to a screen where you must fill in the necessary details (name...) and then you are taken to an empty question screen (see illustration below). On this page the user can add his questions. There is a choice of the following question types which are selected via **field A** (illustration next page):

Lexogoth - Lexical Toolbox

Create questions for test

Questionnaire details

Name: La condition	Author: MF	Date: 23/05/2026	Subjects: la condition: expressions cour	Number of items: 0
------------------------------	----------------------	----------------------------	--	------------------------------

Question data: question and answers

Question type: Multiple choice **A**

Points: 1 **B**

Text: Type your question **C**

Option A:

Option B:

Option C:

Option D: **D**

Correct option: enter the correct option here (A, B, C, D)

[Previous question](#) [Next question](#) [Delete question](#) [Save test](#) [Close](#)

(a) multiple choice: The user enters the question in question text field (field C) and then provides a minimum of 2 and a maximum of 4 answer options (field D). Next, he chooses the number of points to be awarded to the question (for a correct answer), that is field B. In field E, the user enters the correct option, i.e. the letter A, B, C or D. Note, for a question to be accepted as valid, the fields Question text, Option A, Option B, Correct option must be completed; the fields Option C and Option D are not mandatory. An example:

Question data: question and solutions

Question type:

Number of points:

Question text:

Complete the sentence: La banque peut changer l'intérêt de base à tout moment
_____ vous en informer

Option A:

Option B:

Option C:

Option D:

Correct option:

(b) The question type Exact completion is self-explanatory: in the question text the user types a question for which an exact answer must be formulated. With this type of question, the user must pay close attention: when someone takes the test, the given answer is compared with the correct answer and both **must match exactly**, otherwise the answer is marked as incorrect. For example, the following pairs are not exact matches: “Albert”-”albert” (difference: capital letter and lowercase letter), “albert “-“albert” (difference: name is followed by a space in the first word and not in the second).

Question data: question and solutions

Question type:

Number of points:

Question text:

Give the correct form of the verb in: L'Iran expulsera les inspecteurs des Nations Unies à moins que les Etats-Unis ne (lever) _____ leurs sanctions avant le 21 février, date butoir fixée par le parlement iranien.

Correct answer:

(c) Open question: here the examiner can ask an open-ended question and type the answer under solution. This question type is impossible to correct reliably without human intervention and **every answer is automatically marked as correct**.

Question data: question and solutions

Question type:

Number of points:

Question text:

Dans quelles situations, selon toi, certains jeunes peuvent-ils être tentés de "se biturer" ? Quelles pourraient être les conséquences de ce comportement ?

Solution:

- principales raisons: volonté d'appartenir au groupe, oublier le stress, problèmes personnels
- conséquences parfois graves-> comportements à risque (accidents de la circulation, coma, grossesse,...), addiction, gueule de bois,...

(d) **Matching:**

Question data: question and solutions

Question type: Matching

Number of points: 1

Question text: What are the most appropriate combinations ?

Column 1: Left column

Column 2: Right column

Row +

Row -

	Left column	Right column
1	pourvu que	si
2	pour que	afin que
3	avant que	dès que

In this type of question, the user is expected to match items from two columns; useful, for example, for testing knowledge of synonymy or linking related terms. The column names are editable and you can add up to a maximum of five rows. Note that when entering (setting up) the question, you should place the correct answers next to each other. In test mode, they will then be scrambled:

Question

Match what belongs together:

Score: 5

French expressions	English translations
des gaz à effet de serre	an energy company
une société énergétique	a climate footprint
une empreinte climatique	to run on petrol
l'énergie renouvelable	renewable energy
rouler à l'essence	greenhouse gases

(e) The final question type is *DropDown*: here, the examiner can have a user choose a correct answer via a dropdown menu, as shown in the illustration below:

Question

Choose the correct answer:

Score: 1

charger
charger
recharger
décharger

sa voiture électrique à domicile

Setting up such a question is done via the following screen:

Question data: question and solutions

Question type:

Number of points:

Question text:

Find the right answer:

Sentence:

gens aigris parce qu'ils sont condamnés à la misère, {} ouais, être dans un environnement pareil, ça doit être l'enfer... {}, si t'as fait le minimum pour prétendre à un boulot plus correct niveau éducation et salaire, l'ambiance est beaucoup plus bon enfant...

	Placeholder #	Options (comma-separated)	Correct option
1	0	à condition que, si, pourvu que	si
2	1	alors, comme, quand	alors
3	2	par contre, sans que, surtout, parce que	par contre

In the *Question field*, the examiner fills in a question of the type *Choose the correct answer / what fits best in the sentence / ...* Next, in the *Sentence field*, he types the sentence in which one (or more) dropdown menu should appear and at the spot where the user must choose a word, the examiner types the following: {}. The Toolbox will then automatically update at the bottom of the screen and generate a row. In the table you see three fields: the first field Placeholder # keeps track of the position of the various {} in the sentence, in the *Options* field the examiner types the various answer options for the dropdown menu and in the third field *Correct option* the examiner selects the correct answer.

Note that regarding the various {} (placeholders), we start counting from zero. That is to say that in the following sentence:

- Je ne {} pas que tu {} faire ça.

the first brace is assigned index number 0, the second is assigned index 1, etc...

2.3.2 Editing Tests and Importing Data from a Database or the Lexogoth Corpus

This allows you to open and edit an existing database or add questions by extracting examples from a self-created database or the Lexogoth corpus.

To the left of the question text field, the user will also find two buttons that allow them to copy entries from a database or from the Lexogoth corpus to the **basket** next to **Question text**:

Question type:

Points:

Text:

Type your question

Import from specific Database (.lexdb)

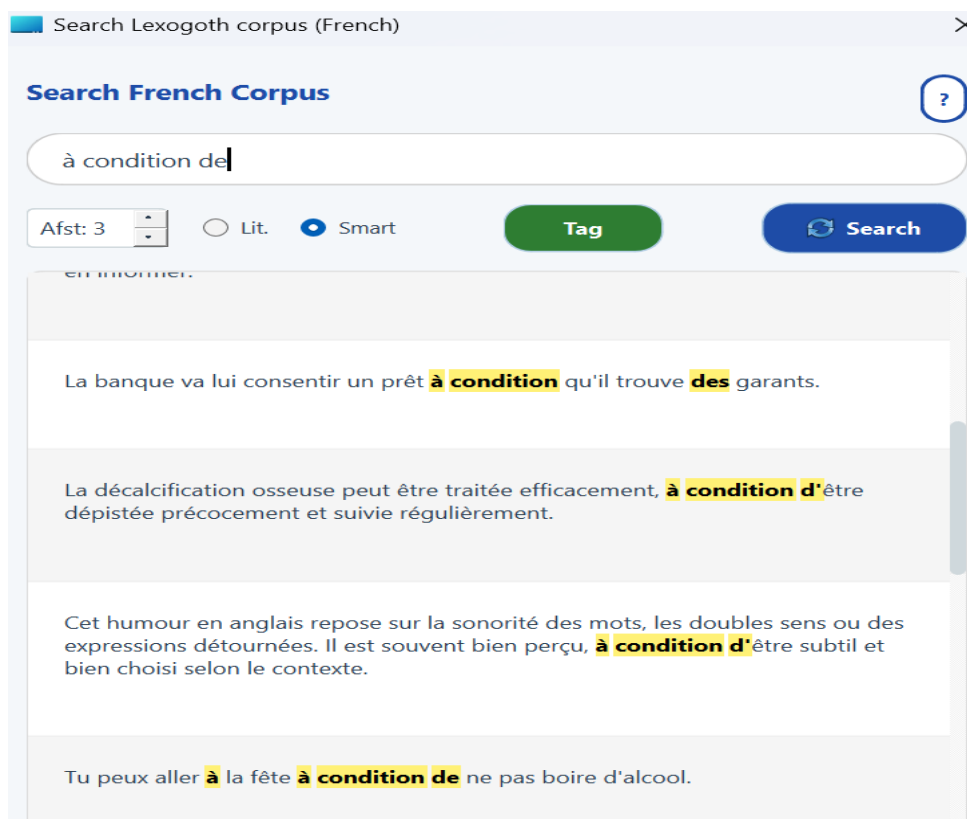
Search the entire Lexogoth corpus

Option A:

Option B:

This direct link between databases, the corpus, and test design is intended to streamline the user workflow.

When creating a test, the user can select any number of (French) sentences from their own custom database or directly from the Lexogoth corpus and copy them to the basket. Selection is performed by double-clicking on a sentence:



The option to search the Lexogoth corpus is quite advanced. You can search by proximity, use wildcards (where one * represents exactly one word), search by grammatical tags such as verb, past participle, conditional, etc. (the green Label button), search for combinations of words, and more. This is extremely useful and time-saving for preparing tests centered around grammatical topics or (lexical) themes.

Through the *Import* from a specific database, the user can integrate their own examples directly into a test. Note that the search options here are *AND*, *OR*, and *EXACT*.

You can then click on the sentences to place them in the basket, after which the user can click a sentence from this basket, which is then copied directly to the Question text field:

My collected sentences

Cet humour en anglais repose sur la sonorité des mots, les doubles sens ou des expressions détournées. Il est souvent bien perçu, à condition d'être subtil et bien choisi selon le contexte.

Tu peux aller à la fête à condition de ne pas boire d'alcool.

[Copy selection](#) [Delete](#)

[Empty basket](#) [Close](#)

Photo above: basket with collected examples

Photo below: example copied directly from basket

Question data: question and answers

Question type:

Points:

Text:

 Cet humour en anglais repose sur la sonorité des mots, les doubles sens ou des expressions détournées. Il est souvent bien perçu, à condition d'être subtil et bien choisi selon le contexte.



Option A:

Option B:

2.3.3 Taking a test and score calculation

With this function you can take a test. When opening a test, the name and ID (class, group,...) of the student are requested, after which the test can begin. The student can then solve the questions and go back and forth between the questions. When you press the *Close test* button, you arrive at the score sheet. There you get an overview of the questions, the given answers and the correct answers and the score per question and total score. To view your results (correct and provided answers) clearly, simply click on a question; a clear pop-up will display the question asked, the answer given, and the score:

Lexogoth - Lexical Toolbox

Take test

Test data

Name: MF **ID:** 6AM **Date:** 23/05/2026 **Start:** 20:09 **Items:** Questions 1 of 1

Question

Choisissez la réponse correcte: Cet humour en anglais repose sur la sonorité des mots, les doubles sens ou des expressions détournées. Il est souvent bien perçu, _____être subtil et bien choisi selon le contexte.

Score: 1

☐ au cas où

☐ en cas d'

☐ à condition d'

☐ à moins d'

[Previous](#) [Next](#) [Close test](#)

Photo above: taking a test...

Photo below: viewing the results

Results

OVERVIEW RESULTS

Score: 1 / 1

QUESTION (CONTENT)	YOUR ANSWER	CORRECT	SCORE
Choisissez la réponse correcte: Cet humour en anglais repose sur la sonorité des m...	à condition d'	à condition d'	1.0

Details result

INHOUD / CONTEXT

Choisissez la réponse correcte: Cet humour en anglais repose sur la sonorité des mots, les doubles sens ou des expressions détournées. Il est souvent bien perçu, _____être subtil et bien choisi selon le contexte.

UW ANTWOORD

à condition d'

CORRECTE OPLOSSING

à condition d'

Behaalde score: **1.0**

[OK](#)

[Save results](#) [Close](#)

You can save your results in a separate file (extension: *.lexresult) so that you can view them again later or export them to a text file for study. Note that the score is calculated per question and is divided over each part of the question. For example, if there is a question worth 2 points consisting of 3 parts, then each part will be worth 2/3 [0.67 points].

2.3.4 Loading test results

Here you can load your past (but first save them after having taken a test) results, display them in a popup, and at the bottom of the screen you have the button Show text view where you have a reader-friendly version of your lesson assignment. You can also export this version to a reader-friendly text file (RTF format) to edit or reread.

Lexogoth Results Browser

La condition

Auteur: MF | Datum: 23/05/2026 | Onderwerpen: Subjects: Subject: Subject: la condition: expressions courantes

QUESTION	GIVEN	CORRECT	SCORE
Choisissez la réponse correcte: Cet humour en anglais repose sur la sonorité des mots, les doubles sens ou...	à conditio...	à conditio...	1.00

TOTAL SCORE: 1.0 / 1

Toggle View

Export RTF

Close

2.3.5 Merging Tests

The user has the ability to combine several tests into a single, consolidated test by using the **Merge tests** button. This feature saves the user time by eliminating the need to transfer questions manually.

Through a popup (shown above), the user can select multiple test files and subsequently enter a new title, author, and subjects in the metadata field:

1. Test Index (Number of questions per file)

La condition.lextest (1 questions)
Test vocabulaire 6ASO .lextest (13 questions)

Total number of questions for export: 14

+ Add test

- Remove selection

2. Metadata for the new combined test

New Title: La condition (Merged)

Author: MF

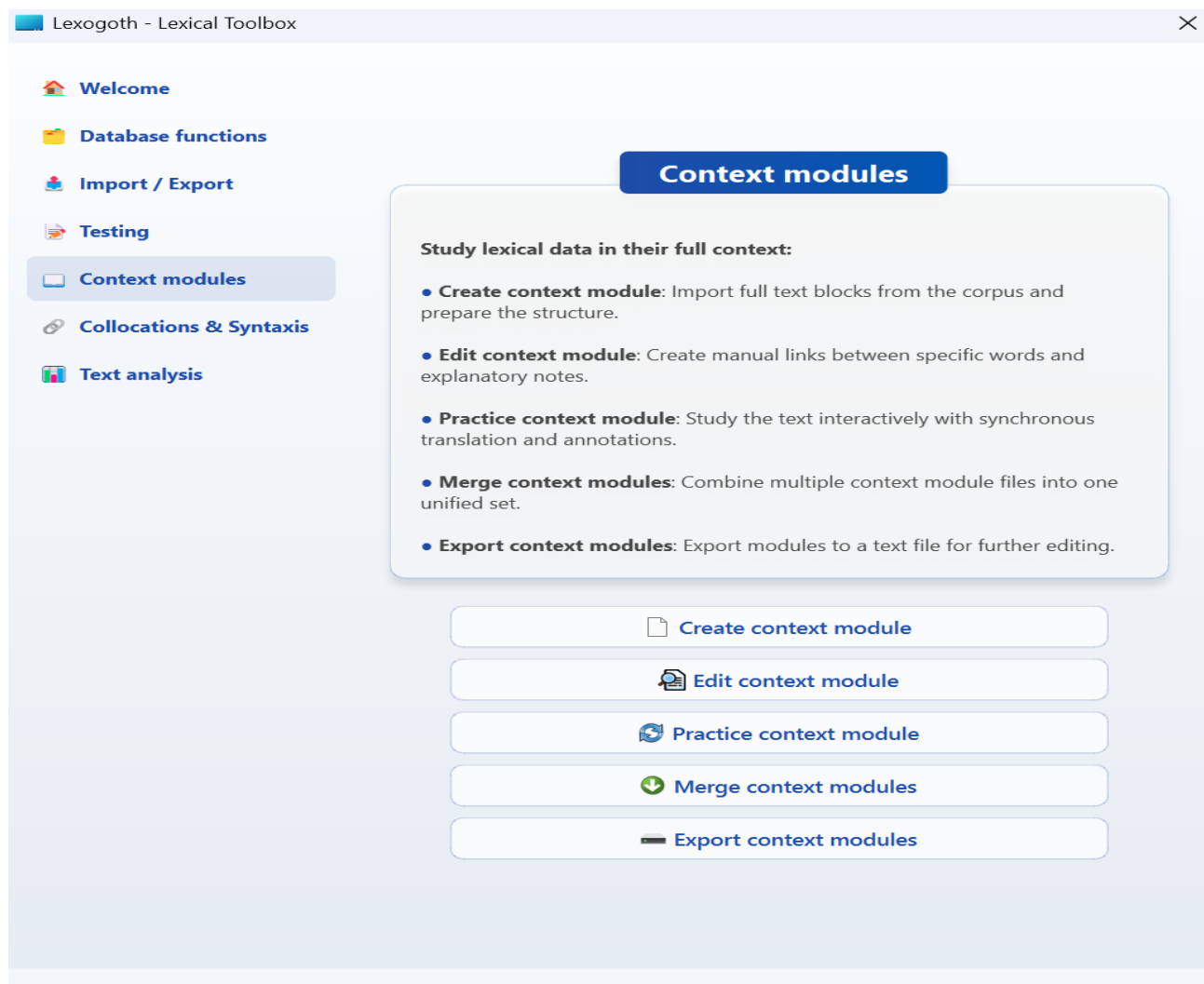
Subjects: **Subject: la condition: expressions courantes**

Merge

Cancel

2.4 Module 4: Context modules

The Lexogoth corpus is constructed from text blocks, each composed of various chunks or sentences. Within these blocks, the chunks or sentences often share a cohesive semantic, thematic, grammatical, or pragmatic relationship. Importing these into the database editor risks losing that cohesion. To ensure that students and instructors can display, study, and practice these coherent blocks, we have provided the **Context Modules** component.



