

Lexogoth



International Version: French - English

Manual

version 4.0, July 2026

Author: Frederik Meul

1. Introduction

1.1 What is Lexogoth and what is the purpose of the program?

1.2 Who is Lexogoth intended for?

1.3 Background on the program's design

1.4 An example: *la confiance*

2. How does the program work?

2.1 Searching the lexical database

2.2 Grammar, vocabulary, and language functions

2.3 Conjugating verbs

2.4 The Search menu

2.5 The Help menu

2.6 The File menu

2.7 Clearing the screen (*refresh icon*)

2.8 Lexical Toolbox^{NEW}

3. How to use the program as a student or teacher

4. Methodology of Lexogoth

4.1 Origin of Lexogoth and primary objective

4.2 Composition of the corpus: on the basis of which criteria were words and structures selected?

4.2.1 Theoretical basis for Lexogoth

4.2.2 Selection and structure of the corpus

4.2.2.1 An empirically based approach

4.2.2.2 Structure of the corpus

4.2.2.3 Structure of the *Entries* in Lexogoth

4.2.3 Language registers, geographical variants, and language varieties

4.3. How does Lexogoth differ from other language apps and websites?

4.4. What is the best way to study vocabulary, and how can Lexogoth help with this?

4.5. An ongoing project

4.6. Bibliography

5. FAQ

6. Reformed spelling

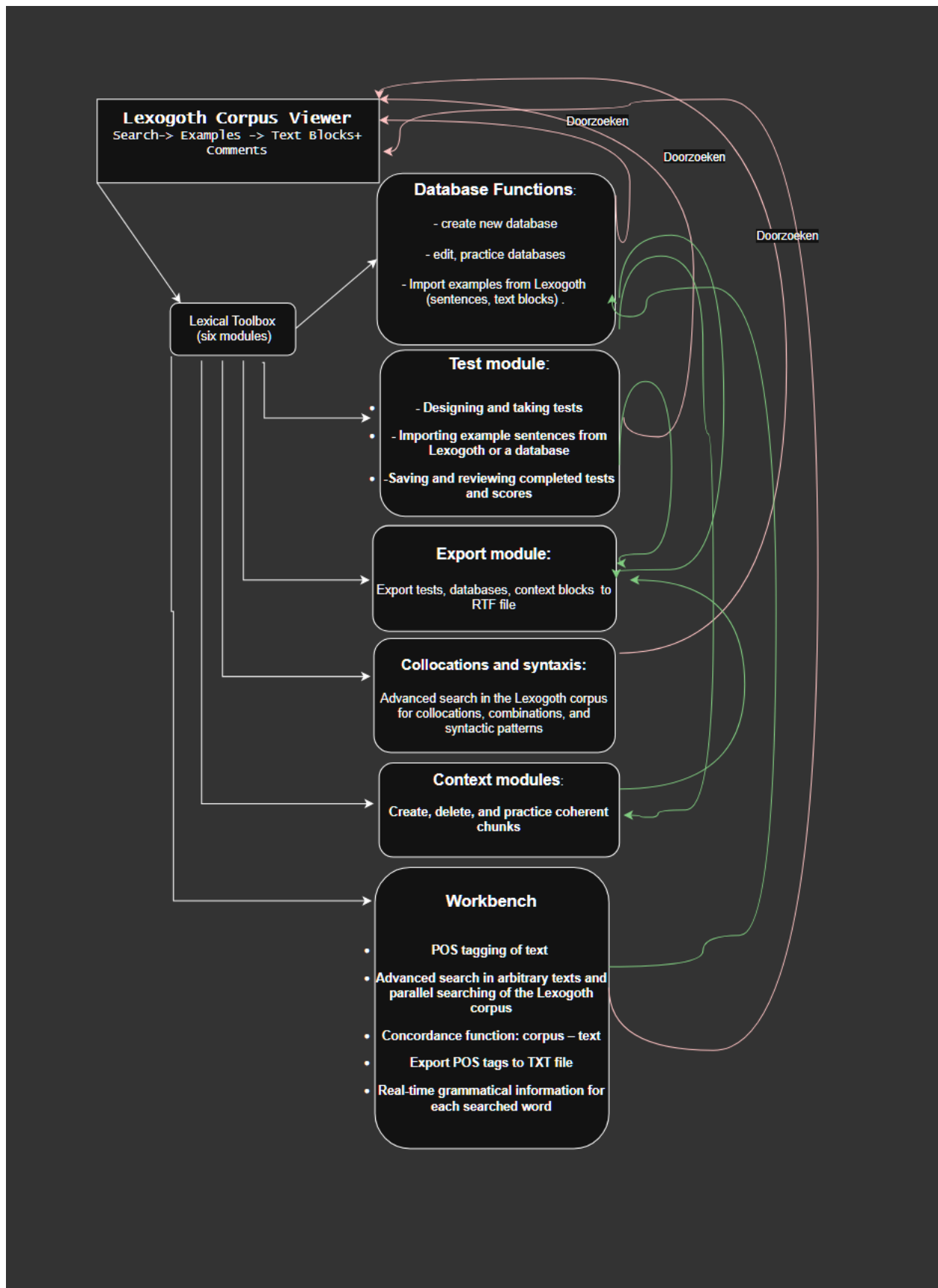
7. Varieties of French: language levels, Belgicisms, and Anglicisms

8. To-Do in Future Updates

9. Versions and Update History

10. Technical Information and Copyright

0. Workflow in Lexogoth



1. Introduction

1.1 What is Lexogoth and what is the purpose of the program?

The technological evolution within Flemish (Dutch-speaking part of Belgium) education, accelerated by the *Flemish Digital Transformation in Education* (called *Digisprong* in Dutch), has yet to fully live up to its pedagogical potential. Due to a lack of a cohesive vision behind its implementation, secondary education is often left relying on digital textbooks, glorified PDFs, and cumbersome cloud systems for automated testing. There is a glaring shortage of learning tools that truly leverage computer technology to create valuable new resources that were previously unattainable.

Lexogoth aims to bridge this gap. Rather than using technology merely to repackage analog materials in a digital format, Lexogoth creates a genuine digital didactic ecosystem. It empowers students and teachers to modernize and streamline the learning process, pushing the technical and pedagogical boundaries of traditional, analog language instruction. In other words, it does not just provide digital support for existing methods; it offers an alternative didactic model for vocabulary acquisition rooted in context, corpus data, and authentic usage.

Lexogoth is a Windows 11 desktop application designed for students and teachers to acquire, practice, and teach advanced French vocabulary, specifically targeting CEFR levels B and C. The program offers a curated selection of frequently used French words, each illustrated with clear, contextual example sentences drawn from both spoken and written language.

Lexogoth is neither a classic reference corpus nor a traditional dictionary; it is a didactically enriched lexical knowledge base. The language data provided has not merely been collected; it has been deliberately selected, structured, and annotated with a focus on learnability, nuance, and context

sensitive usage. Meanings are not just displayed but differentiated, interpreted, and made explicit, enabling users to gain deep insight into semantic nuances, usage conditions, and pragmatic constraints.

Users should note that Lexogoth is not a translation tool. Instead, it functions as an interactive textbook and study instrument (and potentially a reference work in the future) whose ultimate goal is to enable learners to maintain, deepen, and expand their lexical knowledge in a sustainable and efficient manner. The focus is not on quick equivalents, but on understanding, context, and accurate usage.

The core of this alternative learning model is the Lexical Toolbox. It enables selection, annotation, personalization, and active processing; these are precisely the processes that are absent from traditional methods. While the corpus serves as the source of knowledge, the Toolbox acts as the user's personal workshop. It allows learners to build their own dynamic knowledge base by selecting, annotating, and organizing relevant chunks and models from the corpus. This process transforms Lexogoth from a passive reference tool into a personalized learning system that bridges the gap between discovery and independent language production.

1.2 Who is Lexogoth intended for?

Lexogoth is intended for :

Fairly advanced students can use this program, specifically the text corpus, while completing communicative exercises or to support their vocabulary studies. We pay constant attention to the language registers that are inevitably linked to the use of certain words or meanings, ensuring that the student gains a clear understanding of which meaning can be used in which context. Students should also use the program as a reader, selecting specific lemmas on a regular basis and studying the accompanying chunks through reading.

Furthermore, through the *Lexical Toolbox*, students can build their own model databases, manage their collections, and practice their acquired knowledge in a targeted manner using the integrated test and context modules. The Workbench is the latest addition; here, students can paste arbitrary French text and have it analyzed, including POS tagging and verb tense recognition. Subsequently, with just a few clicks, they can look up the words in the Lexogoth corpus to find translations, usage examples, and notes associated with those terms.

Language teachers: Based on the program, they can compile a corpus of clear examples and make it available to their students. Furthermore, they can demonstrate the use of specific words to the entire group via projection. Thanks to the Toolbox functionalities, teachers can also create customized practice modules and vocabulary lists (databases), provide these to students, and generate digital tests that align directly with the material covered in class. Finally, they can design individual remediation paths for grammar and vocabulary using the available lessons (CEFR levels A to B). Teachers and students can also conduct serious studies of French language usage through the advanced search functions and the text corpus concordancer in the Workbench.

Beginning students (CEFR level A): Many examples may be too difficult for students at the beginner level. However, the corpus also contains a significant number of simple chunks or phrases that are extremely suitable for beginners. We have included the entire curriculum for the A1 to A2 trajectory (CEFR) regarding grammar, language functions, and vocabulary in Lexogoth. These approximately 90 lessons are easily accessible via a clearly structured menu. This is very practical for quickly reviewing something during class or while completing an exercise. Beginning students can also benefit greatly from the Workbench: they can paste a text into it, have it analyzed, and then look up unknown words and

structures under the virtual guidance of the corpus searcher to better understand the text (by means of the found example sentences with their translations in a discreet side panel).

1.3 Background of the program design [see chapter 4: methodology for a more thorough treatment of the background]

In our view, words, their meanings, and their combinatory possibilities constitute the greatest challenges for a foreign language learner. For every word, a student must learn a multitude of elements that enable them to use it correctly: pronunciation and spelling, syntactic possibilities, the language register to which the word belongs, its various meanings, and how it combines with other words. All these elements make learning vocabulary in a foreign language a very challenging task indeed. Unfortunately, the complexity of this process is not reflected in the way many textbooks present new vocabulary. All too often, we observe that textbooks and methods limit themselves to creating lists of words and their translations, sometimes accompanied by a short example sentence. We believe such an approach leads to very superficial and passive lexical knowledge in students, and does not allow them to use these newly learned words correctly and fluently in communicative exercises.

To bridge this gap in foreign language education, we created Lexogoth. Through this program, we offer students who have already moved beyond the basic level the opportunity to acquire new vocabulary or deepen their existing knowledge quickly and efficiently. The program provides French students with the opportunity to look up usage examples of French words (and combinations) based on examples from everyday spoken and written language. For many words, there are already numerous examples available in Lexogoth that clearly illustrate the most frequent construction and combination possibilities. **In this way, a student can form a more complete and accurate understanding of how to use certain words and transition more**

quickly to communicative exercises. The current version contains over 45,000 clickable items, each leading to a number of sentences or sentence fragments, also known as chunks. These chunks can be expressions, fixed word groups, common (and less common) word combinations, or shorter, easily learnable example sentences that illustrate both the meanings and the usage of the word.

The idea is that French students should systematically consult the available chunks via the searched lemmas, both when doing productive exercises and when studying vocabulary. For each lemma, the user finds numerous search results that, through frequent consultation, help the language learner build a vast mental network of word combinations.