2.4.1 Creating and Editing Context Modules

When you click the *Create context modules* button, you are presented with a screen where Lexogoth asks you to provide the title, subjects, and author for the file that will contain the context modules.

Once this is completed, you are taken to the page where you can edit or add blocks. If the database is currently empty, the first entry will be blank.

Lexogoth - Lexical Toolbox

×

Create and edit context modules

?

Module: Context M1 | Author: MF | Subjects: la condition

Blok: 1 / 1

FRENCH:

ENGLISH:

SEARCH WORD

NEW BLOCK

EDIT BLOCK

EDIT NOTES

REMOVE DUPLICATES

Show comments

Block 1 loaded.

Previous block

Next block

Import block

Delete block

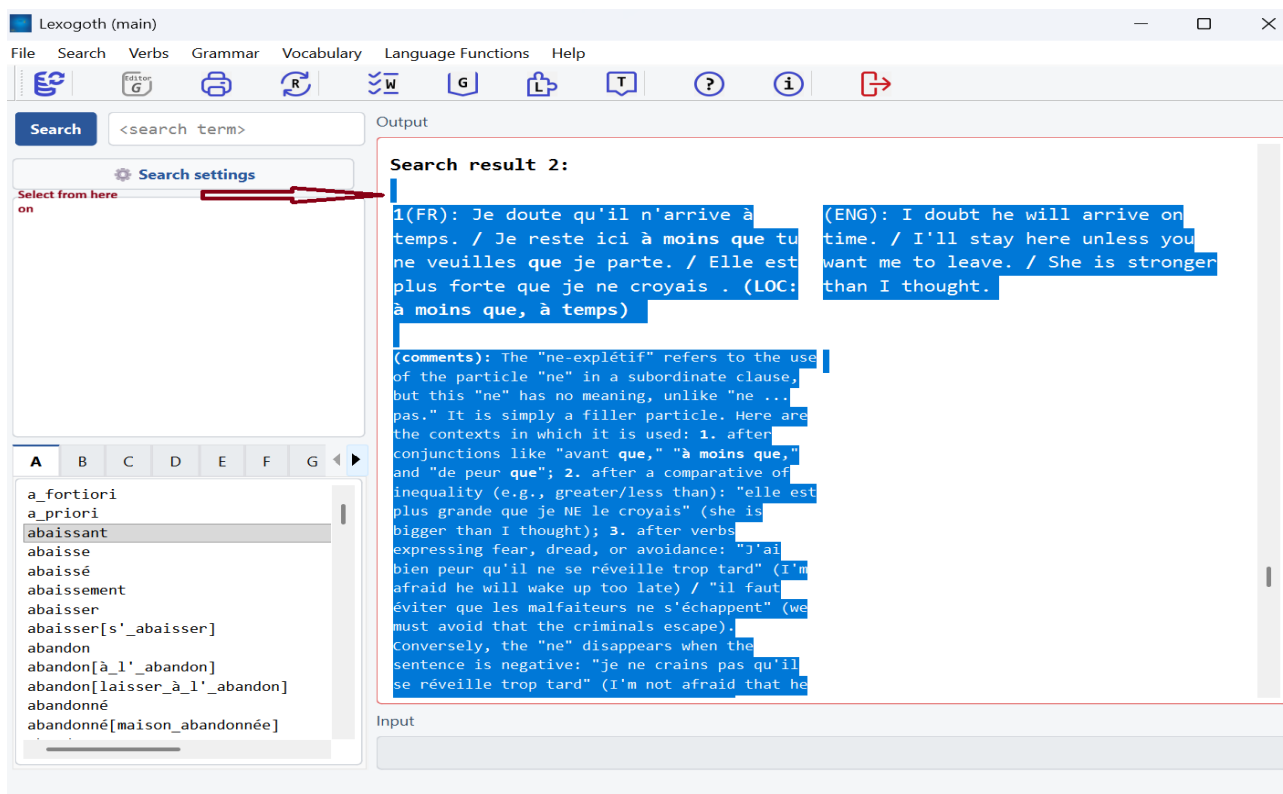
Save all

Close

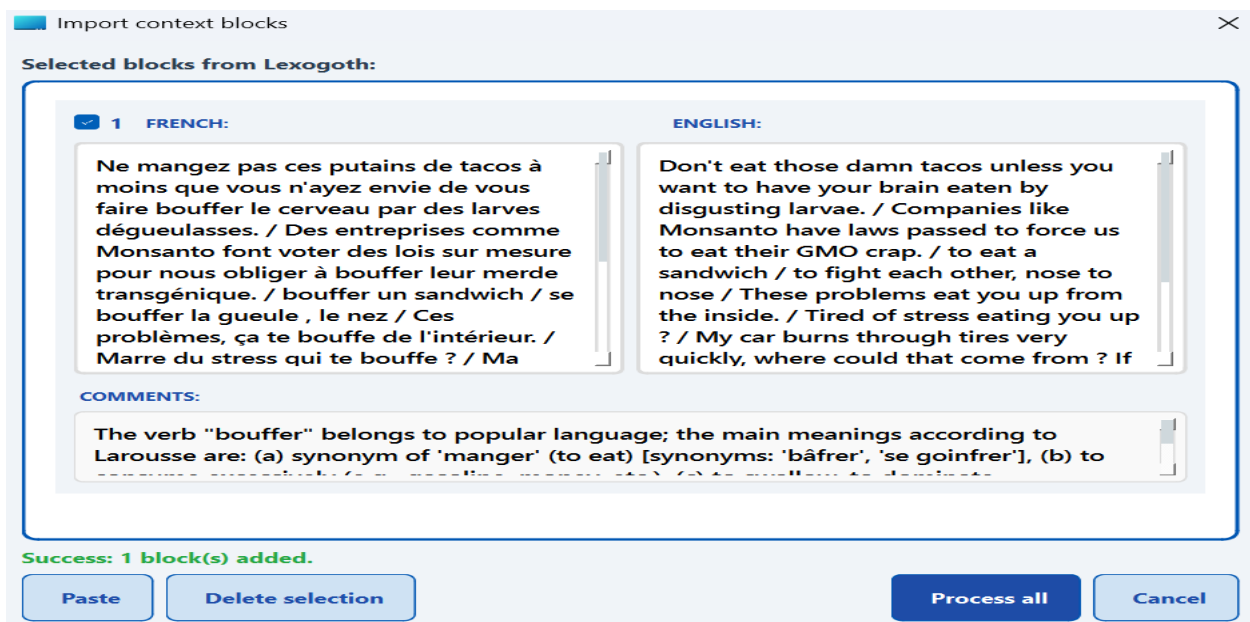
The **French** and **English** fields will display the respective French and English text blocks. Since the example above is still empty, the database needs to be populated. To avoid processing errors, we have provided an import routine that allows users to easily add text blocks from the Lexogoth corpus to a context module. The procedure is similar to that of the database editor:

(a) **Make a selection in the Lexogoth corpus:** The minimum permitted selection consists of all French sentences, followed by their English translation, optionally accompanied by a field for comments.

The order here is important, and the user must be careful not to copy extraneous metadata. Selection is performed just below “Search result 2” in the Lexogoth output screen:



(b) Once everything is selected, go back to the Toolbox where you have chosen the option **Context Modules** (sidebar) > **Edit/Create Context Modules** (buttons), and then press the **Import Block** button at the bottom, which will immediately display a popup screen. There, press the **Paste** button:



In this popup, you can still remove items (blocks) if you wish. Otherwise, click **Process all**, and all blocks will be placed into a database.

(c) Once in the database, the user can browse through the blocks. If they click on a sentence with the mouse, the entire sentence (or chunk) will be

highlighted in light yellow, and the corresponding translation will automatically appear in yellow in the adjacent window. This allows the user to easily find the specific sentence they are looking for.

Lexogoth - Lexical Toolbox

×

Create and edit context modules

?

Module: Context M1 | Author: MF | Subjects: la condition

Blok: 10 / 10

FRENCH:

Ne mangez pas ces putains de tacos à moins que vous n'ayez envie de vous faire bouffer le cerveau par des larves dégueulasses. / Des entreprises comme Monsanto font voter des lois sur mesure pour nous obliger à bouffer leur merde transgénique. / bouffer un sandwich / se bouffer la gueule, le nez / Ces problèmes, ça te bouffe de l'intérieur. / **Marre du stress qui te bouffe ?** / Ma voiture bouffe du pneu très vite, ça peut venir d'où ? Si la géométrie est bonne, c'est juste que tu roules comme une merde. / une voiture qui bouffe plus d'essence qu'un avion / se laisser bouffer par sa belle-mère, par son patron / bouffer du curé (LOC: à moins que, ce putain de)(EXP: se bouffer la gueule, le nez)

ENGLISH:

Don't eat those damn tacos unless you want to have your brain eaten by disgusting larvae. / Companies like Monsanto have laws passed to force us to eat their GMO crap. / to eat a sandwich / to fight each other, nose to nose / These problems eat you up from the inside. / **Tired of stress eating you up ?** / My car burns through tires very quickly, where could that come from ? If the geometry is correct, it's just that you drive like crap. / a car that drinks more gas than a plane / to allow yourself to be overwhelmed by your stepmother, by your boss / to have a deep resentment of the church

SEARCH WORDNEW BLOCKEDIT BLOCKEDIT NOTESREMOVE DUPLICATES

Show comments

Block 10 loaded.

Previous blockNext blockImport blockDelete block

Save allClose

(d) At the bottom, the user will find the usual **Previous Block / Next Block** navigation buttons. To expand the database, one can import new blocks via **Import block** or remove blocks from the database via **Delete block**. The **Save all** and **Exit** buttons save the database and close it.

(e) Above the navigation buttons, the user will find a few other useful options for editing the database:

SEARCH WORDNEW BLOCKEDIT BLOCKEDIT NOTESREMOVE DUPLICATES

Show comments

Block 10 loaded.

Previous blockNext blockImport blockDelete block

Save allClose

- **Search word:** Here, the user can search for one or more words in the current block, and the results will be highlighted in the block that is currently in focus. Note that this search finds all forms of the word (e.g., searching for the term "aller" will also find "vais," "allait," "allé," etc.). When searching for multiple words, it searches for all of them.
- **New block:** This allows the user to manually add a new text block. To keep the database consistent, the parity between French and English sentences is strictly enforced. In other words, there must be an equal number of French and English sentences. To remind the user of this, we have added a counter that tracks this equality for them:



The screenshot shows a dialog box titled "Add new block" with a close button (X) in the top right corner. Inside the dialog, there is a blue header text that says "Enter the texts (separate with '/'):". Below this, there are three input fields. The first is labeled "French:" and contains the placeholder text "French text...". The second is labeled "English:" and contains the placeholder text "English text...". The third is labeled "Comments:" and contains the placeholder text "Comments...". Above the "Comments:" label, it says "Segments: 1". At the bottom right of the dialog, there are two buttons: "Add" and "Cancel".

- **Edit block:** Here, the user can modify individual sentences within a block. Note: only the sentence currently highlighted in yellow (and its translation) can be changed here. This is again to keep the source database free of errors.
- **Edit comments:** Here, the user can modify the comments associated with a block.
- **Show comments button:** Clicking this opens a popup displaying the comments for the current block. While browsing the database, the comments popup stays synchronized with the blocks in the database, so the user does not need to click the button repeatedly. If the user does not wish to see comments while editing, they can simply close the window.
- **Delete duplicates button:** Use this to remove duplicate blocks from the database.

2.4.2 Practicing with Context Modules

Finally, there is the option to practice with text blocks using the **Practice context module** button.

This component allows the user to simply browse through the database and learn through reading—this is the **Show all** option available in the drop-down menu at the top of the screen. Here, too, the user can click on a sentence, and the translated equivalent will automatically be highlighted in yellow:

The screenshot shows the 'Practice context' window of the Lexogoth - Lexical Toolbox. The window has a title bar with the application name and a close button. Below the title bar, the title 'Practice context' is centered, followed by a help icon. The interface is divided into two main columns: 'FRENCH:' and 'ENGLISH:'. The 'FRENCH:' column contains a long paragraph of text with several sentences. The 'ENGLISH:' column contains the corresponding English translation of the French text. The text in the English column is highlighted in yellow. At the bottom of the window, there are four buttons: 'Previous block', 'Next block', 'Show comments', and 'Close'. The 'Previous block' and 'Next block' buttons have left and right arrows respectively. The 'Show comments' button has a speech bubble icon. The 'Close' button has an 'X' icon. The window also displays 'TRAINING: Context M1 | Author: MF', 'Mode: Show all', and 'Block: 1 / 10'.

Lexogoth - Lexical Toolbox

Practice context ?

TRAINING: Context M1 | Author: MF Mode: **Show all** Block: **1** / 10

FRENCH:

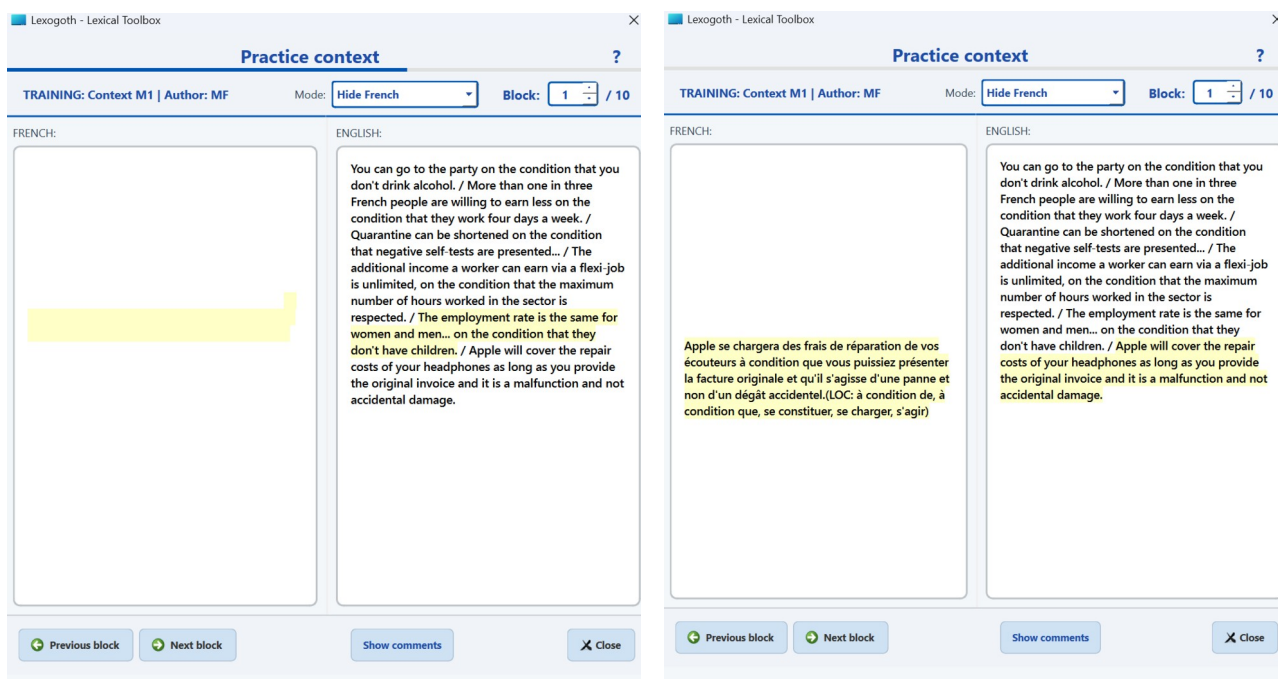
Tu peux aller à la fête à condition de ne pas boire d'alcool. / Plus d'un Français sur trois est prêt à gagner moins à condition de travailler 4 jours par semaine. / La quarantaine pourra être écourtée à condition de présenter des autotests négatifs... / Les revenus complémentaires qu'un travailleur peut se constituer via un flexi-job sont illimités, à condition de respecter le nombre maximal d'heures de travail applicable dans le secteur. / Le taux d'emploi est le même chez les femmes et les hommes ...à condition de ne pas avoir d'enfants. / Apple se chargera des frais de réparation de vos écouteurs à condition que vous puissiez présenter la facture originale et qu'il s'agisse d'une panne et non d'un dégât accidentel.(LOC: à condition de, à condition que, se constituer, se charger, s'agir)

ENGLISH:

You can go to the party on the condition that you don't drink alcohol. / More than one in three French people are willing to earn less on the condition that they work four days a week. / Quarantine can be shortened on the condition that negative self-tests are presented... / The additional income a worker can earn via a flexi-job is unlimited, on the condition that the maximum number of hours worked in the sector is respected. / The employment rate is the same for women and men... on the condition that they don't have children. / Apple will cover the repair costs of your headphones as long as you provide the original invoice and it is a malfunction and not accidental damage.

Previous block Next block Show comments Close

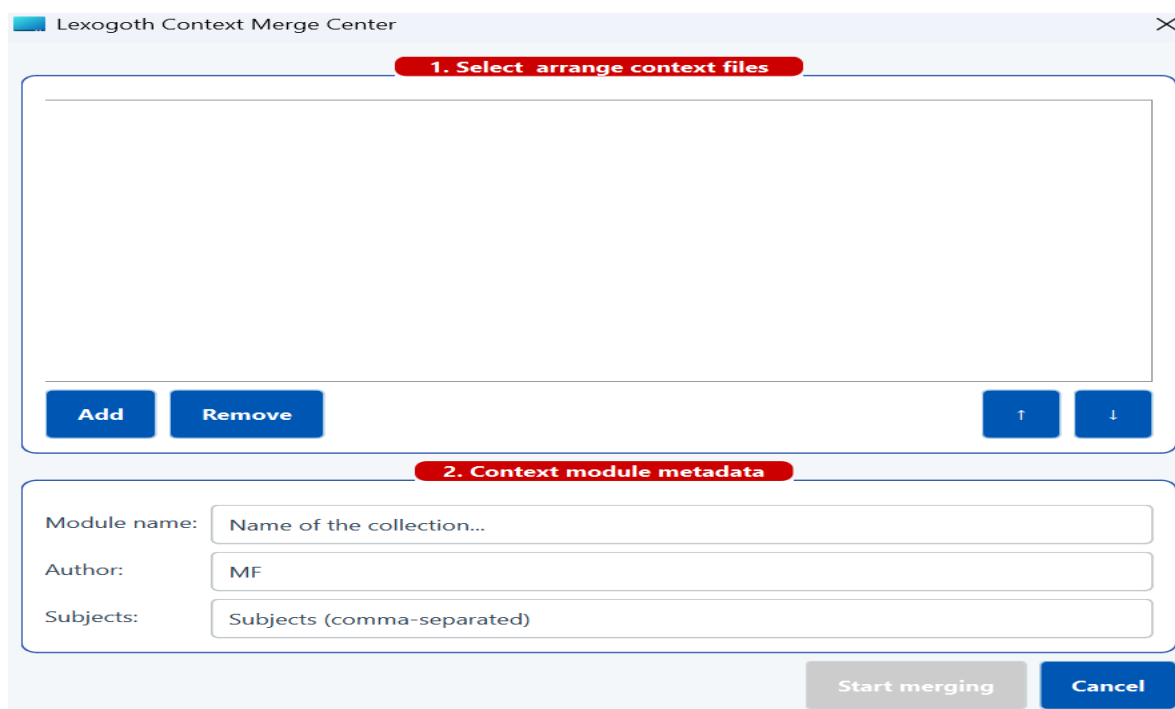
In addition, the user can choose to hide one of the languages (French or English). When you hide a language, the Toolbox starts a timer (currently set to approximately 30 seconds) that begins after you select a sentence. You then have x amount of seconds to guess the answer, and once the time expires, the answer (translation) appears in the hidden field next to it:



If desired, the user can also use the comments window here. Note that the file type used for context modules is .lexcb.

2.4.3 Merging Context Modules (Blocks extracted from Lexogoth or custom blocks compiled from own databases)

The user also has the ability to merge several block files into a single file. In the sidebar, select Context modules and then click the Merge context modules button. You will then see a popup asking you to select the files (extension: .lexcb) you wish to merge and to fill in the metadata field:



2.4.4 Exporting Context Blocks

The user also has the ability to export context blocks to a text file (RTF), allowing for easy editing in their preferred word processing software. Upon clicking **Export context modules**, the user can select a file (.lexcb) and toggle the **Export English** and **Export comments** options on or off. Once a file is selected, a preview of the blocks is loaded to allow for further inspection of the file.

Lexogoth Context Export Center

Source file

C:/Users/meul_/Desktop/kladwerk/Context M1.lexcb **Browsing...**

Export Options

French (Always) ☒ Export English ☒ Export comments

Select	French text (Preview)
☒	Tu peux aller à la fête à condition de ne pas boire d'alcool. / Plus d'un França...
☒	Je doute qu'il n'arrive à temps. / Je reste ici à moins que tu ne veuilles que j...
☒	Ne mangez pas ces putains de tacos à moins que vous n'ayez envie de vous faire b...
☒	crier au scandale / Le 13 janvier dernier, une femme a fait scandale à l'aéropor...
☒	A cause d'un fâcheux contretemps on ne pourra jamais terminer le projet avant la...
☒	Le roi doit sanctionner et promulguer les lois. / édicter une loi / publier une...
☒	Je partirai à moins que vous ne vouliez que je reste. / On ira en train, à moins...
☒	Je vais partir en vacances de ski à moins qu'il neige chez nous. / J'y vais tout...
☒	(a) Malheureusement, son tir face au FC Barcelone a heurté le poteau , alors qu'...

Cancel **Export to RTF**

2.5 The Collocations and Syntactic Patterns Module

2.5.1 Purpose of the module

The **Collocations and Syntax** module allows the user to perform targeted searches across the entire Lexogoth corpus, rather than being limited to locally created databases. This approach enables users to conduct their own corpus research and study specific linguistic phenomena on a large scale.

All three search fields are equipped with an **auto-correct function** to prevent frustrating typos that could hinder search results. When a search term is not found in the Lexogoth lexicon, the user is presented with a table containing the most likely alternatives. These suggestions are selected based on **Levenshtein distance**.

Finally, the user can export the content of the results window to an editable text file in **RTF format**.

2.5.2 Three Search Queries: Collocations, Co-occurrences, and Syntactic Patterns

The user can search the entire Lexogoth corpus in the following four ways:

Lexogoth - Lexical Toolbox

Search queries

1. Collocations

Enter lemma...

Search collocations

2. Co-occurrences

Words (separated by spaces)...

Max distance: 2

Zoek co-occurrences

3. Syntactic patterns
& individual words

Patterns or tags...

Max. dist: 1 Lit. Smart

Tag... Search syntaxis

4. Metadata & Comments

Search in annotations...

Search comments

Back

?

Export

Selected: 0 sentences

View selection

Clear selection

- collocations

Through this method, the user can find specific word combinations that are registered in Lexogoth as fixed, interesting, or frequent combinations, expressions, or collocations via (for the user) invisible Q-tags. For example, if you type "arriver" in the search field, you will receive the following response:

The screenshot displays the Lexogoth - Lexical Toolbox interface. On the left, the 'Search queries' sidebar is active, showing four categories: 1. Collocations, 2. Co-occurrences, 3. Syntactic patterns & individual words, and 4. Metadata & Comments. The 'Collocations' section is selected, with 'arriver' entered in the search field and the 'Search collocations' button clicked. The main results area on the right shows 'Resultaten voor: arriver 14 tags'. A list of 14 collocations follows, each with 'arriver' in blue and a translation in gray. The first collocation is 'arriver → aller au fait'. The 11th collocation is 'arriver → en arriver là', which is highlighted in yellow. Below this, a sentence is shown: 'C'est bien triste d'en arriver là. Et j'ose même pas imaginer la scène quand les secours sont arrivés sur les lieux.' The words 'en arriver là' are highlighted in yellow. Below the sentence is a red circled 'i' icon and the text 'S #13405002'. The 12th collocation is 'arriver → en arriver à'. The 13th is 'arriver → ne pas arriver à la cheville de'. The 14th is 'arriver → verbe'. The 15th is 'arriver → y arriver'. The 16th is 'arriver → ça n' arrive qu' aux autres'. At the bottom, there are buttons for 'Back', '?', and 'Export'. The status bar at the bottom indicates 'Selected: 0 sentences' and has buttons for 'View selection' and 'Clear selection'.

Lexogoth - Lexical Toolbox

Search queries

1. Collocations

arriver

Search collocations

2. Co-occurrences

Words (separated by spaces)...

Max distance: 2

Zoek co-occurrences

3. Syntactic patterns & individual words

Patterns or tags...

Max. dist: 1 Lit. Smart

Tag... Search syntaxis

4. Metadata & Comments

Search in annotations...

Search comments

Resultaten voor: **arriver** 14 tags

arriver → aller au fait

arriver → arriver au fait

arriver → arriver deuxième

arriver → arriver en tête

arriver → arriver à

arriver → arriver à ses fins

arriver → arriver à échéance

arriver → en arriver au point de

arriver → en arriver là

C'est bien triste d'en arriver là. Et j'ose même pas imaginer la scène quand les secours sont arrivés sur les lieux.

It is very sad to end up like this. And I don't even dare to imagine the scene when the rescue teams arrived on the scene.

i S #13405002

arriver → en arriver à

arriver → ne pas arriver à la cheville de

arriver → verbe

arriver → y arriver

arriver → ça n' arrive qu' aux autres

Selected: 0 sentences

View selection Clear selection

Back ? Export

Note that you can click on the base word (in blue) to see the found examples along with their translation. If there is a red circled **i** next to a sentence, it means that a comment from the Lexogoth corpus is available; click on it to read it. The gray **S:#number** serves to select the sentence: if you click on it, the sentence is added to your selections to potentially be exported later to a database or a Word file.

- Co-occurrences

This search method allows the user to find words that appear together in a sentence, and you can define the distance between the two words yourself, ranging from zero (directly next to each other) to 10 (ten words between them). A search for "réchauffement climatique" [distance 2] yields these results:

The screenshot shows the Lexogoth - Lexical Toolbox interface. On the left, the 'Search queries' panel has four sections: 1. Collocations (with 'arriver' entered), 2. Co-occurrences (with 'réchauffement climatique' entered and 'Max distance' set to 2), 3. Syntactic patterns & individual words (with 'Patterns or tags...' entered and 'Max. dist' set to 1), and 4. Metadata & Comments (with 'Search in annotations...' entered). The right panel displays the results for the Co-occurrences search. It shows three example sentences with the search term highlighted in yellow. Each sentence is preceded by a gray index number in a box and a red 'i' icon. The sentences are: 1. 'Planter des arbres: une panacée pour contrer le réchauffement climatique?' (ID: #10701001), 2. 'Il y a encore un an, on parlait de réchauffement climatique, mais depuis, le gouvernement italien a fait marche arrière sur le Green_Deal.' (ID: #9520000), and 3. 'Le réchauffement climatique va s'accélérer et s'accroître à l'échelle de la planète.' (ID: #13343002). At the bottom, there are buttons for 'Back', 'Export', 'View selection', and 'Clear selection', along with a status bar indicating 'Selected: 0 sentences'.

Here too, you can add example sentences to your selection by clicking on the gray index number; you can read comments by clicking on the red i.

- Syntactic patterns & individual words

With this search method, the user can look for interesting syntactic patterns in the corpus or simply search the entire corpus for individual words. Here, too, it is possible to define a distance between the words. Additionally, the user can use the asterisk (*) as a wildcard for a word. Some examples of valid search queries and their meanings:

- **avoir être +par:pas [distance 1]**: this allows you to find all passive past participles.
- **en +par:pre [distance 1]**: the Toolbox searches for all *gérondifs* where the distance between "en" and the present participle is a maximum of 1 word.

- **tellement * que [distance 0]**: this finds all patterns consisting of "tellement" + any random word + "que".
- **+ind:imp[actief]**: this finds all active imperfect indicative (*indicatifs imparfaits*) forms.