Naturally, the reader is confronted with a large amount of data here, but that is precisely the strength of Lexogoth: by repeatedly exposing the student to common, recurring constructions, a natural assimilation takes place. To accelerate this process from passive recognition to active mastery, the program features a powerful *Lexical Toolbox*. This module acts as an interactive bridge: students can select the most relevant models and chunks from the corpus and save them in personal databases. Through this active processing—the selecting, annotating, and systematic practicing of this self chosen data—word combinations are durably anchored and can be immediately and correctly deployed in communicative exercises. The other modules of the Toolbox are also aimed at the specific acquisition of contextual chunks: via the Context module, students can gradually bundle sentences or chunks that belong together (semantically, syntactically, thematically, etc.) into meaningful units that are much easier to study. Furthermore, the reader can practice their reading skills via the Workbench: users can now paste any arbitrary French text into the text field and then very easily look up structures and words in context (Lexogoth corpus examples) to grasp a text more quickly.

1.4 An example: 'la confiance'.

A concrete example to clarify how Lexogoth can be useful for students of French: A capable student who knows the individual words "confiance" and "avoir" might come up with the sentence: "Elle a de la confiance dans ses amies". There's not much to criticize here – the student has constructed a sentence using the lexical (vocabulary lists!) and syntactic knowledge at their disposal, and it may seem correct to them.

Unfortunately, such a sentence is rarely encountered in real usage, and most native speakers would find it rather unnatural. With Lexogoth, however, the student can find several authentic user examples that they can study and then apply in communicative exercises with only minimal adjustments. For the word "la confiance", for instance, we might offer the following chunks or example sentences:

- (1) T'as raison, il me faisait confiance et j'ai trop parlé, j'ai merdé.
 - (2) être digne de respect, de foi, de confiance, de mépris
 - (3) déborder d'enthousiasme, de passion, de joie, de confiance
 - (4) Elle n'a plus confiance dans l'avenir.
 - (5) Je n'ai en elle qu'une confiance limitée.
 - (6) faire confiance à l'équité d'un médiateur
 - (7) L'enquête montre que la confiance dans la politique a encore chuté depuis 2020
 - (8) Il est temps qu'elle reprenne confiance en elle.
 - (9) Il n'a plus confiance en lui.
 - (10) Elle manque de confiance en elle, cela se ressent
- (+ nog 30-tal andere voorbeelden)

However, it is not the intention to force a language learner to suddenly memorize all these examples that illustrate the correct use of "confiance"; that would be a difficult and, for many, unachievable task. The student can, however, read through the examples carefully and choose one or two models that best fit their needs. Frequent use and rereading of the same examples throughout the duration of the course will likely lead the student to gradually internalize these chunks and systematically build a large network of usable lexical knowledge. After all, these chunks are not only about the

lemma “*confiance*”, but also about, for example, the use of “*manquer de*.”

Another example where Lexogoth can be useful is with the problem of “*linking words*” (adverbs, conjunctions) used to meaningfully link sentences or parts of a text. When learning words such as “*toutefois, cependant, pourtant,*” many students will wonder where in a sentence these words can typically be placed. Due to a lack of sufficient meaningful (and correct) examples, a scrupulous student may stick to the example sentence provided in the textbook and consistently place the word in just one position, while a more adventurous student might assume that these words can be used in the same positions as in their own native language – an understandable but potentially incorrect assumption. Through *Lexogoth*, a student gains access to multiple correct examples that can be studied and that allow them to independently and easily adjust their understanding of a word's proper use.

The program also includes a complete overview of the content typically found in a French (as a foreign language) course for the full A level of the Common European Framework of Reference for Languages (CEFR, Levels A1–A2). We provide an overview of the main grammar topics, vocabulary, and language acts addressed at this level. For the components *Vocabulary* and *Language Acts*, we also have a whole series of (communicative) exercises, but these have not been included in the program. We want to emphasize once again that this second part primarily serves a supportive role, aiming to help students quickly review or learn certain parts of the material independently and efficiently.

2. How does the program work?

2.1 Searching the lexical database

The primary goal of the program is to study words in their context of use. How can we quickly find a word or expression among all these examples? There are essentially **two search methods** in the program, both of which are accessed via the search window at the top left [see photo 1]:

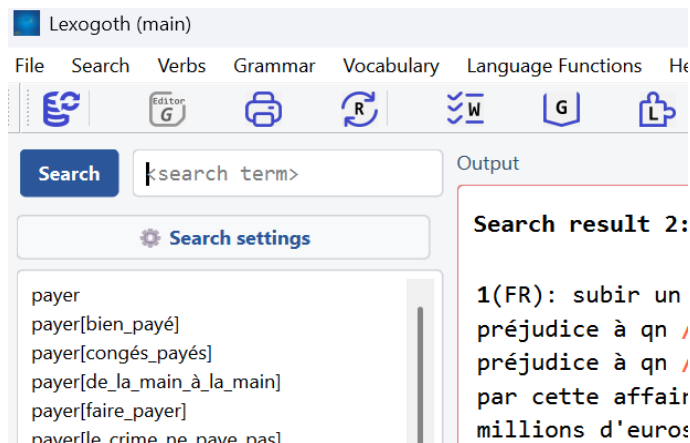
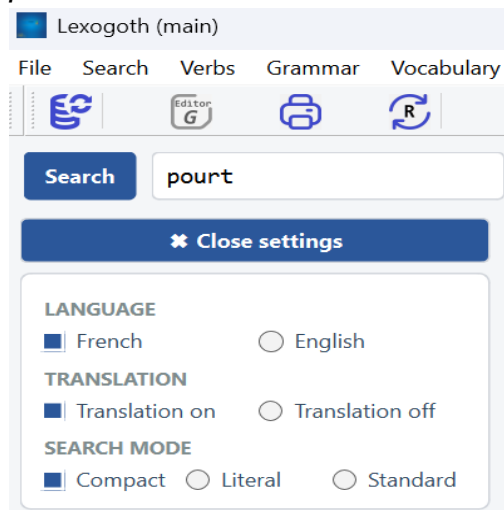


photo 1

* Method 1: directly entering one or more search terms

When the user presses Search settings, they can select various options for the search:

photo 2



First, the user can choose the language in which the search is to be performed: French or English. Additionally, one can choose whether or not

to display the translations. Finally, there is the search mode, where the user can choose between *Compact*, *Literal*, and *Standard*.

The **Standard option** allows the user to search for one or more words (in all their forms) within the various text blocks that make up the Lexogoth corpus. As soon as all the searched terms appear in a single block, the block is displayed and the searched words are highlighted against a yellow background. This is the classic **AND search query** applied to an entire text block. Some examples:

- Search term: “aller” → Display of all text blocks in which a form of “aller” occurs, such as “vais, va, irez, allé,...”.
- Search term: “voiture électrique” → Display of all text blocks in which at least one form of both words occurs, e.g., “voiture ... électrique” / “voitures ... électriques” / “électriques ... voiture”. Note that the distance between or the order of the terms is not taken into account here.

The **Literal search** option displays the blocks in which the literal search term was found. This is ideal for searching for invariable words such as “parce que, bien que, vite,...”.

The final search option is **Compact**: this option is a form of smart searching where the user can type in one or more terms, and the engine then looks up all forms of each term within a maximum distance of three words, not necessarily in sequence. For example, when searching for “dans le but de”, you will find all blocks in which “dans le but de” and “dans le seul but de” appear.

For more advanced and complex search options (wildcards, syntactic patterns, grammatical labels) that operate on isolated sentences, we refer the user to the Workbench in the Toolbox, where they can extensively query the entire corpus, as well as the notes section, using much more complex options suitable for advanced students and teachers. We chose not to integrate those search capabilities here because they run counter to the purpose of Lexogoth: exploratory reading and learning of words in context blocks, provided with extensive notes on usage. For editing, practicing, and didactically “preparing” Lexogoth material, the user should turn to the Lexogoth Workbench, which can be opened via the first button on the toolbar, just above the search button.

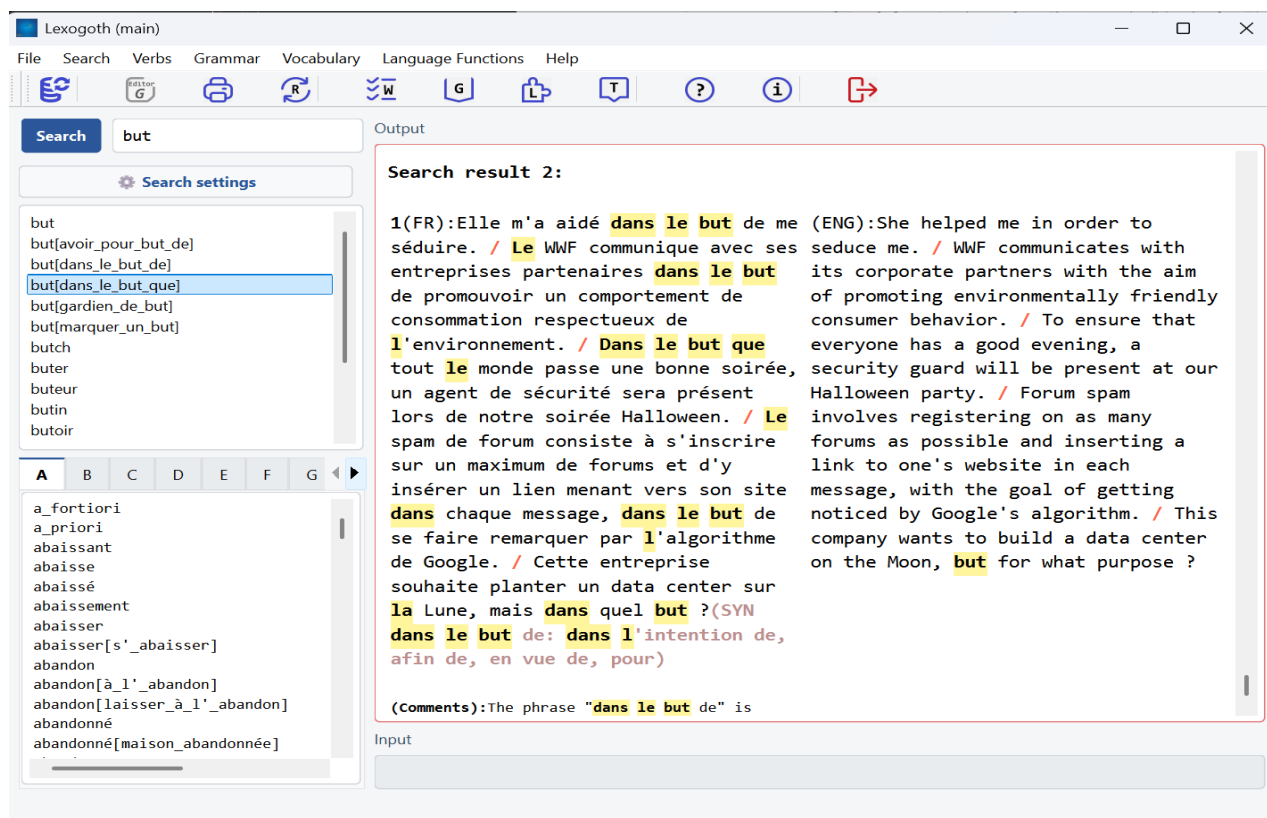
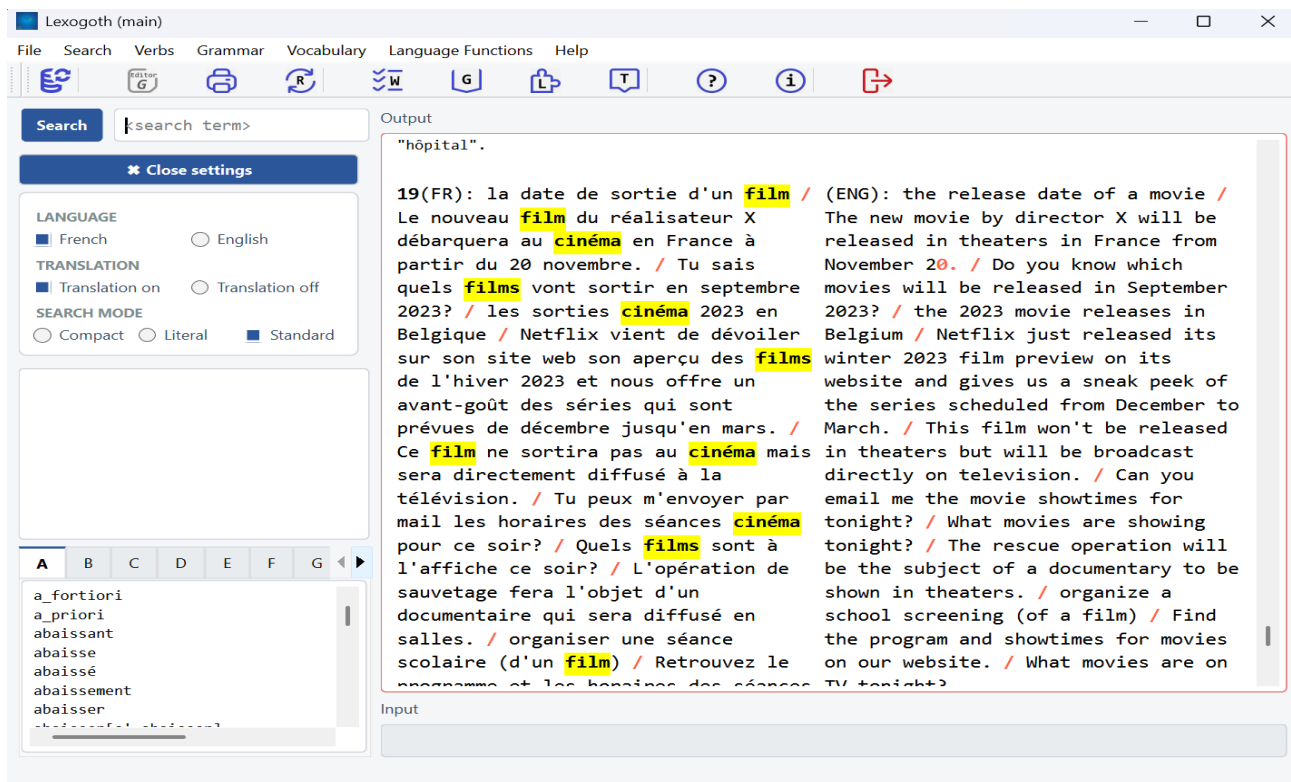