Note that the **Tag...** button serves to copy grammatical tags directly into the search field, ensuring the user does not make awkward typos. The labels used here are different (more extensive) than the labels used in the database module. An example for "en +par:pre":

The screenshot shows the Lexigoth - Lexical Toolbox interface. On the left, under 'Search queries', there are four sections: 1. Collocations (with 'arriver' and a 'Search collocations' button), 2. Co-occurrences (with 'réchauffement climatique', a 'Max distance' of 2, and a 'Zoek co-occurrences' button), 3. Syntactic patterns & individual words (with 'en +par:pre' in the search field, 'Max. dist' of 1, radio buttons for 'Lit.' and 'Smart' (selected), and buttons for 'Tag...' and 'Search syntaxis'), and 4. Metadata & Comments (with 'Search in annotations...' and a 'Search comments' button). At the bottom left are 'Back', a help icon, and 'Export' buttons. On the right, the 'Results for en +par:pre: 455' are displayed. Three results are shown, each with a French sentence, an English translation, and a tag ID. The first result is 'se délasser en lisant un roman' (ID: 0), the second is 'Les message qu'on veut faire passer en brûlant nos factures d'énergie, est que nous n'allons pas faire les frais de la pandémie et du réarmement de l'Ukraine.' (ID: 1), and the third is 'Depuis votre gîte, vous pourrez découvrir le charme des Ardennes en faisant une excursion dans la région.' (ID: 2). At the bottom right, there are buttons for 'View selection' and 'Clear selection', and a status bar indicating 'Selected: 0 sentences'.

Lexigoth - Lexical Toolbox

Search queries

1. Collocations

arriver

Search collocations

2. Co-occurrences

réchauffement climatique

Max distance: 2

Zoek co-occurrences

3. Syntactic patterns & individual words

en +par:pre

Max. dist: 1 Lit. Smart

Tag... Search syntaxis

4. Metadata & Comments

Search in annotations...

Search comments

Back ? Export

Results for en +par:pre: 455

se délasser en lisant un roman

to relax while reading a novel

[i] [S] ID: 0

Les message qu'on veut faire passer en brûlant nos factures d'énergie, est que nous n'allons pas faire les frais de la pandémie et du réarmement de l'Ukraine.

The message we want to send by burning our energy bills is that we will not bear the cost of the pandemic and the Ukraine's rearmament.

[i] [S] ID: 1

Depuis votre gîte, vous pourrez découvrir le charme des Ardennes en faisant une excursion dans la région.

From your gîte, you can discover the charm of the Ardennes by taking an excursion in the region.

[i] [S] ID: 2

Selected: 0 sentences

View selection Clear selection

Selection of the keywords for the Syntactic patterns & individual words search mode:

Select Tag (+TAG)

Step 1: Choose the form (default is Active). Step 2: Double-click a label.

Voice:

☒ Active
 ☐ Passive

Code	Beschrijving
(1) Word classes	
+NOM	Noun (homme, femme, chaise,...)
+ADJ	Adjective (grand, méchant, petit,...)
+ADV	Adverb (vite, en effet, effectivement,...)
+PRE	Preposition (près de, dans, pendant,...)
+PRO	Pronoun (je, il, nous, eux,...)
+DET	Determiner (le, les, ce, ces, nos, ton,...)
+VER	Verb (parler, affichez, comprendrez,...)
(2a) Simple tenses	
+ind:pre	Indicatif présent (je finis, nous croyons,...)
+ind:imp	Indicatif imparfait (je finissais, nous croyions,...)
+ind:fut	Indicatif futur simple (je finirai, nous croirons,...)
+cnd:pre	Conditionnel présent (je croirais, nous croirions,...)
+ind:pas	Indicatif passé simple (je finis, nous crûmes,...)
+sub:pre	Subjonctif présent (je finisse, nous croyions,...)
+sub:imp	Subjonctif imparfait (je finisse, nous crussions,...)
+imp:pre	Impératif présent (finis, croyons,...)
+inf	Infinitif présent (finir, croire,...)

Cancel

Insert

- Metadata & Comments

This search function is designed to provide direct access to the editorial layer of the Lexogoth corpus. While the other options focus on raw data and sentence structures, this section allows you to search through expert explanations, style guides, and usage advice.

- **Goal:** To find specific linguistic explanations and advice based on keywords.
- **Smart Correction:** The system is intelligent; if there is a typo in your search term, a pop-up with suggestions will automatically appear to guide you to the correct result. This also applies to the other three search options.
- **Visual Register Recognition:** Important nuances in language usage—such as terms like "formal, informal; familiar, literary, polished, everyday,

standard language, vulgar language..."—are immediately highlighted in the results with a red underline.

- **Direct Context:** From any comment, you can use the **voorbeelden>>>** link (highlighted in red) to switch directly to the relevant sentences in the French source material to see the theory in practice.

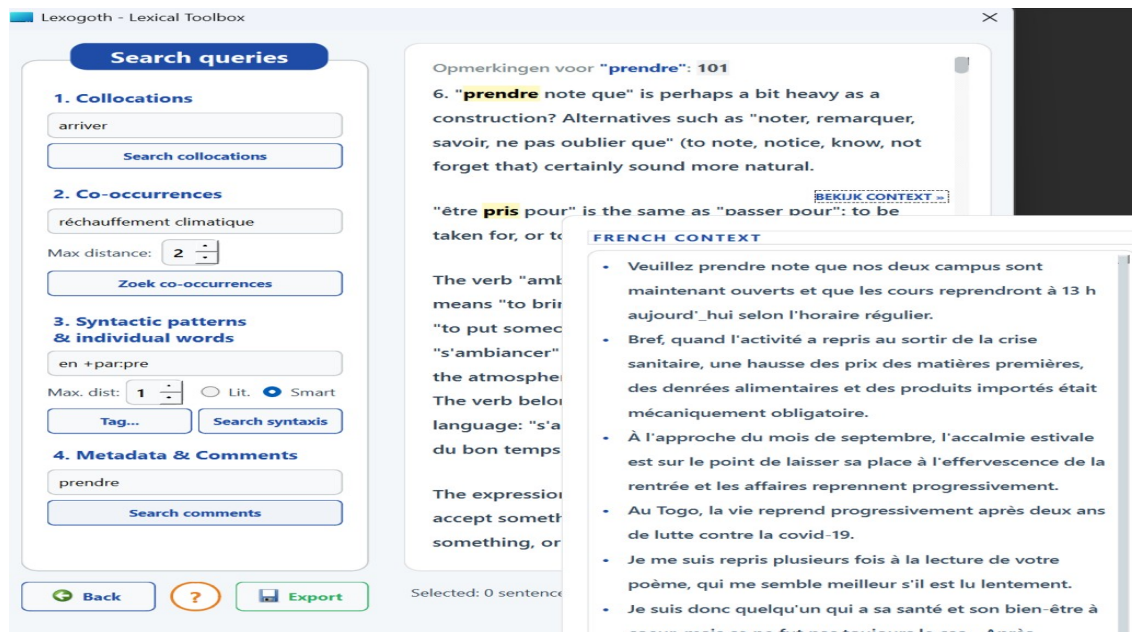


Photo above: An example of a search query for "prendre": all comments from the Lexogoth corpus in which "prendre" is mentioned are displayed. The example sentences can be consulted by clicking **voorbeelden>>>**.

2.5.3 Selecting items and reading comments

After performing a search, the results appear in the overview. To export sentences or save them to your own database, you must first select them. You can do this using the interactive labels next to each result. These labels all take the form "#number":

Selon le représentant, de toute évidence, l'absence de responsabilité pour Israël, y compris sur la question de la protection des enfants, a enhardi ce pays à commettre d'autres violations en toute impunité. <i>According to the representative, it is clear that the lack of accountability for Israel, including on the issue of child protection, has emboldened this country to commit further violations with total impunity.</i> [i][S]ID: 0	Selon le représentant, de toute évidence, l'absence de responsabilité pour Israël, y compris sur la question de la protection des enfants, a enhardi ce pays à commettre d'autres violations en toute impunité. <i>According to the representative, it is clear that the lack of accountability for Israel, including on the issue of child protection, has emboldened this country to commit further violations with total impunity.</i> [i][✓]ID: 0
Selon un sondage, en France, 74 pour cent des couples enterrent leur vie de célibataire avant de se marier; ce chiffre montre que l'enterrement de vie de célibataire est une coutume bien ancrée en France. <i>According to a survey, in France, 74 percent of couples hold a bachelor party before getting married; this figure shows that the bachelor party is a deeply rooted tradition in France.</i> [i][S]ID: 1	Selon un sondage, en France, 74 pour cent des couples enterrent leur vie de célibataire avant de se marier; ce chiffre montre que l'enterrement de vie de célibataire est une coutume bien ancrée en France. <i>According to a survey, in France, 74 percent of couples hold a bachelor party before getting married; this figure shows that the bachelor party is a deeply rooted tradition in France.</i> [i][✓]ID: 1

Labels are underlined in red above. Gray (first photo) means: not selected; green (second photo) means: selected. If there is a red i after the English translation, you can click on it to read the corresponding comment.

How to select a sentence?

For each search result, you will see a gray label on the left-hand side (e.g., [S]ID12).

- **Add:** Click the gray [S] label to select the sentence. The label will immediately turn green, and the text will change to [✓] (**Selected**).
- **Remove:** Did you select a sentence by mistake? Simply click the green label again. The sentence will be removed from your selection, and the label will turn back to gray.

Viewing the selection and the selection counter

At the bottom of the screen, you can see a live status bar indicating how many sentences you have selected (e.g., "Selected: 5 sentences").

- This counter updates immediately with every click, so you know exactly how large your collection is at any given moment.
- If you click the "**View selection**" button, you will get an overview of the sentences you have already selected.

Clearing everything at once

Do you want to start with a clean slate? Use the "**Clear selection**" button:

- All green labels in the overview will immediately revert to their standard gray color.
- The counter will be reset to zero.
- The internal list of selected numbers will be completely emptied.

Maintaining selection while browsing

An important advantage of Lexogoth is that your selection is preserved, even if you perform several different search queries in the same session. You can, therefore, combine results from a search for "condition" with results from a search for "agreement," and then export them together as a single package.

- The selection is only removed when you click "**Clear selection**" or the "**Back**" button.

Reading comments on sentences

If there is a red i inside a red circle after the English translation, you can click on it to read the corresponding comments from the Lexogoth corpus.

2.5.4 Exporting results to a text file or a database (.lexdb)

Once you have made a selection of sentences in the results overview, you can export them for external use or for further processing within the Lexogoth ecosystem. Clicking the **Export** button opens a menu with four specific methods:

A. Save as a Word file (.rtf)

This option is ideal for drafting reports. The selected sentences are exported to a Rich Text Format (RTF) file.

- **Formatting:** The export preserves the visual hierarchy from the browser, including the bold French source text and the italicized English translation.
- **Compatibility:** You can open the file directly in Microsoft Word, LibreOffice, or other word processors.
- **Content:** The interactive elements (such as ID links) are automatically removed for a clean layout.

B. Create a new Lexogoth Database (.lexdb)

Choose this option for durable storage within the native Lexogoth format. An additional dialog box will appear where you can record the following metadata:

- **File Name & Location:** You are required to select a storage folder before you can save the file.
- **Metadata:** You can enter your own name as the author and specify the subjects (e.g., "conjunctions with subjunctive").
- **Encryption:** The file is secured with a specific algorithm. This ensures that the data is only readable within the Lexogoth software and remains protected against unauthorized modifications.

C. Add to an existing Database

This function allows you to build your own subject-specific corpus over a longer period.

- **Merging:** The software opens an existing .lexdb file, decrypts the content, and appends the new selection at the end.
- **Duplicate Check:** The system automatically checks whether a sentence is already present in the target database. Only unique sentences are added to prevent cluttering your database.

D. Add to an existing Context Module

Students and teachers can also add selected examples to existing and new context modules. This feature is very useful if one wishes to group certain items to facilitate studying. For example, you can use various targeted

searches to bundle all conditional expressions into one module, within which there are different examples per block for each expression.

How do you handle this?

After selecting a few examples, click **Export** and choose either **Create new context block** or **Add to existing context block**. Depending on your choice, you may be prompted to enter the necessary details, such as a file name and subjects.

Next, a pop-up will appear asking which comments should be included in the export. Here, the user can decide which data—or if any data at all—should be copied along:

The image shows a 'Choose Export Method' dialog box with a close button (X) in the top right. Below the title bar, it says 'Export 2 selected sentences'. A sub-dialog titled 'DOCUMENT' and 'Lexogoth: Select annotations' is open. It has a close button (X) in the top right. The sub-dialog has a title 'Select Annotations' and a question 'Which comments do you want to add to the new blocks?'. There are two text boxes, each with a checkbox on the left. The first checkbox is checked and contains the text: 'The verb "enhardir" means "to make someone more self-confident, to encourage; to give someone daring or nerve; to make someone bolder". The reflexive version "s'enhardir" means "to become more self-confident, more bold; to dare to". The verb is not often used in current language and seems to belong to very formal or literary language. In standard language, you can usually replace "enhardir" with more common synonyms such as "encourager, inciter, pousser". See also the adjective "hardi" in Lexogoth.' The second checkbox is checked and contains the text: '1. "Un enterrement de vie de célibataire" is a party organized for someone who is about to get married; in Dutch, we refer to it as a "vrijgezellenavond" (Belgium) or "vrijgezellenfeest." The use of the noun "célibataire" has the advantage that it can refer to both the man's party ('enterrement de vie de garçon' or 'EVG') and the woman's party ('enterrement de vie de jeune fille' or 'EVJF'). 2. Note that "jeune fille" refers to an "unmarried woman." 3. The adjective "arrosé" means "where a lot of alcohol is consumed"; synonym: "alcooolisé." 4. "Un SAM" is the French counterpart of the Belgian "BOB," i.e., a'. Below the text boxes are two radio buttons: 'All' and 'Nothing'. At the bottom right are 'Confirm' and 'Cancel' buttons. Below the sub-dialog, there is a button labeled 'Toevoegen aan bestaand Contextblok'.