Photo 3: search for “dans le but de”

The Compact and Standard search modes are primarily intended for finding multi-word expressions or specific combinations (e.g., film and cinéma).

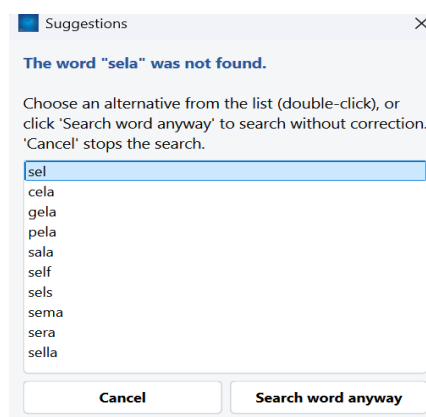
We provide two examples to illustrate their use.

Suppose we want to know how to say "to broadcast, show, etc. a film on television." We can attempt to find the combination of the words "télévision" and "film" somewhere in the corpus. You simply type both words—separated by a space!—and then press search. If the combination appears anywhere in the corpus, the result will be displayed in the main screen, allowing you to see if you have found what you were looking for. Please note that it is virtually impossible to provide every potential combination, so there is a high probability that your search query might not yield results. If that is the case, you will need to launch a broader search query or a search query with only a single search term (for example, "film") in order to analyze all the examples afterward. An example for the search query with "film" and "cinéma":

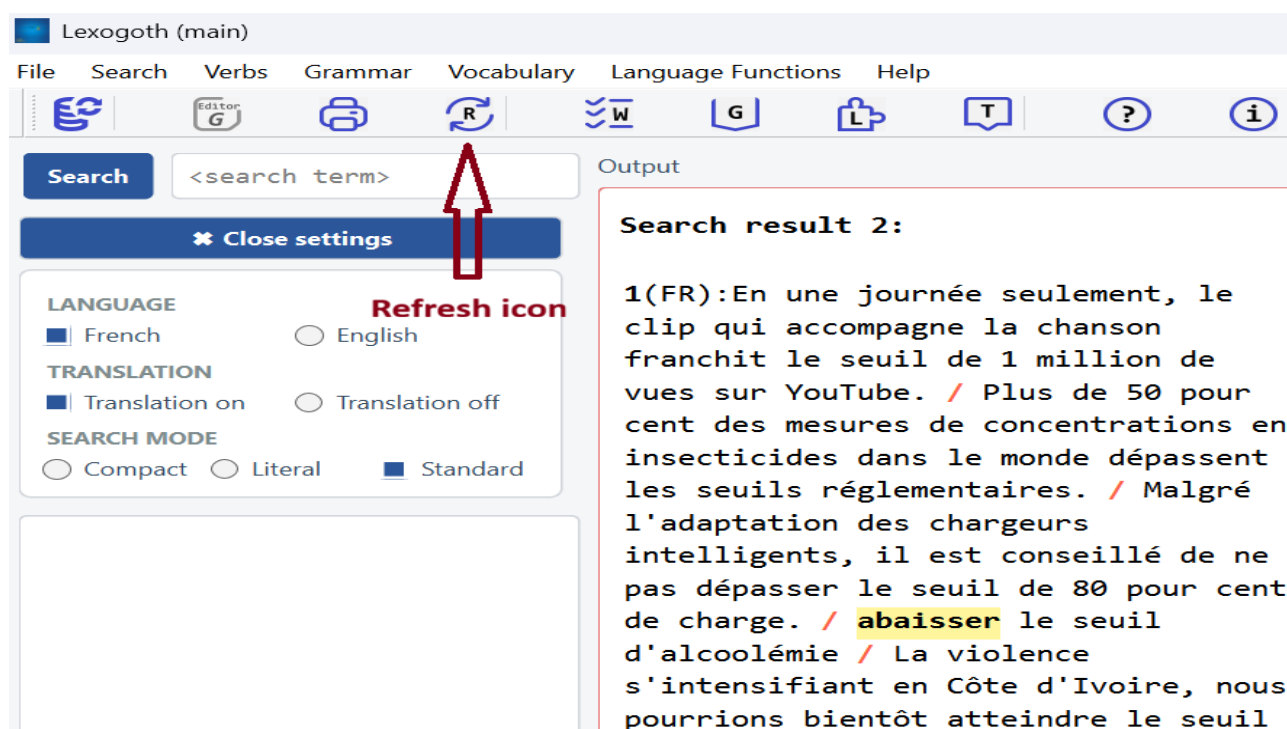


To help the user search more efficiently, suggestions (autocomplete) appear below the search window, which the user can click to go directly to the search results.

If you enter a word that is not recognized, a popup appears suggesting other words that resemble the typed word (calculated via Levenshtein distance). For example, if you enter “probleme” (without an accent), a box appears that allows you to correct your mistake by choosing a word from the suggested list. If you wish to ignore the suggestions (for example, when you were searching for a proper noun), simply click “Search for word anyway”:

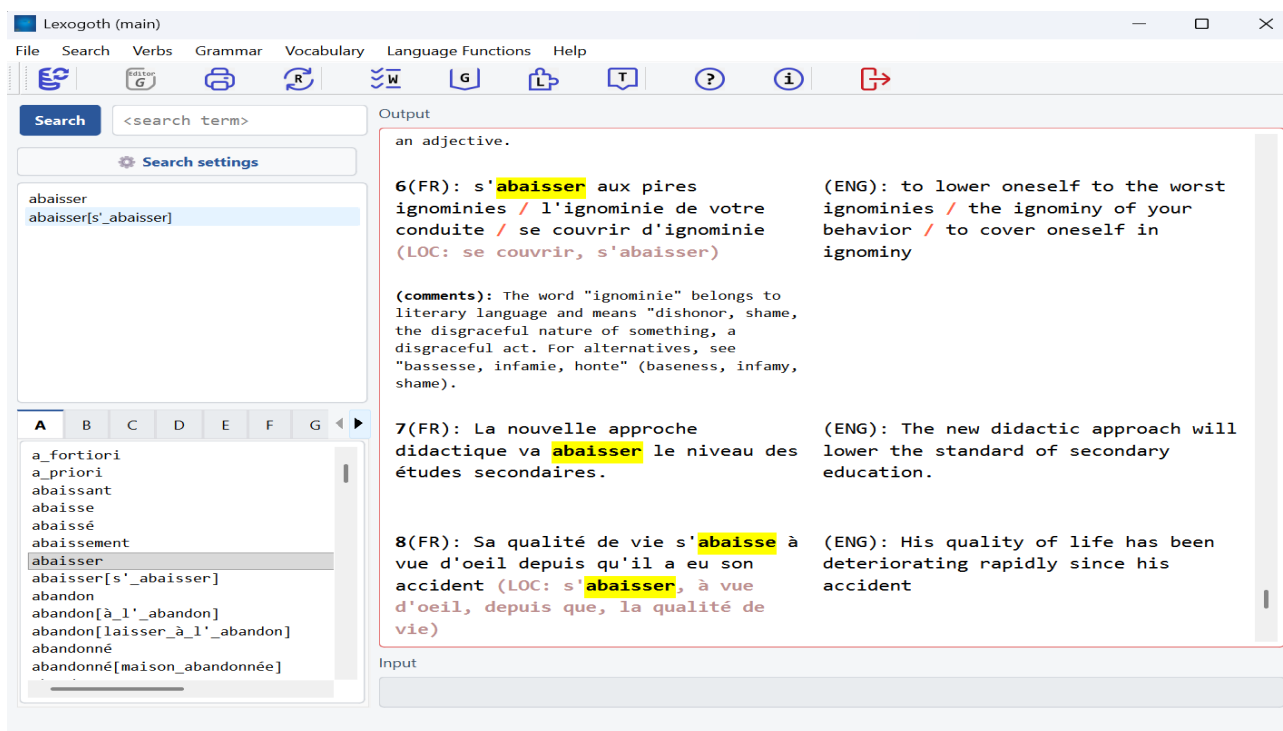


Please note: if this is not your first search, some queries can sometimes yield over 100 pages of results, and if you have performed several searches, the interface may become less responsive when resizing. Therefore, clear the screen from time to time using the refresh button (icon with the letter R in a circle) when launching a new search.



The **first result** returns (under *Search Result 1*), if the word is recognized, the base lemma of the search term and the (possible) word category. This search is performed on the word corpus, which contains a list of French words (approximately 565,000 word forms; English: approximately 35,000 forms). This is useful if, for example, you want to quickly check the "form" of a word: for "vieille," the program displays under *Result 1* that it is a feminine singular adjective derived from the lemma "vieux."