Finally, the user is redirected to the context editor screen:

The image shows the 'Lexogoth Context Editor' screen. It has a close button (X) in the top right. The screen is divided into two main sections. The left section contains a list of text blocks, each with a checkbox on the left. The first checkbox is checked and contains the text: 'Tu peux aller à la fête à condition de ne pas boire d'alcool. / Plus d'un Français sur trois est prêt à gagner moins à condition de travailler 4 jours par semaine. / La quarantaine pourra être écourtée à condition de présenter des autotests négatifs. / Les revenus complémentaires qu'un travailleur peut se constituer via un flexi-job sont illimités, à condition de respecter le nombre maximal d'heures de travail applicable dans le secteur. / Le taux d'emploi est le même chez les femmes et les hommes. / À condition de ne pas'. The second checkbox is checked and contains the text: 'Je doute qu'il n'arrive à temps. / Je reste ici à moins que tu ne veuilles que je parte. / Elle est plus forte que je ne croyais. (LOC: à moins que, à temps) / Selon un sondage, en France, 74 pour cent des couples enterrent leur vie de célibataire avant de se marier; ce chiffre montre que l'enterrement de vie de célibataire est une coutume bien ancrée en France.'. The third checkbox is checked and contains the text: 'Ne mangez pas ces putains de tacos à moins que vous n'ayez envie de vous faire bouffer le cerveau par des larves dégoulassées. / Des entreprises comme Monsanto font voter des lois sur mesure pour nous obliger à bouffer leur merde transgénique. / bouffer un sandwich / se bouffer la gueule, le nez / Ces problèmes, ça te bouffe de l'intérieur. / Marre du stress qui te bouffe ? / Ma voiture bouffe du pneu très vite, ça peut venir d'où ? Si la géométrie est bonne, c'est juste que tu roules comme une merde. / une voiture.'. The fourth checkbox is checked and contains the text: 'crier au scandale / Le 13 janvier dernier, une femme a fait scandale à l'aéroport international de Miami en se dénudant devant tout le monde. / une affaire qui a fait scandale au sein du gouvernement (-/-) un film, un roman, une publication... qui a fait scandale / le scandale a éclaté en décembre 2024. / ça fait deux jours au moins que le scandale a pété. / Le condamné avait passé à tabac le docteur en le frappant avec une bouteille après avoir fait du scandale dans son cabinet en cassant les ampoules, et l'avoir'. The fifth checkbox is checked and contains the text: 'A cause d'un fâcheux contretemps on ne pourra jamais terminer le projet avant la date butoir fixée, / un léger contretemps / L'Iran expulsera les inspecteurs des Nations Unies à moins que les États-Unis ne lèvent leurs sanctions avant le 21 février, date butoir fixée par le parlement iranien. / Quelle est la date butoir pour m'inscrire au concours ? / la date butoir de paiement d'une facture, du loyer mensuel... / savoir travailler sous pression et respecter les deadlines (LOC: à cause de, à moins que) SYN: une date butoir: une date'. The sixth checkbox is checked and contains the text: 'Le roi doit sanctionner et promulguer les lois, / édicter une loi / publier une loi au Moniteur belge / la promulgation d'une loi est suivie de sa publication au Journal Officiel / La loi entre en vigueur le dixième jour après sa publication au Moniteur belge à moins que la loi ne fixe elle-même une autre date. (BOU) LOC: Journal Officiel, à moins que'. The right section contains two text boxes. The first text box contains the text: 'Selon le représentant, de toute évidence, l'absence de responsabilité pour Israël, y compris sur la question de la protection des enfants, a enhardi ce pays à commettre d'autres violations en toute impunité.'. The second text box contains the text: 'Selon un sondage, en France, 74 pour cent des couples enterrent leur vie de célibataire avant de se marier; ce chiffre montre que l'enterrement de vie de célibataire est une coutume bien ancrée en France.'. At the bottom of the screen are buttons: 'New empty block', 'Delete block', 'Help', 'Save', and 'Cancel'.

On the left side, you will find the current context blocks, and on the right side, the selected sentences. You can now drag the sentences from right to left and drop them into the desired block. Once you reach the target destination, you can release the block, and the sentence will be placed within that context block.

- **Adding sentences:** You can add multiple sentences to the same block.
- **Creating new blocks:** If you wish to create a new block, simply press the green button at the bottom: **+New empty block**.
- **Saving:** When you are finished, you can conclude by saving everything by pressing the **Save** button.
- **Automatic formatting:** If a block already contains sentences, Lexogoth will automatically insert a sentence separator ("/") and simply append the new sentence to the existing block.

Below are several screenshots of the export function:

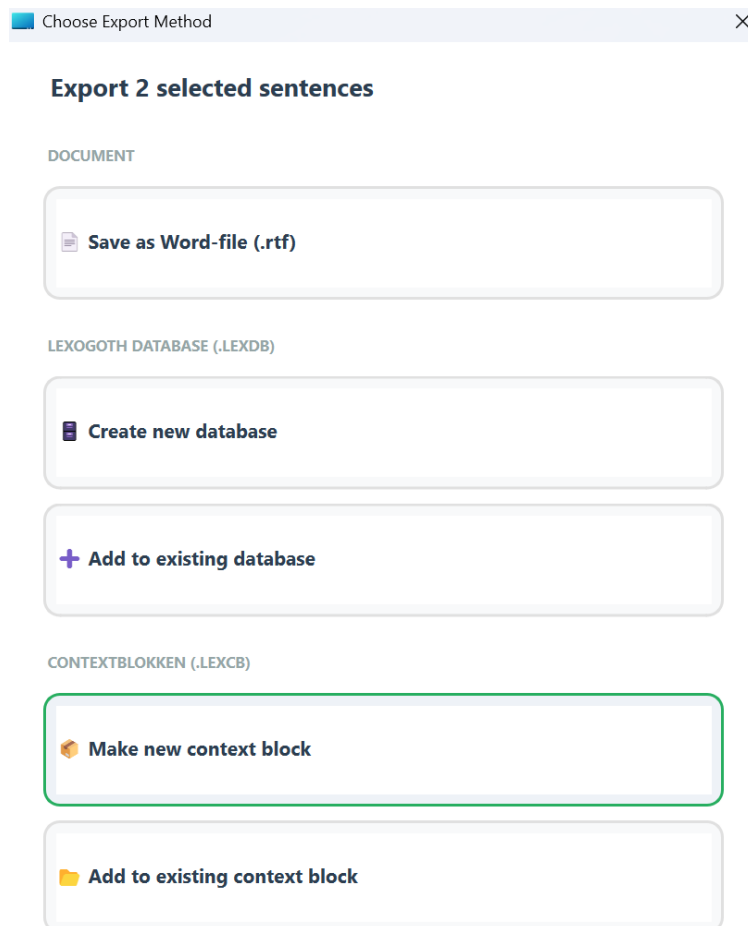
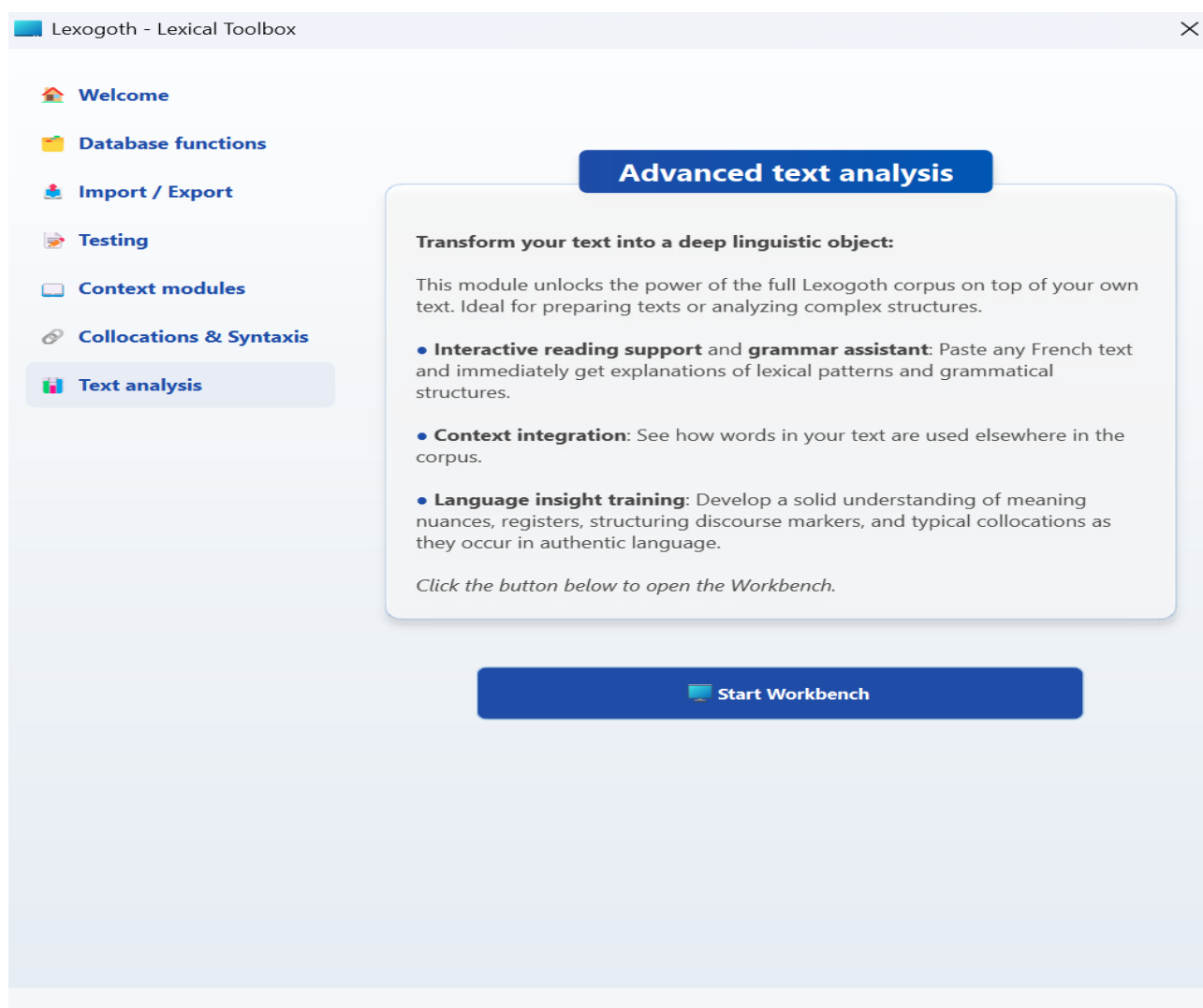


Photo above: Various save options...

2.6 Module 6: The Lexogoth Workbench (Text Analysis)

2.6.1 What is the Workbench and what can you do with it: a concise overview.

The Workbench (also known as the Text Analysis module) can be opened by selecting Text Analysis in the Toolbox side panel and then clicking Start Workbench:



The Workbench allows you to do the following:

- **POS Tagging & MWE Identification:** Copy arbitrary French texts into the text box to assign a POS tag (part-of-speech category) to every word. Furthermore, MWEs (multi-word expressions like *pomme de terre* or *faire partie de*) are clearly identified in the text. This information is conveniently processed in a sub-window below the text.
- **Advanced Text Search:** Search texts using advanced functions: searching by grammatical labels, using wildcards, searching for syntactic patterns, collocations, specific searches for grammatical tenses, morphological patterns, and searching with variable distances between different terms.

- **Parallel Search:** Words from the texts can be looked up in the Lexogoth corpus (both examples and comments) via parallel search queries, which are then displayed in two clear side panels to support text analysis.
- **Didactic Editing:** Numerous options for didactic text editing: adding footnotes to the text, adjusting labels, and exporting text to an easily editable file format (RTF).
- **Concordancer:** Possibilities to place examples found in the text alongside examples from the Lexogoth corpus and compare them via a convenient Concordancer.
- **Pattern Extraction:** Extract patterns and constructions from a text and export them to a text file for further editing.
- **Data Export:** POS-tag texts and write the output to a text file (TXT format) for potential further processing by computer.
- **File Format:** Processed files are saved in a proprietary format, recognizable by the ".lexdoc" file extension.

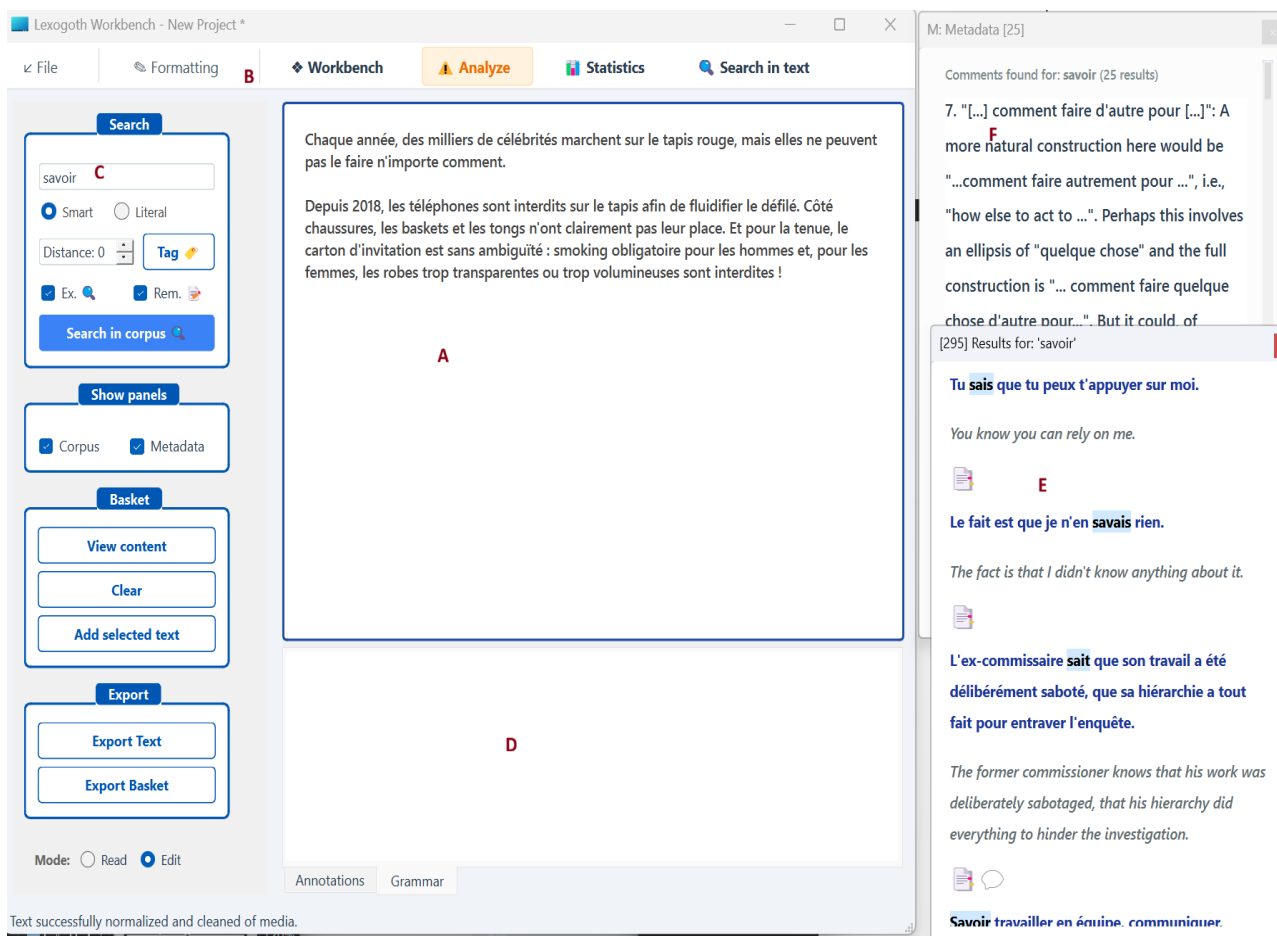
The functions above offer many didactic possibilities for teachers and students of the French language. We have attempted to make the various functions offered by the Workbench as simple as possible in the interface by hiding the sometimes complex indexing logic and search algorithms. This allows the user, even with minimal theoretical knowledge of NLP and corpus linguistics, to utilize many possibilities that were previously only accessible via specialized and sometimes difficult-to-use software. We will discuss these various possibilities in detail below.