Following this, *Search Result 2* is displayed; this is the search result from the text corpus in which the forms of the searched word are shown against a yellow background:



It is also possible to use this search method to find expressions, fixed word combinations, and idioms. Suppose you want to search for the expression “*ne pas y aller par quatre chemins*”; it might be sufficient to type the terms “*aller*” and “*quatre*” to obtain the desired result. Here, too, a positive search result is not guaranteed because each entry in the corpus can consist of multiple chunks, and it is quite possible that one chunk contains “*quatre*” and another contains the verb “*aller*” without both appearing in combination! Please note that you should not use this search method to look for compound words such as “*parce que*” or “*pour que*.” For those, it is best to use method 3. [See also further down: FAQ > Searching for an exact match for multi-word expressions].

* method 2

Finally, you can also search via a classic dictionary. At the bottom left of the interface, you will find various tabs with the letters of the alphabet: select the desired language, click on the first letter of your search term, and then scroll through the list until you reach the desired

word. This method is useful if, for example, you are looking for expressions that consist of multiple words. For the word combination "faire face à," you go to the word "face" and scroll until you reach "face[faire_face_à]." Then, double-click on your desired search term so that the result appears on your screen:

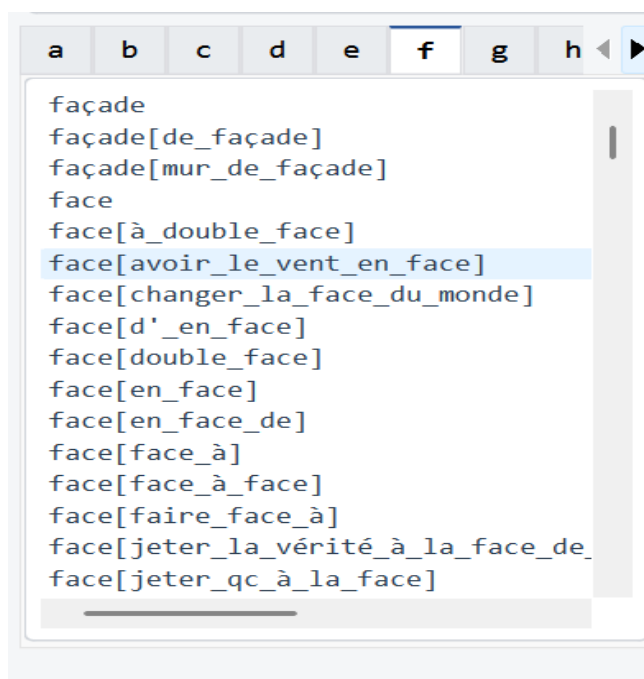


photo 5 : dictionary search

2.2 Grammar, vocabulary and Language acts

In the program, under the respective menus **Grammar**, **Vocabulary**, and **Language Functions**, you can find lessons that cover nearly all topics from the CEFR levels **A1** and **A2**. By clicking on the menu (photo 6), a submenu appears (photo 7), offering an overview of the available lessons and explanations. Select a topic from the list and double-click the item to open it. The lessons will open in the default **external PDF viewer** of the operating system on which Lexogoth is installed. If your PC does not open these files automatically, you will need to set a default program as the PDF viewer in your system settings. For most users, this will be the **Edge browser**.

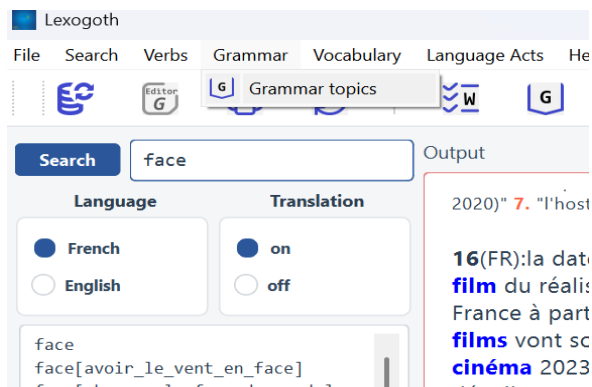


photo 6

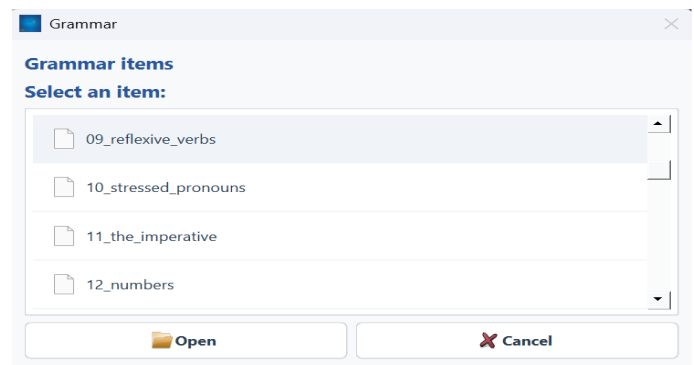


foto 7

2.3 Verbs...

A common challenge for many students is the conjugation of French verbs. By clicking on the *Word Formation* menu and the *Show verb conjugation* submenu, you can view the full conjugation of a French verb. Note that this only works if the verb is present in the database, and you can only view the active forms. The *passé simple*, *passé antérieur*, and *subjonctif imparfait* have not been added. The *passé simple* will be included in a future update. Also, pay attention to the comments provided at the bottom of the conjugation of the requested verb [photos 8 and 9].

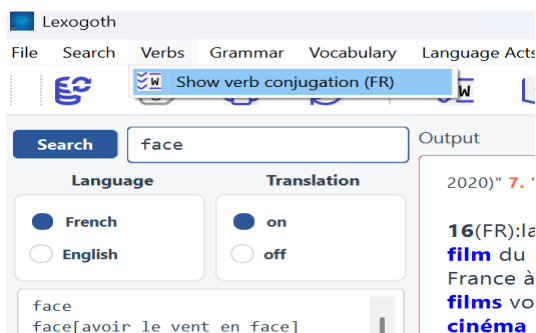


photo 8 : Verb conjugation

photo 9 : result: « acquérir » →

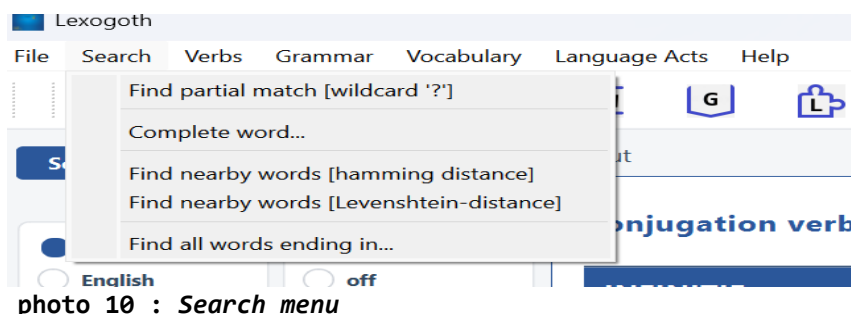
Conjugation verb:	
INFINITIF	
Infinitif :	acquérir
INDICATIF	
Indicatif présent :	j'acquiers, tu acquiers, il acquiert, nous acquérons, vous acquérez, ils acquièrent
Indicatif imparfait :	j'acquerais, tu acquerais, il acquerrait, nous acquérions, vous acquériez, ils acquerraient
Indicatif futur simple :	j'acquerrai, tu acquerras, il acquerra, nous acquérons, vous acquérez, ils acquerront
Indicatif passé composé :	ai acquis, as acquise, a acquis, avons acquises
Indicatif pl-q-pf :	avais acquis, avais acquise, avait acquis, avions acquises

If you enter a verb and it is not recognized, the search screen will disappear, and a message will appear indicating that the verb was not found. You will notice that the options *Show all noun forms* and *Show all adjective forms* have been disabled. We did this because they are actually redundant: the student can easily figure out the plural and/or feminine forms by making the masculine singular form plural and/or feminine (adding an “e” or “s”) and searching for that word. If the program

returns "unknown," it means you have likely entered an incorrect form or that this word form is not included in the word list.

2.4 The *Search* menu: advanced search options

In the search menu (photo 10), you can launch various types of search queries.



With the **Find partial match** option, you can search for words that meet the conditions using wildcards (placeholders; here the wildcard is the question mark). For example, if you search for 'compl??', you will get results such as: *complet*, *complot*, *complus*, *complut*.

The **Complete word [autocomplete]** task is self-explanatory: you type the beginning of a word, and the program searches for all words that match. For instance, if you type 'complet', the results will be: *complet*, *complet-veston*, *complets*...

The **Find all words ending in** option is also self-explanatory: it lets you find all words that end with a specific letter combination. For example, searching for 'ète' will return: *achète*, *adjupète*, *affrète*, *anachorète*, *analphabète*, *arbalète*, *arpète*, *ascète*, *athlète*, *becquète*, *bonnète*, *brevète*, *cacahouète*, *cacahuète*, *cachète*, *cafète*, *caquète*, *centripète*, *circaète*, *cliquète*, *collète*, *complète*, *compète*, *comète*, *concrète*, *coquète*, *corsète*, *crochète*, *curète*, *diabète* [+ many more].

The **Find all nearby words [Hamming distance]** is an experimental feature. It finds words of the same length that differ by a specific number of letters. By setting the distance to 1, you find words with only one letter changed. For example, searching for 'pars' with a distance of 1 will return results like *bars*, *cars*, *gars*, *mars*, *pacs*,.... This tool helps you discover similar vocabulary based on the number of letter manipulations required. Please note that this calculation only applies to words of equal

length. Using this feature allows for advanced pattern matching within the database.

The ***Find all nearby words [Levenshtein distance]*** task is an experimental feature that offers more flexibility than the *Hamming distance*. While Hamming only allows for letter substitutions, Levenshtein calculates the number of operations (including insertions, deletions, and replacements) needed to transform one word into another. Because it handles variations in word length, this logic serves as the basis for most modern spell checkers to suggest corrections for typos. For example, if you search for all words within a distance of 2 from "*parti*", you will get: *arbi, pars, parie, parée, part, paros, art, pâti, mari, pares, marri, ports, parut, barri, cari, parés, sari, paris, partes, serti, mati, port, anti, paré, pari, paras, kari, parti, tari, gari, para, péri, arts, parus, parez, pait, parts, parcs, porte, harts, mardi, parois, patri, perte, pare, paria, sorti, garni, par, hardi, paru, parr, fart, parc, hart, kart, parsi, farsi* [+ more]

2.5 The *help* menu

Here you will find the usual help topics in the submenus: *tutorial* (the document you are currently reading), *links* (references to online dictionaries and grammars), and *About*.

2.6 The menu *File*

This menu contains the option to close the program (*Close program*), print the output screen (*Print current output*).

2.7 A small but important icon: the refresh icon.

Once you have performed several searches, the results can quickly add up to the hundreds and this may slow the interface. Therefore, we recommend users to regularly clear the output screen by clicking the refresh icon [photo 11]. By doing this regularly, you can learn more comfortably, easily resize the program, and avoid accidentally printing hundreds of pages if you wish to print something.