2.6.2 The main screen, menu, left panel, and the two side panels: a concise overview.

We distinguish the following sections (in the photo below):

- **A: The text area:** The screen where texts are copied and analyzed.
- **B: The menu bar:** Features the various functions. On the far right, you will also find the search function for the text.
- **C: [Left panel]:**
 - The search function to query the Lexogoth corpus (and comments).
 - The control box to show or hide independent side windows (windows containing examples and comments from the Lexogoth corpus).
 - **The Basket ("het Mandje"):** Keeps track of which examples you want to export to a text file later. Under **Export**, you can choose to export either the contents of the Basket or the current text from the text area.
 - **Mode selection:** If you check the **Read** option, you can temporarily prevent accidental changes to the text while reading.

- **D: The annotation and grammar screen:** Here you can read the content of footnotes or view (and add) the POS tags for the various words.
- **E: The side panel for examples:** Where the found examples from the Lexogoth corpus are displayed. The speech bubble icon indicates that there is a comment associated with the sentence; the other symbol allows you to copy the example sentence to the Basket.
- **F: The comment panel:** This is where the comments regarding the searched word from the Lexogoth corpus are displayed.

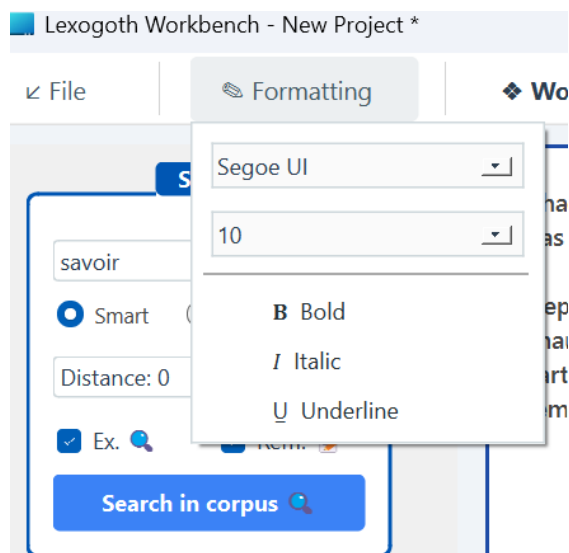
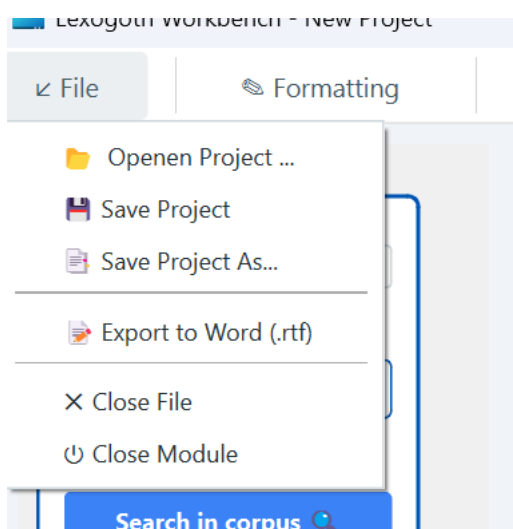


2.6.3 The Menu Bar

2.6.3.1 Menu Bar: File and Format

The **File** and **Format** menu items are self-explanatory:

- **File:** Used for standard operations such as creating new documents, opening existing files, saving your progress, and managing imports or exports.
- **Format:** Provides tools to adjust the visual presentation of your text, such as font settings, text alignment, and other styling options to ensure the document is clear and easy to read.



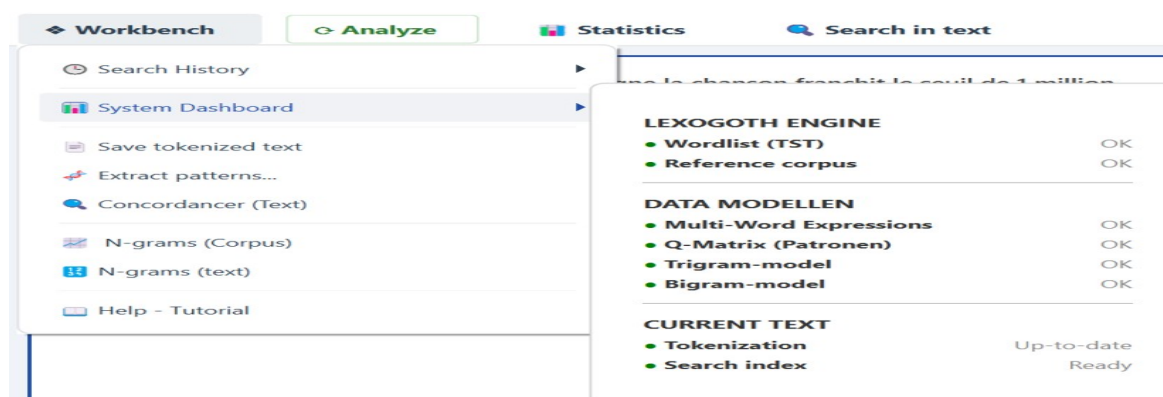
Under the **File** menu, users can open and load a *.lexdoc* file (project), save or close the current project, and export it to a Word-editable file. Note that the *.lexdoc* format is not a standard text file; it is a specialized file type for the Workbench that saves annotations, POS tags (grammatical labels), layout, and the text itself as a single, integrated unit.

In the **Formatting** menu, users can select various options to style the text, ranging from **bold**, *italics*, and underlining to selecting the font type and text size.

2.6.3.2 Menu Bar: Workbench

Via the **Workbench** menu item, the user can:

- **Consult search history:** Review and repeat previous search queries in both the corpus and the text (with a maximum of ten stored searches per category).
- **Access the System Dashboard:** This displays the status of the various data modules required for text analysis. If "OK" appears next to an item, it means the module has loaded correctly. If a module fails to load, search results may be less accurate or potentially fail entirely. In such cases, a reinstallation is recommended.



- **Save tokenized text:** This command allows the user to save a text in .txt format (Windows CRLF - UTF-8). The text is saved as a series of tokens, each accompanied by its corresponding POS tag. For example, as shown below:

=====						
LEXGOTH TOKEN EXPORT & ANALYSIS REPORT						
=====						
Date	:	2026-05-23 22:12				
Total Tokens	:	89				
Unique Lemmas	:	57				
MWE Units	:	1				
Q-Tag Matches	:	1				
Total Idioms	:	2 (cleaned)				

SURFACE	LEMMA	TAG	MOOD/TENSE	PERS	GEN	NUM INFO

Chaque	chaque	ADJ	num	-	s	
année	année	NOM		f	s	
,	,	PUNCT				PUNCT
des	des	DET	indéfini			
milliers	millier	NOM			m	p
de	de	PRE				
célébrités	célébrité	NOM			f	p
marchent	marcher	VER	ind:pre3			p
sur	sur	PRE				
le	le	DET	défini	m		
tapis	tapis	NOM		m	s	
rouge	rouge	ADJ		m	s	
,	,	PUNCT				PUNCT
mais	mais	CON				
elles	elles	PRO				
ne	ne	ADV				
peuvent	pouvoir	VER	ind:pre3		p	
pas	pas	ADV				
le	le	PRO				
faire	faire	VER	inf			
n'	ne	ADV				
importe	importer	VER	ind:pre3			s
comment	comment	ADV				
.	.	PUNCT				PUNCT
Depuis	depuis	PRE				
2018	2018	ADJ	num			
,	,	PUNCT				PUNCT
les	les	DET	défini	m		
téléphones	téléphone	NOM			m	p
sont	être	VER	ind:pre3		p	
interdits	interdire	VER		par:pas		p
sur	sur	PRE				
le	le	DET	défini	m		
tapis	tapis	NOM		m	s	
afin	afin de	PRE	mwe			MWE-Part(1/1) Q-Match: afin de
de	de	PREP	mwe			MWE-Part(2/1) Q-Match: afin de
fluidifier	fluidifier	VER	inf			
le	le	DET	défini			
défilé	défilé	NOM		m	s	
.	.	PUNCT				PUNCT

Extract patterns from text: This command allows the user to select specific interesting words and patterns and automatically extract them from the text. These patterns are then displayed in a clear, organized pop-up window that includes various export options:

Extract Patterns from Text

Enter search terms (max. 5):

alors 0 wrd Start + X

quota 0 wrd Start + X

Start Analysis Cancel

Looking for patterns with “dépasser” and “plus”

And these are the results produced by the search:

Lexogoth Extraction Results

Lexogoth Extractie Rapport

Gegenereerd op 15-07-2026 11:55

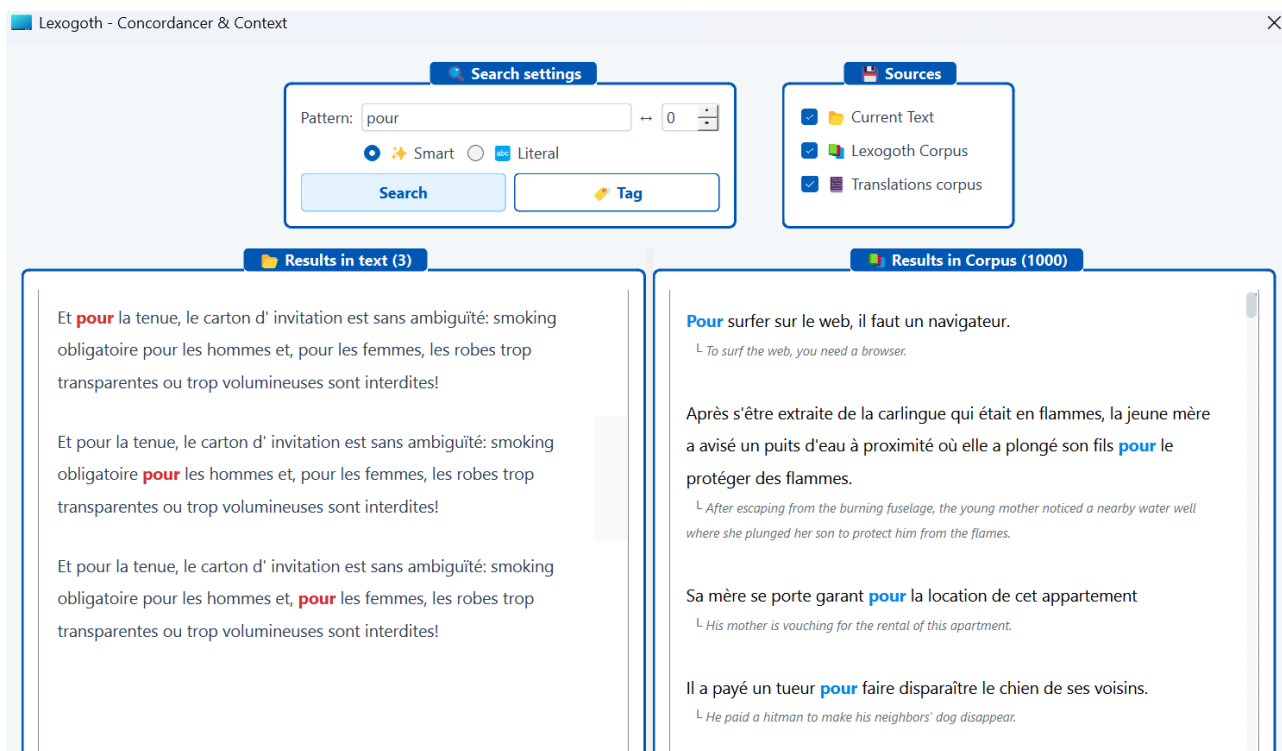
PATROON: alors

- Alors, quand le dernier rapport de Greenpeace réclame à la Belgique, notamment, de réduire ses aides aux industries polluantes pour investir dans l'énergie propre, est-ce aussi simple ?
- Alors, comment ça fonctionne ?

PATROON: quota

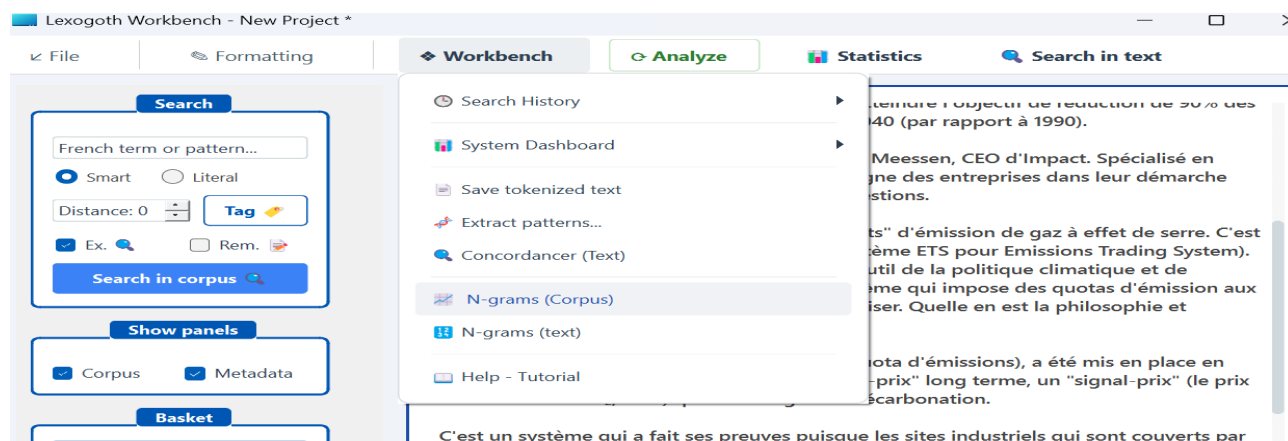
- Et elles peuvent aussi marchander les quotas d'émission de gaz à effet de serre, comme le prévoit le marché européen du carbone.
- En coulisse, s'affrontent deux blocs et deux visions : les détracteurs des quotas (avec en tête l'Italie) qui réclament leur assouplissement voire leur suspension et les pays favorables au système (l'Allemagne, la Suède) qui soutiennent des règles climatiques strictes (baisse des quotas, annulation des quotas gratuits) pour atteindre l'objectif de réduction de 90% des émissions nettes de gaz à effet de serre d'ici 2040 (par rapport à 1990).
- Ce système qui impose des quotas d'émission aux industries, leur permet aussi de les commercialiser.
- Le système ETS (pour système d'échange de quota d'émissions), a été mis en place en Europe depuis 2005.
- En fait, vous avez toute une série de sites industriels à qui on donne un quota d'émissions annuel.
- Ce quota, il va descendre progressivement.

The **Concordancer** function allows users to search for patterns, words, or labels within the text and compare them side-by-side with the same search results from the Lexogoth corpus. To combine these views, the user must check the **Lexogoth corpus** option; if desired, they can also display translations by selecting **Corpus translations**.



Here, too, the user has the option to export the search results to a text file for further study or editing.

In the Workbench menu, the user can also perform N-gram analyses on both the text and the Lexogoth corpus:



When you click the N-gram Analysis menu item, a popup appears showing the N-gram analysis for N = 2, 3, and 4, based on either the text or the corpus. Note that the column headers “Rank”, “Combinations”, and “Frequency” are clickable, allowing you to sort the data:

Corpus N-Gram Analysis

EXPLORATION OF RECURRENT PATTERNS (N-GRAMS)

Double-click a pattern to view the sentences and add a specific sentence to the Basket.

Live-filter results by search term (e.g. 'stade', 'police', 'il est')...

Word pairs (N=2)

Three-word combinations (N=3)

Fixed patterns (N=4)

Rank	Combinations (N-gram)	Frequency
124	le temps de	25
125	on ne peut	25
126	par rapport à	25
127	qu' il s'	25
128	que je n'	25
129	se mettre en	25
130	au long de	24
131	droits de l'	24
132	en matière de	24
133	en raison de	24

View Context / Sentences

Close

When the user then double-clicks an item, all examples are displayed in a clear, well-structured popup:

Sentences with N-gram: "par rapport à"

Found contexts (25 unique sentences):

L' indice général est en hausse de quatre points **par rapport à** la même période l' an dernier.

Si ce film peut faire avancer le schmilblick et un peu ouvrir les mentalités **par rapport à** la différence, c' est gagné.

Les défaillances d' entreprises sont paradoxalement à leur plus bas historique, avec 36 pour cent de faillites en moins en 2020 **par rapport à** 2019.

Le taux de chômage des 15-30 ans enregistre une légère hausse **par rapport à** l' année dernière.

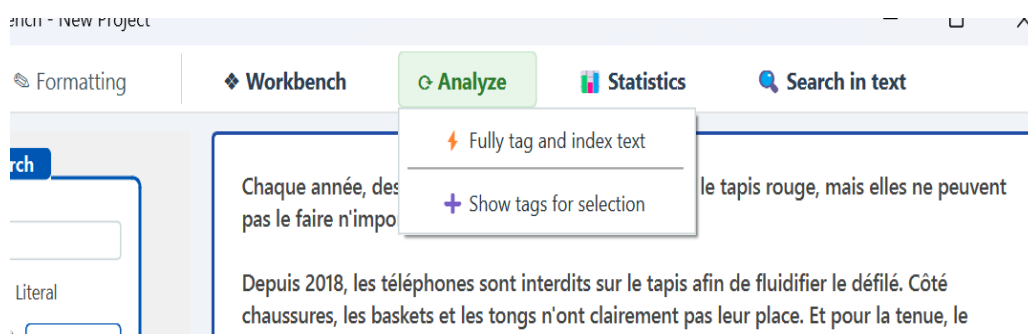
Add selected sentence to Basket

Close

The user can then double-click one or more examples to copy them into the Basket, allowing the examples to be exported later. Note that N-gram analysis for texts only becomes truly meaningful once the input texts are sufficiently large.

- **Help:** Finally, there is the **Help** menu item. Clicking this will launch an external PDF viewer (depending on your operating system's default settings) to display a short, clear tutorial on using the Workbench.

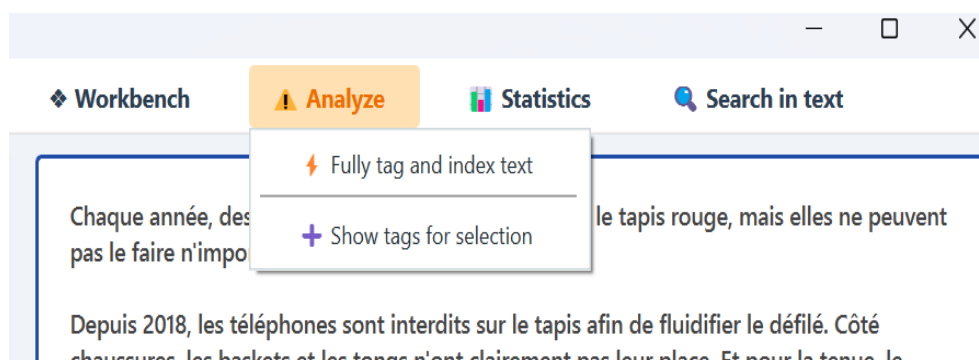
2.6.3.3 Menu Bar: The *Analyze* Button



Here, the user has three options:

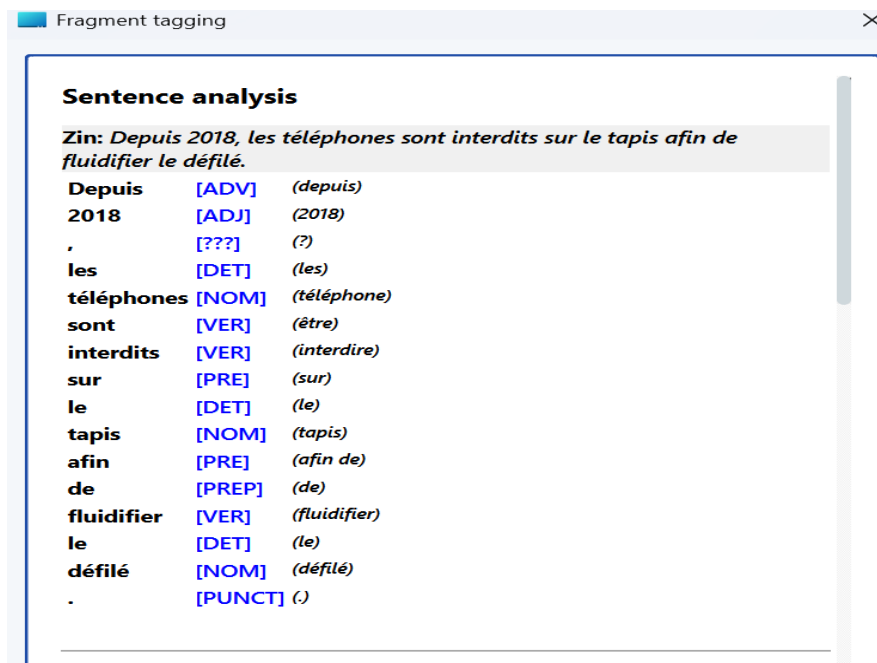
- The option to fully tag and index text: in this process, the text is completely scanned again, a grammatical label is assigned to each word, and the MWEs (expressions consisting of multiple words) are recalculated. After this tagging procedure, the entire text is indexed, meaning it is prepared to be searched efficiently and at lightning speed.

If the user makes changes to the text (deleting, adding, correcting, copying and pasting words, etc.), the index (position) of the words will shift and the search function will no longer be able to correctly locate the searched word. As soon as the text has been changed, you will notice that the *Analyze* button turns orange. We recommend that the user regularly chooses the item to fully tag and scan the text so that the word index remains up to date:



Because a user tends to forget this occasionally, we automatically have the scanner re-tag and re-index the entire text upon activation of "Search in text," ensuring that search queries always run on a perfectly indexed and tagged text.

The option "Show tags for selection" allows the user to tag individual sentences, and the result is then displayed in a clear, organized pop-up:

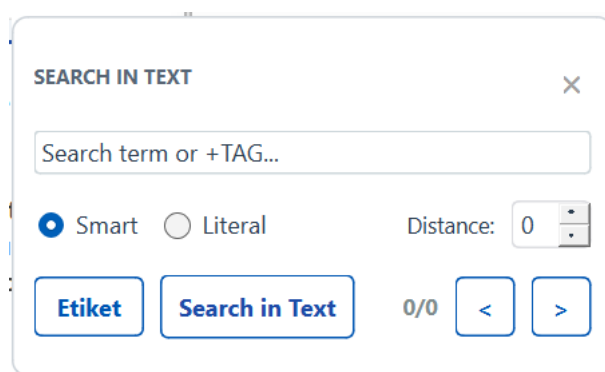


2.6.3.4 Menu Bar: The *Statistics* button

Via this button, the user can quickly view several figures regarding the analyzed text, such as the number of sentences, the average sentence length, the number of tokens, etc.

2.6.3.5 Menu Bar: The *Search in text* button

This function is very important for anyone wishing to perform an in-depth analysis of the text; as soon as the user clicks the button from the menu, the text is automatically tagged and indexed once more:



This small search window contains many search options that we will now discuss in detail.

(1) You have the basic options *Smart* and *Literal*:

- **Literal:** Searches for a literal search term. If you type “était,” for example, it will only search for that form in the text and not the derived forms “étaient, sommes, ...”.
- **Smart:** The search function performs a broader task: every search term is expanded to all its forms. If you then type “était,” it will search for all forms of “être” in the text.

(2) If you want to search for multiple terms, you can also enter the **maximum distance** between the different terms via the **Distance** drop-down box. For distance 0, the terms must be adjacent; for 1, there can be a maximum of 1 word between the terms. For example, if you enter the search term “tellement que” with distance 1, it will search the text for all constructions of “tellement que” and “tellement + RANDOM WORD + que”.

(3) The user can also perform two types of searches with **wildcards**:

- If you type the wildcards separately from a word, the wildcard counts as an individual word. For example, the search query “ne * * pas” means: search for all sentences where “ne pas” occurs with exactly 2 words in between.
- If you type wildcards attached to a word (or part of one), you can search for morphological patterns. There is the prefix search where the wildcard is at the end: if you type “donn*”, the program will scan the text and retain all words starting with “donn” (prefix). For the suffix search, we do the opposite; here we search for words ending in a certain combination. For example, the search query “*re” will be interpreted as “search for all words that end in ‘re’” (suffix).

(4) It is also possible to perform searches based on **grammatical tags**:

- **Part of speech:** To do this, the user must click the **Tag** button and then select their desired part of speech. We recommend always using this selection window so that no typos can occur.
- **Grammatical tenses:** We also offer the possibility to search for grammatical tenses in the text. For example, the user can search for all *passés composés* in a text. For this, it is also best to use the Tag selection window to ensure no mistakes are made. Note: for grammatical tenses, you must also choose between active and passive.

Select Tag (+TAG)

Step 1: Choose the form (default is Active). Step 2: Double-click a label.

Active

Passive

Code

Description

Code	Description
(1) Word classes	
+NOM	Noun (homme, femme, chaise,...)
+ADJ	Adjective (grand, méchant, petit,...)
+ADV	Adverb (vite, en effet, effectivement,...)
+PRE	Preposition (près de, dans, pendant,...)
+PRO	Pronoun (je, il, nous, eux,...)
+DET	Determiner (le, les, ce, ces, nos, ton,...)
+VER	Verb (parler, affichez, comprendrez,...)
(2a) Simple tenses	
+ind:pre	Indicatif présent (je finis, nous croyons,...)
+ind:imp	Indicatif imparfait (je finissais, nous croyions,...)
+ind:fut	Indicatif futur simple (je finirai, nous croirons,...)
+cnd:pre	Conditionnel présent (je croirais, nous croirions,...)
+ind:pas	Indicatif passé simple (je finis, nous crûmes,...)
+sub:pre	Subjonctif présent (je finisse, nous croyions,...)
+sub:imp	Subjonctif imparfait (je finisse, nous crussions,...)
+imp:pre	Impératif présent (finis, croyons,...)
+inf	Infinitif présent (finir, croire,...)

Cancel

Insert

(5) **Targeted part-of-speech search:** It is a problem if you enter a search term like “durée”: that single term can be both a verb and a noun. After all, the current smart search expands search terms to all possibilities by default. That is why we provide a targeted search. If you want to find the noun form of “durée,” you simply type “durée=NOM” in the search field, and it will search for all instances of the noun and completely ignore examples of “durée / durer / dure / ...”. Naturally, you can also combine this way of searching with the other search options mentioned above. You can add the following tags to a word:

“=NOM” / “=VER” / “=ADJ” / “=ADV” / “=DET” / “=PRO” / “=PRE”

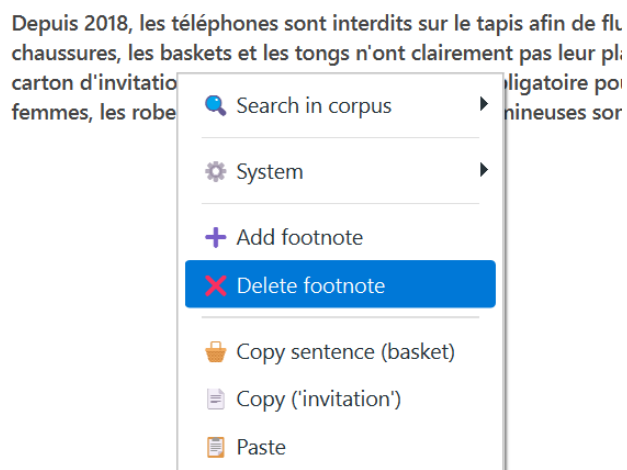
Note that for labels in this search query, you do not use the "Tag" (*Etiket*) window, but type them directly into the search field; also, always use capital letters for these labels!

Examples of valid search queries are:

- **compris=ADJ:** this allows you to search for the use of *compris* as an adjective.
- **+VER revue=NOM [distance=4]:** this allows you to search for *revue* as a noun with a verb within the 4 preceding words.

Note that these search queries may sometimes contain results that deviate from the requested search term. This is due to the POS-tagger, which may sometimes place words in the wrong category (e.g., occasionally tagging *Revue* as a noun while it is a verb). Normally, this does not happen often because our POS-tagger achieves a correctness percentage of 97 percent (for main tags), which means that 97 out of 100 examined words receive a correct label, but 3 do not. We verified these figures manually on various French-language texts from news media (RTBF), for which we manually checked 1000 tokens each time.

2.6.4 The context menu (right-click in the text field)



An overview of the different functions:

(a) **Search in corpus** (Lexogoth): If you are positioned on a word, the option *Search in corpus* will search for the word or the selected words in the Lexogoth corpus. The results will then appear in the right side panel (examples).

(b) Under **System**, you can choose to fully *re-tag* and *re-index* the text, or you can choose the option *Tag selection* to tag a specific portion of text and display it in a pop-up window.