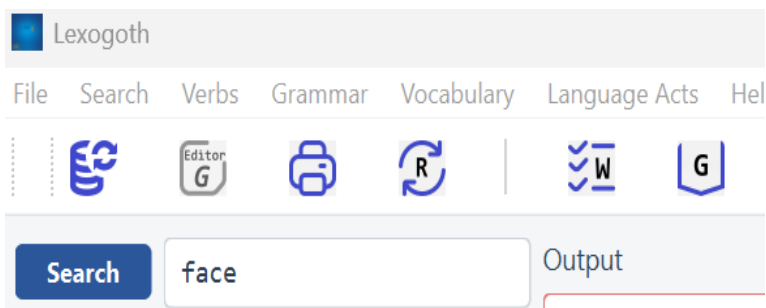
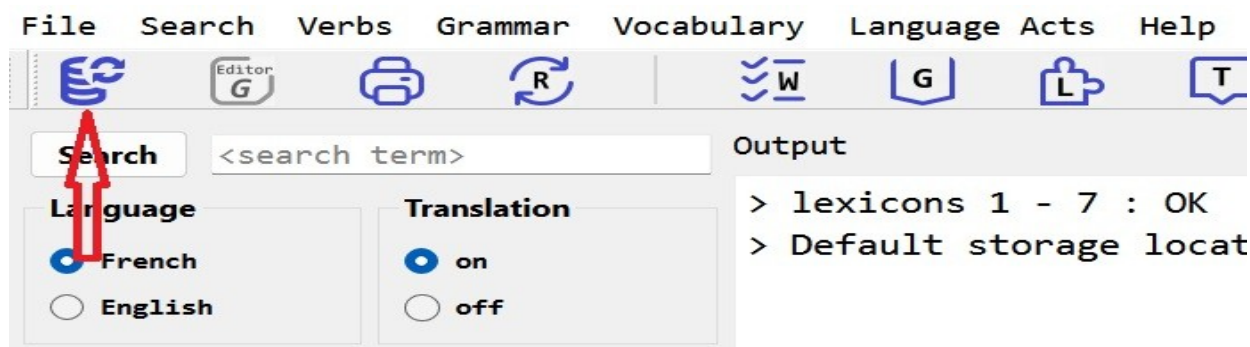


Photo 11 : (Icons from left to right: Open Lexical Toolbox, open Golgoth editor [disabled in International version], print output screen, **refresh screen (clear screen)**, conjugate verb, Vocabulary lessons)

2.8. The Lexical Toolbox



In the illustration above, clicking on the first icon will open the *Lexical Toolbox*. This module is new and is being rolled out from *Lexogoth International v3*. The *Lexical Toolbox* of *Lexogoth* is a module for students and teachers of French, designed to practice vocabulary and lexical patterns using authentic corpus data. The program is built around three main modules:

In the illustration (above), clicking the first icon will open the Lexical Toolbox. This module is new and is being rolled out starting with Lexogoth v4.5. The Lexogoth Lexical Toolbox is a module for French students and teachers, designed to practice vocabulary and lexical patterns using authentic corpus data. The program is structured around three main modules:

(1) Database functions: These allow users to create lexical databases, edit existing databases, and practice the databases by hiding the English or French translation. This also includes several important corpus tools such as n-gram analysis, advanced search options, searching for word combinations, and more.

(2) Import / export: This module forms the heart of the Toolbox and allows users to import entries directly from Lexogoth and filter them using advanced search options (AND/OR/Exact string match/Proximity Search). It also enables exporting databases, tests, and test results to an editable RTF text file.

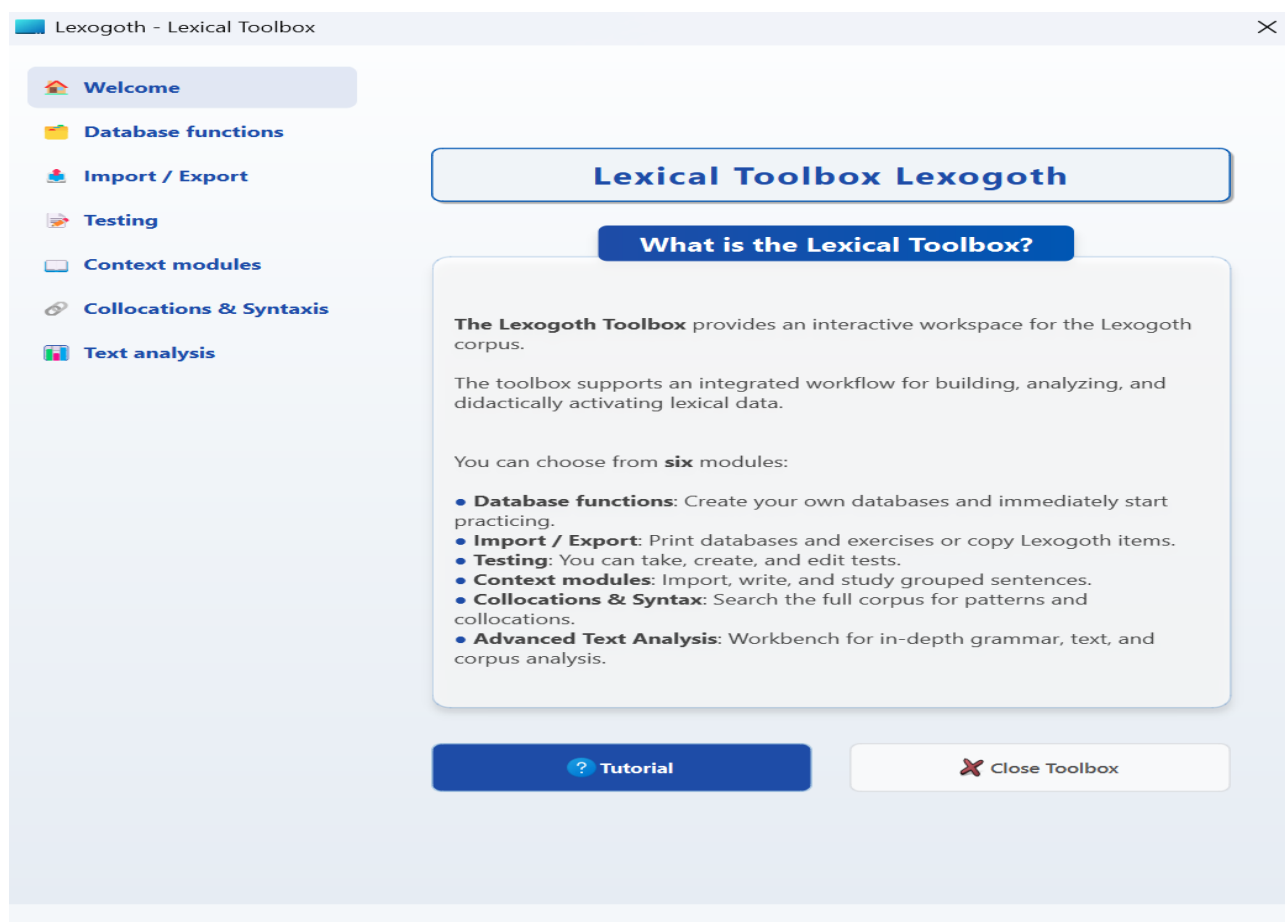
(3) Testing: Users can create and take various types of tests, including multiple choice, matching, exact fill-in-the-blank questions, and dropdown questions.

(4) Context modules: Through this module, students and teachers can compile their own learning blocks according to their own criteria (semantic, thematic, or syntactic) to more easily illustrate or practice specific lexical fields or syntactic constructions.

(5) Collocations and syntax: Here, one can perform advanced searches on the Lexogoth corpus: searching for fixed expressions, syntactic patterns, grammatical tags, searching with wildcards, searching with fixed or variable distance between words, and so on. The results found can easily be exported to databases, context modules, tests, or text files (RTF) that can be easily edited further in Word or LibreOffice.

(6) Lexogoth Workbench: This module allows users to paste arbitrary (French) texts and subsequently tag them (providing words with a grammatical label): it is an extensive tool that can be used for many things: linguistic analysis, didactic support during reading. We have provided a direct link to the Lexogoth corpus via convenient side panels so that words can be looked up in the Lexogoth corpus for illustration with a single mouse click. There are also numerous export possibilities. A complete description can be found in the separate Lexogoth Toolbox tutorial.

An illustration of the startup screen:



Consult our separate manual on the website for detailed Toolbox instructions [<https://english.lexogoth.be>]

3. How to use the program as a student or teacher?

- As a student, you can use Lexogoth in the following ways:

(1) If you need to prepare a speaking or writing exercise on a specific topic, you can launch some searches using key words related to the topic you are working on. With the returned sentences, you can then begin to build your texts. The results provide you with both linguistic and content-related assistance. Linguistically, this means you get an overview of the possible constructions of the words you can use; the thematic support is found in the example sentences, which expand the lexical field you're working within. For example, if you want to work on "réchauffement climatique" (climate change), you can get quite far with

results from search terms like "réchauffement (climatique)", "pollution", "température", "santé", and "problème".

(2) You can also use the program to study vocabulary: you can look up and study sentences that illustrate the correct use of difficult words (such as connectors and verbs). This approach to studying will lead to a much deeper understanding of the language.

(3) Additionally, Lexogoth can also be used as a kind of reading and learning book. The language learner who wants to expand their knowledge of French over a longer period can dedicate a few moments each day to exploring a few words. Since the same words will appear regularly, they will gradually internalize them through reading. The reader will, over time, develop more affinity with the French language, and their expression capabilities will grow increasingly expansive.

(4) You can independently refresh your knowledge of certain language topics covered in the ERK A-level (and part of B-level) through the numerous lessons on 'grammar, vocabulary, and language functions' available in the program.

(5) The program also includes a module that shows the conjugation of most French verbs; if in doubt, you can always call up the conjugation of the desired verb with the click of a button.

(6) While working on short tasks in class, you can also use the Lexogoth editor with the autocomplete function for French; using this can help you write difficult words correctly.

(7) Build personalized databases with the *Toolbox*: You can create your own lexical databases by importing French text blocks and their English translations directly from Lexogoth.

(8) Take interactive tests: You can take tests featuring various question types, such as multiple choice, exact completion, matching, and dropdown menus.

(9) Refine your learning material: Use search filters during import to target specific words or grammatical patterns (like the subjunctive) to ensure your study databases are relevant.

(10) Export for offline study: You can convert your databases, tests, and test results into reader-friendly RTF text files to edit or print them using standard word processing programs.

As a teacher, you can use Lexogoth in the following ways:

- (1) You can use the program in class as a tool to explain specific lexical problems to students; for example, you can easily demonstrate the difference between "*prêter*" and "*emprunter*" with the numerous example sentences available. The program is also very useful for exposing the more complex constructions of some words through various examples.
- (2) The example sentences in the program can also be used when preparing tests or lessons. For instance, the teacher can prepare a list of key words and representative examples for their students ahead of speaking or writing exercises, making it easier for them to complete the exercises successfully (see the example "*réchauffement*" above).
- (3) Lexogoth can also serve as a quick and easy tool for looking up language problems (synonyms, use of certain complex constructions, useful examples, meanings, translations, language registers, etc.) that require a fast answer during class.
- (4) The teacher can also use *Lexogoth* to design remedial learning paths for students who have fallen behind. The main grammar and vocabulary topics, along with numerous language functions for the A-level of the European Framework, are included in the program. If a student has difficulties with any of these topics, the teacher can guide them in the right direction by having them review the relevant "grammar, vocabulary, or language function" chapters via the program.
- (5) Create specialized study databases: Teachers can build custom lexical databases by importing French text blocks and their English translations directly from Lexogoth.
- (6) Curate content with advanced search: Use the specialized "AND," "OR," and "Exact match" search filters to quickly extract specific grammatical patterns (like the subjunctive) or thematic vocabulary from large imported text blocks.
- (7) Design various test formats: You can create and edit tests using multiple question types, including multiple choice, exact completion, matching, and dropdown menus.
- (8) Manage and refine materials: Easily merge different databases, remove duplicate entries, and rank items by length to organize content from simple to complex.
- (9) Generate printable materials: Export databases, tests, and results into editable RTF (Rich Text Format) files to create handouts or printed exams.

- (10) Monitor and review student performance: Load saved test results (*.lexresult files) to view a detailed table overview of scores, given answers, and correct answers for each student.
- (11) Annotating reading texts for students so that they can process and understand unseen texts more independently. (*Workbench* from the Toolbox).
- (12) Studying specific structures, semantic and syntactic patterns, or evolutions within modern French via the various search options.

4. Methodology of Lexogoth

In this chapter, we explain how Lexogoth was developed, the linguistic principles that guide it, and how we compiled a rich corpus of authentic French language. This text offers language teachers, foreign-language education specialists, and curious users a deeper insight into the vision and approach behind Lexogoth.

4.1. The Origin and Primary Goal of Lexogoth

Lexogoth arose out of a frustration familiar to every language learner who has ever had to study vocabulary: the struggle to find the right word combinations and the realization that memorizing word lists alone is not enough to participate fluently in communicative tasks.

We believe this difficult transition is largely due to poor study methods and insufficient practice with vocabulary. Most foreign language students begin their learning journey by memorizing vocabulary lists, which allows them to quickly and efficiently acquire a small base lexicon. While this approach is useful in the early stages, many textbooks—particularly in French—continue to rely on word lists year after year. This persistent focus eventually handicaps students: when they are required to perform meaningful communicative tasks later on, they soon discover that a mix of grammar rules and isolated words is simply not enough to use the language effectively.

Learning through lists prevents students from forming connections between the words and expressions they encounter. In other words, learners end up with a collection of disconnected lexical items, without the semantic network needed to tie them together. This lack of internal structure leads to shallow memorization, which in turn makes students forget new vocabulary much faster than if it had been embedded in a network of associations. Word lists also tend to ignore the inherent complexity of words. A word is more than just a definition or a translation: to use it properly, one needs to know which complements it takes, which meanings are frequent and which are not, what register it belongs to, in which contexts it can be used, as well as its synonyms, antonyms, and derivatives.

To make matters worse, vocabulary practice in many French textbooks is still limited to crosswords, fill-in-the-blanks, or short completion

exercises—far from enough to build solid lexical competence. Because of this rigid reliance on word lists, combined with a lack of rich, meaningful input that could illustrate both the meaning and use of new words, students progress in the foreign language much more slowly than they should. For further reading on vocabulary acquisition and these issues, see: Nation (2001), Thornbury (2002), Laufer (2022), Bentolila (2010).

In our view, vocabulary should always be taught in context, with students exposed to numerous examples that clearly demonstrate how words are used and that help them build a robust semantic network more quickly. It is from this perspective that we developed **Lexogoth**: our goal is to provide students of French with a representative corpus of high-frequency words, expressions, and collocations in varied contexts. This enables learners to remember and practice these lexical items more effectively and more efficiently. Below, we explain the methodology we used and how the corpus was developed.

4.2. Composition of the corpus: based on which criteria were words and structures selected?

2.1 Theoretical Basis for Lexogoth

The development of Lexogoth is closely linked to two major trends in foreign language education that gained serious momentum in the 1990s:

- The Lexical Approach by Lewis (1993, *The Lexical approach: The state of ELT and a way forward*)
- Corpus linguistics: the use of large text corpora for linguistic research or language applications such as machine translation.
- *Data-driven Learning* (often abbreviated as *DDL*): a didactic approach in which students take on the role of a researcher and discover patterns and regularities themselves within large amounts of **authentic** language data.

Let's start with Lewis's Lexical Approach. One of the key principles of this approach is that, in the context of foreign language learning, language should no longer be seen as a collection of sentences built from grammatical structures into which lexical items are inserted at will (the traditional grammar-based approach to language teaching). Instead, language should be understood as a collection of ready-made expressions, compound words, and regular word combinations that are stored in the learner's memory and can be used directly and quickly, with only minimal adjustments depending on the communicative context.

Put simply, according to Lewis, it is not grammar but rather the network of lexical units that forms the foundation of successful foreign language teaching. These lexical units, often called *chunks* in this context, can be defined as all individual and compound words, frequent word combinations, expressions, and collocations. Lewis's idea is that language teachers should start from lexical units as the basis for instruction and provide students with rich and varied input through authentic texts and conversations, supported intensively by concordancers and dictionaries.

The Lexical Approach with its focus on chunks offers several advantages; here are a few:

- Using chunks improves foreign language learners' writing skills (Albaqami, 2022, *The Role of Lexical Chunks in Promoting English Writing Competence among Foreign Language Learners in Saudi Arabia*).
- Native speakers also rely heavily on pre-formed chunks or phrases for communication (Pawley and Syder, 1983, *Two puzzles for linguistic theory: Nativelike selection and nativelike fluency*).
- There is evidence that students who learn chunks demonstrate greater fluency and better speaking skills, while those who focus on individual words struggle to produce sentences spontaneously (Boers, Eyckmans, Kappel, Stengers, 2006, *Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test*).

Of course, Lewis's proposal to base language teaching predominantly on lexical knowledge has also drawn criticism. Here are some key points:

- On what basis should word combinations and expressions be selected, and how can a course be designed around them? And assuming one selects the most frequent words, how can an entire year's curriculum be organized around all the authentic material related to those combinations? (Alison Wray, 2002, *Formulaic Language and the Lexicon*)
- It is relatively easy and manageable to learn large word lists and store them in long-term memory, but what about word combinations where students are expected to study dozens of common collocations for a single word? This seems practically and humanly impossible, as it would require an excessive amount of time and effort from students (Alison Wray, 2002).

- What about grammar? Grammar, especially syntax, determines how we construct correct sentences. Why should it be sidelined when it provides one of the most important mechanisms for creating new sentences? (Rod Ellis, 2003, *Task-Based Language Learning and Teaching*)

A second important development that is fundamental to Lexogoth is that of corpus linguistics, a linguistic discipline that emerged in the 1960s and that systematically collects and examines large amounts of language data (usually via computer) to identify trends. Some well-known (English-language) corpora that can be found online include The Corpus of Contemporary American English (COCA) and The British National Corpus (BNC). Due to the increasing accessibility of computers in the 1990s, more people, including language didacticians, gained access to various corpora that were previously only available within research institutions. In this way, people were able to start thinking extensively about a potential deployment of text corpora in foreign language education, which is where the idea of data-driven learning originated (Johns, 1990). From the 2000s onwards, various initiatives and attempts followed to bring authentic text corpora to students and integrate them into the classroom through the creation of, among other things, electronic dictionaries supplemented with corpus material, concordancers to study word usage/syntactic constructions, and syllabi with expressions and word combinations based on their frequency.

Finally, a third important development upon which Lexogoth relies is the didactic approach of *data-driven Learning (DDL)*, a subset of *CALL (Computer Assisted Language Learning)*. In their paper (2024), which provides a state-of-the-art overview of *DDL*, Boulton and Pascual Pérez-Paredes describe *DDL* as a form of inductive learning in which the student acts as a language researcher. Instead of passively consuming rules, the learner analyzes authentic corpus data to independently "discover" patterns and register differences. Such an approach to teaching foreign languages seemed very promising, and many researchers in the scientific literature regarding the *DDL* approach were extremely positive. Yet, we see today that the *DDL* approach has not become a standard component of modern foreign language education. The reasons why teachers (and students) have not (yet) embraced this study approach can be found in *Exploring the barriers to data-driven Learning in the classroom: a systematic qualitative synthesis* (2025, Applied Corpus Linguistics 5, Amelie Xiaohan Sun, Atsushi Mizumoto), where the authors provide an in-depth analysis of the literature surrounding *DDL*, specifically regarding the obstacles hindering its widespread acceptance.

The obstacles are divided into two groups: internal barriers inherent to the method and external barriers.

According to the authors, the main internal obstacles and challenges are:

- **Time:** Searching for and filtering results (concordances) is a very time-consuming task, especially when the corpus yields hundreds of examples per query.
- **Interface:** Most concordance programs are quite complex and designed for linguists rather than students who want to learn a foreign language.
- **Other source materials:** Many students resort to more familiar sources for research, such as dictionaries and search engines. Added to this is the recent rise of AIs, which allow students to get direct answers to their questions.
- **A mismatch between learning needs and the complexity of corpus tools:** Sometimes, especially at the beginning of their learning journey, students have only simple goals: looking up a simple word, consulting a translation, etc. For such purposes, complex corpus tools are entirely unnecessary.

The authors also cite external factors that explain the low success rate of DDL:

- Numerous students have only limited language proficiency and ICT skills.
- Some learners have very poor autonomy, a low frustration threshold (for uncertainty), and insufficient analytical capabilities.
- Most teachers have never learned how to use corpus tools and are under great time pressure to complete their curriculum.

The authors rightly emphasize that successful implementation of DDL requires much more than just a few "technical" solutions. With the development of Lexogoth, we have attempted to formulate an answer to these objections.

By using Lexogoth, students no longer have to deal with the drawbacks that inevitably come with the individual use of such tools:

- Concordancers return search results with thousands of example sentences mixing all meanings of the searched lemma. If a language learner has to sift through all these examples and distinguish the

different uses based on dictionary definitions, learning will be very slow. In Lexogoth, we have done this heavy work for the students, clearly distinguishing meanings and providing representative examples.

- Definitions in online dictionaries are often too abstract for learners and offer little guarantee of correct usage due to insufficient clear examples. In Lexogoth, we provide clear examples and explain many definitions in English to ensure learners fully understand.
- Concordancers are sometimes specialized programs that are difficult to learn to use. Lexogoth's user interface is simple and intuitive.
- The seemingly infinite lists of examples offered by corpora are not suitable as direct study material: too many examples, no information on register, no suggestions for synonyms or paraphrases, no sociocultural explanations for certain phenomena (such as *verlan*), etc.

Finally, we want to emphasize that Lexogoth is not a language course or method like Lewis originally envisioned with the Lexical Approach, but a tool to help students learn frequent word combinations and words in context more quickly, providing authentic material with which they can systematically and independently practice and reinforce their lexical knowledge.

How does Lexogoth address these problems ?

- The Lexogoth corpus (468,000 French tokens, 349,000 tokens of didactic notes; May 2026) is a didactically processed corpus rather than raw concordance output:

For each lemma:

- meanings are clearly distinguished,
- the most representative sentences and chunks are selected,
- socio-cultural, syntactic, semantic, and pragmatic notes are added. Students no longer need to sift through hundreds of examples to understand the correct usage. The time required is significantly reduced. This didactic processing (clustering, combined with notes) also ensures that students do not have to discover the analysis of

the various meaning aspects of a word themselves; the heavy lifting has already been done for them.

As a result, students no longer need to sift through hundreds of examples to understand the correct usage. The time required is significantly reduced. This didactic processing (clustering, combined with notes) also ensures that students do not have to discover the analysis of the various meaning aspects of a word themselves; the heavy lifting has already been done for them.

- a very accessible interface

Lexogoth hides the complexity typically associated with corpus tools:

- advanced options are available but not mandatory,
- default settings avoid decision fatigue,
- contextual help panels guide the user,
- the Toolbox shows full sentences instead of limited snippets (concordances).

This approach significantly reduces the cognitive load.

- integration of corpus data into all modules

Because students too often resort to dictionaries, translation tools, or AIs, we have integrated the ability to search the corpus into as many modules as possible, the most important being the test module, the database module, and the Workbench. This integration is such that the need to quickly consult external sources is significantly reduced. **Online dictionaries often offer overly abstract definitions without sufficient context or usage instructions.** In Lexogoth, meanings are clarified with concrete examples and, where necessary, accompanied by English explanations. A tool like Lexogoth also avoids the problems associated with AI bots: these bots generally provide good answers, but they are essentially “black box” answers: you do not know where they come from and they are often not reproducible, which is an essential component for a learning tool.