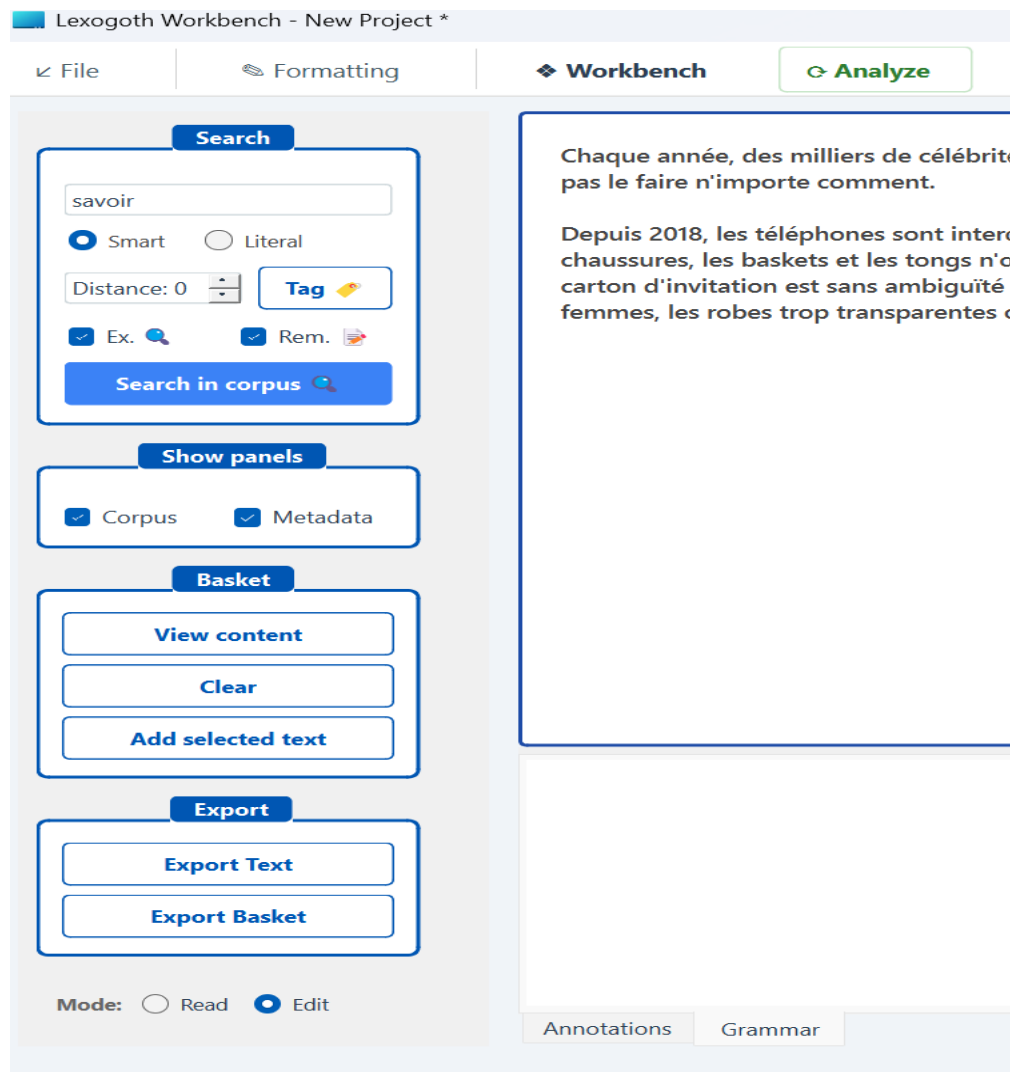
(c) With the options **Add footnote** and **Delete footnote**, you can add and remove annotations (footnotes) to the text. As soon as you click the option, the tab for footnotes appears below the text field. Note that you can also remove footnotes via this screen and add multiple footnotes per word.

(d) With the option **Copy sentence (basket)**, the user can copy a selected sentence to the Basket for later export to a text file. The option **Copy sentence (standard)** allows the user to copy a sentence to the Windows clipboard in order to subsequently paste it into any other location.

(e) The option **Paste**: this allows the user to copy texts into the Workbench text field. For safety reasons, we have incorporated strict validation that normalizes all letters (such as æ) and characters (such as apostrophes, quotation marks, etc.) that could potentially cause processing problems into characters recognized by Lexogoth. Excessive whitespace between words is also reduced to 1.

2.6.5 Left panel & separate side panels

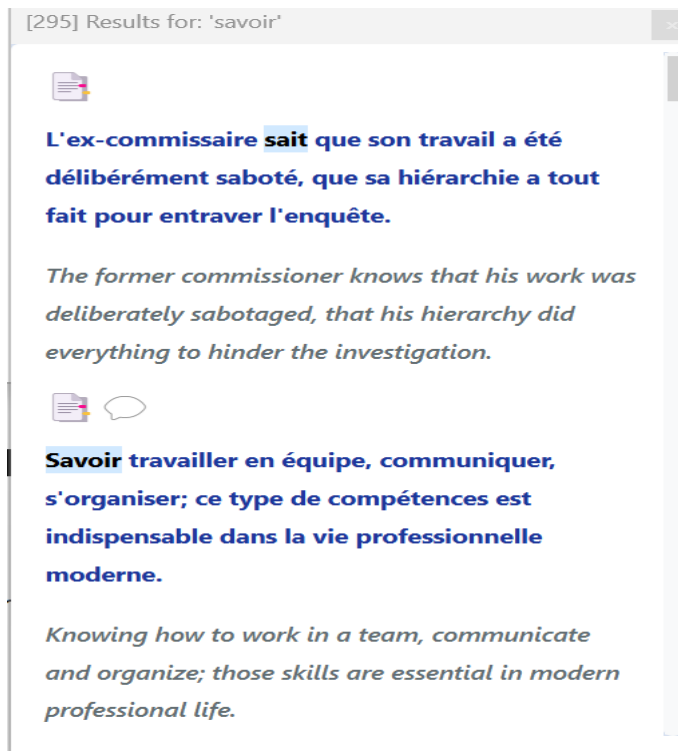
2.6.5.1 The Search zone



The left panel is divided into 5 sections:

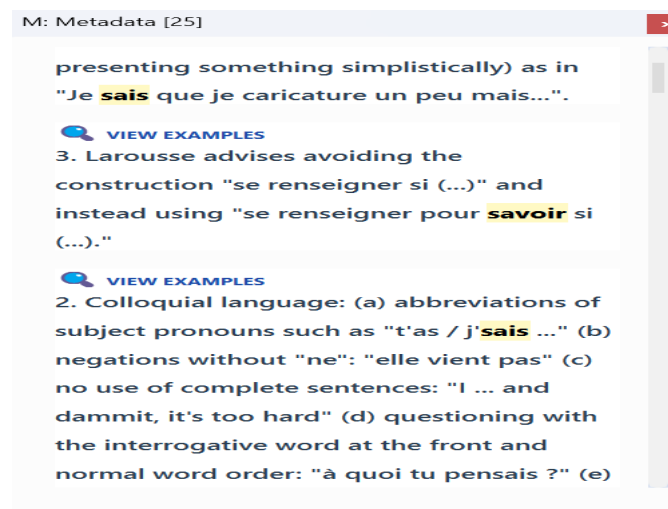
(a) The top section, Search, allows the user to search the Lexogoth corpus. The user can type one or more terms or patterns here, and the examples from the corpus (not the text) are then displayed in the separate right side panel. The search logic (the various ways to search the corpus according to distance, labels, wildcards, combinations, etc.) is exactly the same as for the text search that we have described in detail above. The user can simultaneously search the notes section of the corpus with the same search term, and the results for that query will then appear in the other side panel. Both side panels also have certain functionalities that we will briefly illustrate below.

- The *Examples* side panel:



In the photo above, the user sees the results for the search query “savoir”: the French example sentences are in blue with the translation below them. Below each pair, you will find the icon : if you click on this, the example sentence is added to the **Basket** for further editing and export. Some sentences are also provided with the icon : if you click on this, a pop-up appears with comments belonging to the text block to which this sentence belongs.

- The **Comments** side panel (ensure that “Rem.” is checked in the search field) displays all comments belonging to the searched lemma, and each comment section is also accompanied by **View examples**. If the user clicks on this, they will see the sentences that belong to these comments. In that pop-up as well, the user can select sentences and add them to the Basket:



2.6.5.2 The "Show Windows" zone

Below the search field, the user can check or uncheck the Corpus and Metadata fields. Both options are connected to the separate Examples and Comments side panels. If the interface becomes too cluttered, the user can easily hide both screens by unchecking them here.

2.6.5.3 The Basket

In this zone, you can choose to:

- View and edit the contents of the Basket, i.e., your selected sentences.
- Completely remove the contents of the Basket via Clear (*Leegmaken*).
- Add a sentence selected from the text field to the already present examples.



2.6.5.4 The export zone

In this section, the user can:

- Export the text as it appears in the text field (including layout features) to an RTF file, along with any accompanying footnotes. The RTF format is editable in any common word processor, from MS Word to LibreOffice—useful for preparing teaching materials. The option is Export text.
- Export the contents of the Basket to an easily editable text file via Export Basket.

2.6.5.5 Mode: read or edit

At the very bottom, the user can determine how they approach the text: if they only want to read, they select Read mode. In this mode, the text field is protected against unwanted, accidental changes that could, for example, disrupt the search index. Edit mode allows the user to make intentional changes to the text.

2.6.6 Using grammatical information and footnotes while reading

Once the text has been tagged and indexed, users can click on any word in the text field with their mouse. At the bottom, the Grammar tab will then appear, displaying the necessary grammatical information about the word and any phrase of which it is a part—specifically, the grammatical tense and the MWE (multi-word expression). For this, you must have the Grammar tab selected:

The screenshot shows the software interface with the following components:

- Top Navigation Bar:** File, Formatting, Workbench, Analyze, Statistics, Search in text.
- Left Sidebar:**
 - Search:** Search bar (savoir), Smart/Literal radio buttons, Distance: 0, Tag button, Search in corpus button.
 - Show panels:** Corpus, Metadata checkboxes.
 - Basket:** View contents (3), Clear, Add selected text buttons.
 - Export:** Export Text, Export Basket buttons.
- Main Area:**
 - Text:** Le Festival de Cannes se termine ce soir ! L'occasion de revenir sur les règles strictes imposées sur le tapis rouge !
Chaque année, des milliers de célébrités marchent sur le tapis rouge, mais elles ne peuvent pas le faire n'importe comment.
Depuis 2018, les téléphones sont interdits sur le tapis afin de fluidifier le défilé. Côté chaussures, les baskets et les tongs n'ont clairement pas leur place. Et pour la tenue, le carton d'invitation est sans ambiguïté : smoking obligatoire pour les hommes et, pour les femmes, les robes trop transparentes ou trop volumineuses sont interdites !
 - DATA:** Clicked word "interdits".
VERB CLUSTER:

Name tense	indicatif présent passif	Sequence	"sont interdits"
Main verb	interdire	Voice	Passive

EDIT:

Lemma	interdire (Verbe)	Word class	Verbe
Tense/Mood	participe passé	Gender	masculine
Number	plural		
 - Annotations:** Grammar tab selected (indicated by a red arrow).

If you click the Annotations tab, you will see the field containing the footnotes. Footnotes in the text are recognizable by a blue underline and a light blue background:

Depuis 2018, les téléphones sont interdits sur le tapis afin de fluidifier le défilé. Côté chaussures, les baskets et les tongs n'ont clairement pas leur place. Et pour la tenue, le carton d'invitation est sans ambiguïté : smoking obligatoire pour les hommes et, pour les femmes, les robes trop transparentes ou trop volumineuses sont interdites !

As soon as you see such an indicator in the text, you can place your cursor there and select the Annotations tab to read or edit the footnote.

Finally, please note that you can also right-click the magnifying glass icon and choose the word from the tabs that you wish to search for in the corpus.

2.6.7 For what didactic or linguistic purposes can we use the Workbench?

The Workbench can be used to tag any arbitrary French text and have all MWEs (Multi-Word Expressions) or Q-tags recognized. This provides the user with reading support and makes it easier to recognize specific patterns.

Furthermore, the reader can consult the Lexogoth corpus directly from the editor to learn the translation or usage of a word from the text.

Search queries can be executed very precisely, allowing users to easily and quickly search a complete text for specific word combinations, as well as syntactic and morphological patterns.

Through the Workbench, users can also conveniently prepare a French text for didactic use (e.g., by adding notes from the Lexogoth corpus) and subsequently export it to a universal text file.

The tool can also be used for research purposes: for instance, it is possible to study the use of tenses, word meanings, etc. Moreover, the concordance function allows you to compare word usage and meanings from the current Lexogoth corpus with those in the imported text.

Finally, the user can create a tokenized file of the imported text and use it for further analysis in other programs (such as syntactic parsers).

2.6.8 Technical explanation of the POS-tagger

- The central function within the POS-tagger is *indexText*. This function analyzes raw (unprocessed) text and transforms it into a list of tokens with grammatical tags in eight distinct stages:

1. **Tokenizer:** The text is split into separate lexical objects. In this step, the text is split at spaces, elisions, and inversion hyphens (e.g., “j’ai” becomes “je”, “ai”; “un rendez-vous” becomes “un”, “rendez-vous”; and “dis-tu” becomes “dis”, “-”, “tu”).
2. **Lemma Lookup:** The lemma (base form) of each token is looked up via a ternary search tree. The surface form (as it appears in the text) is also linked to its potential grammatical category.
3. **MWE Identification:** Multi-word expressions (such as “bien que” or “à condition que”) are identified and combined into a single token.
4. **Heuristics:** In the fourth stage, we use several heuristic rules to determine unknown words (those not in the search tree) or words with an absolutely certain tag, often using morphological characteristics.
5. **Statistical Routine:** A statistical routine is applied to the found tokens to resolve tag ambiguities. We use a **Viterbi algorithm** to calculate the

most likely pattern (i.e., the pattern with the highest score at the end of the routine).

6. **Manual Correction:** In this step, we review the results of the previous step and apply manual corrections using clear rules to rectify the assignment of certain tags.
7. **Q-Tagger:** The Q-tagger does not intervene in the assignment of a category but identifies groups of words (interesting patterns) found in the text. These often involve *locutions* (such as *faire partie de*).
8. **Tense Analysis:** After this pass, we analyze grammatical tenses in two moments via a dependency parser. This allows us to assign tense labels to verb groups (e.g., *Indicatif passé composé, actif*) and attempt to identify triggers for the subjunctive.

After tokenization, we proceed to indexing—making the tokens accessible for the search function. To ensure fast searching on large texts, we build a ternary search tree based on the literal text forms and their position (index) in the text.

- How did we treat and prepare the Lexogoth corpus for searching?

To make searching in the corpus (approximately 460,000 French tokens) fast and efficient, we performed the following:

1. **Read a POS version** of the corpus.
2. **String Pool Mechanism:** To prevent excessive memory usage, we utilize a string pool mechanism: we reduce the text of each token (lemma, tense, surface form, etc.) to a numerical ID (uint32_t) and transform a token into a set of integers [surface form, lemma, tense, etc.].
3. **Hashtable (Inverted Indexing):** We place these IDs into a hashtable. This allows for lightning-fast searching via a key (uint32_t), where the values (positions or indices in the corpus) associated with that key are returned immediately.

3. Tips for smooth operation

- **Practice materials:** Soon, you will find several databases and tests on the Lexogoth website that you can download as illustrations to help you get up to speed quickly.
- **Database size:** When compiling databases (individual sentences), both students and instructors should carefully consider their purpose. If you want a database to illustrate certain constructions or common word usages, feel free to compile one with several thousand examples; the excellent filter functions will allow you to find whatever you are looking for quickly and efficiently. However, if you are compiling a

database intended for study, it is best to limit it to a few dozen examples to ensure the material remains manageable and learnable for students.

- Importing and searching: Pay close attention to how you use the search options during import: each search option serves a specific type of query. The same applies to the search options (6) within the database editor under the search menu.
- File management: To edit databases (extension .lexdb), test results (extension .lexresult), and tests (extension .lexresult), you should always use the Toolbox. To prevent file corruption, we have encoded our native file formats, ensuring that a user cannot accidentally corrupt a database.
- The Lexogoth philosophy: Users should not lose sight of the original Lexogoth philosophy: learning words and word combinations in context, both in a narrow context (the sentence) and a broader context (blocks of text). This coherence risks being lost if the user only limits themselves to studying individual items from a database. We recommend using database items for focused study and to examine specific constructions in detail. Once the detailed work is complete, the user can combine the relevant items and notes—which form a very important part of the Lexogoth corpus—into contextual blocks or modules.

4. Technical data

The Lexogoth Toolbox is a part of Lexogoth and was added to the program in January 2026.

- Version 1 (January 2026)
- Version 2 (February, 2026)
- Version 3 (March, 2026)
- Version 4a (May, internal 2026)
- Version 4 (July, 2026): see main manual for details on upgrades

© Lexogoth, Frederik Meul
January 2026