- for beginners and advanced learners

For beginner learners, the corpus also contains vocabulary for the A1-A2 levels of the CEFR. Complex examples can easily be skipped. This keeps the search function usable for all learning levels.

External obstacles: what can Lexogoth solve and what can it not?

For beginner learners, the corpus also contains vocabulary for the A1-A2 levels of the CEFR. Complex examples can easily be skipped. This keeps the search function usable for all learning levels.

External obstacles: what can Lexogoth solve and what can it not?

- **Teacher training:** teachers will need additional training to become familiar with DDL principles and integrate them into their lessons; Lexogoth lowers the threshold, but it certainly does not replace training.
- **Mindset change:** DDL requires tolerance for uncertainty and variation in language use. Lexogoth supports this through rich, realistic examples and by making variation visible. However, students will still need to dare to step away from a too conservative attitude towards language rules. This is also something that is best taught in teacher training.
- **ICT skills:** the installation (via MSIX) and interface have been kept deliberately simple so that even users with limited digital skills can get started with it.
- **Different learning styles:** Lexogoth offers multiple processing formats, such as databases, context blocks, annotated texts, and export functions, so that students and teachers can choose what suits them best.

These three developments, *the Lexical Approach*, *corpus linguistics*, and *data-driven Learning*, have served as the methodological and theoretical foundation upon which we have drawn inspiration.

We start from the premise that, once a student has reached a certain level of proficiency, language acquisition, and specifically the acquisition of a foreign language, should revolve primarily around learning authentic and context-bound lexical units (frequent word

combinations, phrases, fixed expressions, etc., which we have carefully curated using tools from corpus linguistics). Lexogoth combines the many advantages of a concordancer, a dictionary, and a textbook into a single program. The software also provides useful advice regarding the use of words and word combinations (frequency, clarification of meanings, grammatical issues, register, pronunciation, etc.).

By combining authenticity, didactic mediation, user-friendliness, and integrated functionality, Lexogoth aligns very closely with what Bax (2003) calls "integrative CALL." In this stage of CALL, technology is no longer viewed as an external tool, but as a natural part of the learning process: inconspicuous and fully embedded in daily pedagogical practice. Lexogoth fits within this vision because the corpus, the Toolbox, and the Workbench do not function as separate modules, but as integrated components of one coherent didactic ecosystem. The technological complexity remains largely "invisible": the interface is simple, the heavy corpus logic is hidden, and the user works directly with meaningful, authentic language data without major technical barriers. Furthermore, within Lexogoth, corpus data is enriched with relevant grammatical and semantic information, making collocations, syntactic patterns, and meaning variants immediately visible. These "interventions" address the classic limitations of traditional corpus-based tools and make it possible to process authentic input in a didactically responsible manner. In this sense, Lexogoth represents a "guided" form of data-driven learning within an integrated CALL environment: it combines the power of corpus linguistics with a user-friendly interface and a workflow that aligns with the reality of the classroom.

Finally, we would like to emphasize that Lexogoth is not a teaching method, as Lewis originally envisioned with the Lexical Approach, but a tool designed to help students learn frequent word combinations and words in context more rapidly, and to provide them with a wealth of authentic material with which they can systematically and autonomously practice and review their lexical knowledge.

4.2.2 Selection and Construction of the Corpus

4.2.2.1 An Empirically Based Approach

As we mentioned above, explanatory (or translating) dictionaries are not ideal tools for foreign language learners because, in our view, too many essential meaning components of lemmas—necessary for proper usage—are not made explicit enough. For example, this concerns the selection of complements by a verb or the actual usage in everyday language, rather than the “virtual” usage found in dictionaries or textbooks. For instance, Larousse gives the following definitions for the verbs *frôler* and *friser* (Le Robert en ligne):

(a) [frôler] 1. Toucher légèrement en glissant, en passant. 2. Passer très près de, en touchant presque. → [raser](#). *La voiture a frôlé le trottoir. au figuré: Frôler le ridicule.*

(b) [friser] ... 2. Passer au ras de, effleurer. (voir: frôler, raser) 3. Approcher de très près. Elle frise la soixantaine. Cela frise le ridicule.

Based on the definitions above, someone wanting to use *frôler* and *friser* might conclude that the two words can largely be considered synonyms. However, after checking the corpora we use, we found that *friser* and *frôler* can indeed be considered synonyms in a figurative sense (with abstract complements), but not in their literal meaning (i.e., followed by concrete complements), as we could hardly find usage examples where *friser* is used in that sense. Although the definitions are largely identical, it would be incorrect to assume their actual usage is the same. For example, you will rarely, if ever, come across sentences like *un motard alcoolisé qui frisait dangereusement les passants dans le centre-ville de...*

To prevent language learners from encountering such problems and to help them learn word meanings more smoothly, we have therefore chosen to always verify the words, structures, and their different meanings included in our corpus based on frequency. We did this using two tools:

- (a) Consultation of the *French Web 2023* corpus from SketchEngine (covering all language registers)
- (b) Consultation of the RTBF corpus (French-speaking Belgian television)

and news channel; standard language), and the *Le Monde* corpus (French newspaper; standard and polished language)

Afterwards, we consulted the following reference works to verify whether the lemma or word combination was included and to map out the main meanings of the lemma:

- Le Robert (dictionnaire en ligne)
- Larousse (dictionnaire en ligne)
- Dictionnaire de l'Académie française (en ligne)

After this step, we searched for useful examples illustrating the meanings currently used in French (regardless of register). We aimed to gather as broad a range of relevant examples as possible and consulted the following resources:

- The *French Web 2023* corpus from SketchEngine
- The website www.languefrancaise.net, specifically the general dictionary and *Bob, l'autre trésor de la langue* for colloquial or informal language
- French-speaking Belgian media: RTBF (public broadcaster), Belgian newspapers: *Le Soir*, *La Libre Belgique*, *La Dernière Heure*; French newspapers: *Le Monde*, *Charlie Hebdo*, *Le Figaro*, *Le Parisien*, *Libération*; Swiss newspapers: *24 heures*, *Le Matin*, *La Tribune de Genève*; French-language media from Madagascar, Canada, Congo (DRC), Haiti, Morocco, and Algeria
- Official government websites from Belgium and France, and to a lesser extent Canada and Switzerland; websites of companies from the Francophone world; websites of international organizations such as the European Union, United Nations, NATO, IMF, etc.
- Numerous internet forums (especially for familiar and spoken language); some of the main sources include JeuxVideo.com (mainly youth language), Forum Santé – Doctissimo, French-speaking Reddit channels, AlloCiné (film reviews), SensCritique (film reviews). This also includes numerous blogs ranging from hair styling and

technology to international politics (covering all language registers)

- French subtitles of recent films and TV series; we ensured that subtitles were created by native French speakers rather than AI-generated to guarantee authenticity
- Websites offering language advice: *Banque de dépannage Linguistique* (Canada), the *Dire et ne pas dire* section of the Académie française, notes from Larousse (online dictionary), *Projet Voltaire*, the WordReference.com forum, and the contrastive Dutch-French grammar resource *Grammar Warrior*

Next, we selected examples that appeared frequently in various sources and included them partially (as chunks) or fully (as sentences) in Lexogoth. We made sure the examples clearly illustrated the existing (abstract) meanings and, in the case of verbs, that a clear selection of the expected complements was used. To illustrate this, let's return to the example of *frôler*:

... sombrer dans la paranoïa communiste / Cela expliquera peut-être mieux sa méfiance qui frôle la paranoïa. / frôler le ridicule; l'illégalité, le burn-out, l'overdose / frôler la mort, la faillite / une balle lui a frôlé le cœur / Une grosse pierre, lancée par un inconnu, lui a frôlé le visage et a percuté la voiture derrière lui. / Le taux de participation aux élections frôle les 60 pour cent. / La police a interpellé un motard alcoolisé qui frôlait dangereusement les passants dans le centre-ville de Lille. ...

Because not everything can be conveyed through examples alone, we regularly add extensive notes concerning register, compatibility with specific complements, synonyms, etc., to better guide learners toward correct usage. For "frôler" and "friser," for instance, we added the following information:

-
1. In the figurative sense of "narrowly escaping" or "bordering on," you can also use "friser" as a synonym: "*friser la mort, la faillite, le burn-out, les 50 pour cent,...*"
 2. In the literal meaning (grazing past something, lightly touching), where the object of "frôler" (as in "*la balle lui a frôlé le nez*") is something concrete, the equivalence between "frôler" and "friser" seems to

no longer hold, even though Larousse and the Académie still consider both words to be synonyms in this sense. In this meaning, the verb "**friser**" is barely used, if at all; therefore, you will not encounter phrases like: "...*frisait les passants...*" or "...*frisé le visage...*"

If a meaning did not appear in dictionaries (e.g., recent Anglicisms, vulgar language, neologisms) but had a decent frequency, we still chose to include the word in our corpus. Conversely, if a word or meaning was recognized by one or more dictionaries but appeared infrequently or not at all in our consulted corpora, we chose not to include it.

Sometimes, we also found it necessary to compress some dictionary definitions because the differences between them were too subtle to be clearly distinguished.

4.2.2.2 Structure of the Corpus

The Lexogoth corpus (July 2025) consists of two parts: first, we included the vocabulary from the main lexical fields for levels A1 and A2 of the Common European Framework of Reference for Languages (CEFR) as short, useful chunks (phrases or word combinations) specifically aimed at functional language skills for beginners. This part accounts for roughly 15 percent of the total corpus. We then expanded this vocabulary by adding the main and most common meanings and uses. This first level represents about 15 percent of the total number of examples.

The second and larger part of the corpus consists of an in-depth exploration of existing lemmas and their usage in familiar and everyday spoken language, covering roughly the B level of the CEFR. An example for the verb *descendre*:

(a) monter l'**escalier** / **descendre** l'**escalier** / croiser qn dans l'**escalier** / **descendre**, monter l'**escalier** quatre à quatre (b) L'incendie était terrifiant, nous confie une locataire. C'est la fumée qui m'a réveillée, j'ai réussi à **descendre** par l'**escalier** avec mon fils. / L'ascenseur ne fonctionne pas depuis ce matin; j'ai dû **descendre** par l'**escalier**. / **descendre** par l'**escalier**, par l'ascenseur / **descendre** de l'**escalier**, de l'échelle, d'un arbre ... / **descendre** du métro, du train, de la voiture, de son vélo.... / Où est ton frère? Il est dans sa chambre. Dis-lui de **descendre** immédiatement pour dire bonjour à mamie / monter dans sa chambre (c) prendre l'**escalier** pour monter à son appartement au 4e étage / Je suis monté en prenant l'**escalier**. / Pour monter sur une échelle, il faut monter une marche à la fois.

(REMARKS) 1. "**Descendre l'escalier**" or simply "**descendre**" without a complement means "to come down the stairs." This is the most frequent construction. Note that this is a transitive use of "descendre" and requires the auxiliary verb "**avoir**": "*J'ai descendu l'escalier.*" If "descendre" does not have a direct object (COD), you should use "**être**": "*Je suis descendu. / Elle est descendue par l'escalier. /...*" The same rule applies to "monter." 2. "**Descendre de...**" is primarily used to mean "to come down from (an object) that you are sitting, standing, or climbing on," as in "*descendre d'un arbre, d'une échelle,*" or "to get out of/off a vehicle," as in "*descendre de son vélo, du train,...*" Here, the auxiliary verb is also "**être**". 3. "**Descendre par l'escalier,...**": When specifying how (by which way) someone came down, it is often to contrast it with another method: "*on descend par l'escalier ou par l'ascenseur?*" 4. "**Une mamie**" or "**mamy**" mean "grandmother," while "**papy**" or "**papi**" refer to a "grandfather." For the plural, you can add an "s" to the words: "*des mamies, mamys, papis, papys.*" 5. "**Quatre à quatre**" means "head over heels, very quickly," and is often used with verbs of movement like "*grimper, monter, descendre, franchir,...*"

For this second part, we also systematically add meanings and uses that we consider to be at an **advanced level (CEFR level C)**. For these entries, we regularly include notes to help learners master the correct usage and study new structures. Below is another example for the verb "descendre":

*...Je risque de me faire **descendre en flèche** sur ce forum, mais est-ce que je suis le seul à trouver ce film débile ? / Cherche pas, à ce prix-là, c'est du vol et ton appli va se faire **descendre en flèche**. / La presse argentine n'a pas attendu longtemps pour **descendre en flammes** la formation dirigée par Jorge Sampaoli.*

(REMARKS):... 2. "**En flèche**" can be used as an adjective phrase ("*une montée en flèche*") or as an adverbial phrase ("*monter en flèche*"). 3. "**En flèche**" is most often used with verbs of "rising" or "increasing," such as "*monter, augmenter, grimper,*" and means "at rocket speed." It is also used quite frequently with verbs in the semantic cluster of "starting," such as "*partir, démarrer,*" etc. Finally, the combination is also used with verbs that express a decrease, such as "*baisser, descendre,*" although to a much lesser extent. See the next note, however. 4. "**Descendre en flèche**" is sometimes used incorrectly in place of the informal expression "**descendre en flammes**," which means "to heavily criticize, tear someone down, or rip them apart." For example, the text above would have been better written as: "*...de me faire descendre en flammes...*" and "*...va se faire descendre en flammes.*"

In the second part, we also systematically include words and structures that frequently appear in the media, internet forums, and government websites. By being exposed to authentic and current material, language learners can understand media and literature on important social issues and also gain the tools to express themselves about these topics.

4.2.2.3 Structure of the Entries in Lexogoth

The way the entries in Lexogoth were designed lies at the heart of our innovative approach: we combine the functionality of a dictionary, a concordancer, and a textbook. It was a deliberate choice to bring together lexical precision, authentic usage, and a didactic dimension, enabling students of French to acquire frequent lexical combinations (chunks) more efficiently and rapidly, through intelligent repetition or targeted consultation. For this reason, users will encounter different types of entries in the corpus:

- (a) **Entries with an exhaustive treatment of a single lemma:** these entries are closest to the dictionary function and illustrate the main uses and meanings of a word, often organized in a clear and structured way.
- (b) **Entries organized around a specific theme,** such as fraud, climate change, or electric mobility. These entries provide a solid basis for building and expanding a lexical network.
- (c) **Entries focusing on a language problem** at the grammatical, lexical, or pragmatic level, which students of French often encounter when using the language.
- (d) **Simple entries** that list words or focus on a single expression or combination. These entries are primarily designed to support the development of a basic vocabulary.

Thanks to this varied structure of the entries and of the corpus as a whole, users can acquire new vocabulary and frequent lexical patterns in an effective and flexible way.

4.2.3 Language Registers, Geographic Variants, and Language Varieties

In Lexogoth, we strive to highlight the diversity of the French language as broadly as possible. For instance, we pay a lot of attention to the geographical variants of French, specifically Belgicisms. Belgicisms are French words, structures, etc., that are specific to the French used in Belgium but have little spread elsewhere. We dedicate less attention to

other varieties of French, such as those from Canada, Switzerland, Luxembourg, Congo, and Algeria.

Anglicisms –that is, Frenchified English (especially American) words and structures– also receive extensive coverage in our work, and wherever possible, we provide synonyms from the standard language.

Although the focus in Lexogoth is primarily on the standard language, we devote a lot of space to informal language–the language that French speakers use in their everyday interactions. We always mention in our notes whether an expression belongs to informal language, and here too we try to provide synonyms or alternative formulations from the standard language. We also address vulgar expressions and the uniquely French phenomenon of *verlan*, always providing standard-language equivalents where possible.

4.3. How Lexogoth Stands Out from Other Language Apps and Websites

Many language apps and websites focus on the fast and easy acquisition of elementary vocabulary through a classic approach based on isolated words, often accompanied by an image, a photo, or a translation. While this method can be useful for beginners or when learning specialized jargon, these apps often fall short in developing the active, productive vocabulary needed for correct and fluent language use. Lexogoth was designed to bridge that gap, distinguishing itself on the following key points.

Authentic Context: Many popular language apps (Duolingo, Drops, Babbel, etc.) base their vocabulary instruction on repeating isolated words or very simple sentences that often feel unnatural (e.g., "I can fly"). The problem is that these apps fail to show learners how a word is used in its natural environment—that is, in real written and/or spoken language. They don't teach which complements often appear with a word, what its register is, which socio-cultural elements might limit its use, or which of its meanings are frequent and which are not. Lexogoth operates on the principle that a word is only truly known when a user understands how it can be used in context. We aim to teach this by presenting users with exclusively authentic text fragments, with a special focus on frequent collocations and idiomatic expressions.

- Relevant Words and Structures: Where other apps often rely on thematic lists (like "living," "food and drink," "getting around," "daily life") for students to memorize, Lexogoth consists of a carefully, manually compiled corpus. This corpus contains words and word combinations that are current, relevant, and used with a relatively high frequency. To

ensure this, we consistently consulted recent text corpora when including new entries.

-Deep Anchoring of Lexical Knowledge: Traditional language apps try to keep their users engaged by offering short sessions (e.g., Drops), presenting vocabulary learning as a game (with streaks, badges, achievements, etc.) through simple fill-in-the-blanks, translation, or pronunciation exercises. Lexogoth takes a different approach: the app contains no exercises in the classic sense. The goal is for users to deeply anchor patterns and correct word combinations through repeated and attentive reading of the examples. This process, which closely resembles how native speakers use their language, will ensure that the acquired knowledge can be easily retrieved and used during communicative situations. In other words, students will less often need to rely on the time-consuming and cognitively demanding word-by-word sentence construction and will be able to react more quickly and correctly.

- Attention to Diversity in Language: Registers, Variants, and Sociocultural Context: Most apps pay little to no attention to the different language registers tied to the use of words or specific meanings. A lack of attention to these nuances leads to inappropriate or unnatural language use. The same applies to the geographical variants of French and the socio-cultural components linked to certain word usage, as is the case with *verlan* language. Lexogoth provides the necessary attention to this diversity, so users not only learn the correct meanings of a word but can also use it in an appropriate context.

-No AI-Generated Material, but Empirically Validated Data: Numerous language learning apps and methods use generative AI to quickly and cheaply create learning material. This is a smart approach, but the quality of the generated sentences is not always reliable, as they often create unnatural sentences or use word combinations that are rare. In Lexogoth, the examples are not AI-generated but are carefully and manually selected from relevant corpora of recent native speaker usage. We must add that we did use AI tools like DeepL, ChatGPT, and Gemini to speed up the translation process into English. However, the translations were manually checked and updated where necessary to guarantee a high level of accuracy for the user.

4.4 What is the Best Way to Study Vocabulary, and How Can Lexogoth Help?

Vocabulary acquisition is a difficult and sometimes painful process for language learners. It's not enough to simply know the translation of a word; you also have to learn a host of other things, such as the type of complement that follows a verb, the correct preposition after a noun, the language register to which a meaning belongs, the pronunciation, and more.

We believe Lexogoth can help ease and accelerate this learning process . For many common French words and combinations, the corpus provides multiple, clear usage contexts that can save the user a lot of time (searching in dictionaries, deciphering difficult definitions, etc.) and give them the means to reliably and systematically learn and review vocabulary.

However, users must not fall into the trap of complacency, as effective vocabulary acquisition must also be paired with quality instruction and highly targeted productive exercises (speaking, interaction, writing). But that, of course, is the task of the teacher...

4.5 An Ongoing Project

Currently (July 2025), Lexogoth contains a little over 40,000 clickable items. We are actively working to fill the gaps that are (most certainly) still present in the work. This process will undoubtedly take several years, and we plan to release a major update every three months.

4.6. Bibliography

Albaqami, M. (2022). The Role of Lexical Chunks in Promoting English Writing Competence among Foreign Language Learners in Saudi Arabia. *Arab World English Journal*, 13(2), 1-15.

Bax, S. (2003). CALL—past, present and future. *System*, 31(1), 13-28.

Boers, F., Eyckmans, J., Kappel, J., & Stengers, H. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Journal of Second Language Writing*, 15(4), 227-246.

Bentolila, A. (2010). Le vocabulaire: pour dire et lire. In A. Bentolila, *Parle à ceux que tu n'aimes pas: Le défi de Babel* (pp. 157-174). Odile Jacob.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34.

Laufer, B. (2022). Formulaic Sequences and Second Language Learning. In P. Szudarski & S. Barclay (Eds.), *Vocabulary Theory, Patterning and*

Teaching: Studies in Honour of Norbert Schmitt (pp. 95-108). Multilingual Matters & Channel View Publications.

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. LTP.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). Longman.

Thornbury, S. (2002). *How to Teach Vocabulary*. Pearson Education.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

5. FAQ

- Can I also learn English with this program as a French speaker?

It is possible to improve your English using this program, but we do not recommend it. Lexogoth was designed primarily for non-native speakers learning French, not the other way around. This choice means that almost all the focus is on French lexical issues, and difficulties related to the English language (for example, for French speakers) are not addressed, or are only marginally addressed. English plays a purely supportive role here, used to help students acquire a large part of the French lexicon as efficiently and easily as possible; this is why the English lists are significantly less comprehensive than the French ones. We have provided far fewer synonyms for English lemmas and did not systematically differentiate between the different labels for the same words (for example, “drink” as a noun and as a verb). Additionally, English and French belong to different language families, and a balanced treatment of both languages in one program would require a different approach.

- What is the difference between Lexogoth and a classic (online) dictionary or a concordancer?

A classic dictionary provides all the meanings that a word has or has had in a language. The major advantage of a dictionary is its completeness, but unfortunately, dictionaries often provide only a very limited number

of usage examples, and for some meanings, they don't offer any examples at all. This is a significant disadvantage for language learners, who cannot make much progress without sufficient input and will find it more difficult to assimilate new words.

A concordancer is a computer program that runs search queries on a large corpus of texts (e.g., all newspaper articles from the last 10 years) and returns the requested search results. If you enter, for example, "par exemple" into a concordancer, the program searches for all example sentences in which "*par exemple*" appears. This can be a very useful tool, but it has some disadvantages: the number of results can become enormous (depending on the size of the corpus, even into the thousands), and the quality of the results is far from ideal. Moreover, the different uses of a lemma are not always ideally illustrated; it is often up to the user to guess what the differences may be between the various examples, if any.

In *Lexogoth*, we try to combine both approaches to the lexicon as follows: (1) we only provide the relevant (more common) meanings of a lemma and omit outdated ones, (2) we offer clear, relatively short, and studyable examples of correct usage of a lemma, construction, or expression, (3) we provide extra explanations for lemmas that may pose difficulties for language learners; for example, we pay attention to the language registers used or to the grammatical "peculiarities" of certain words or constructions, so that the learner has sufficient knowledge to use the lemma appropriately, (4) we also pay attention to the possible combinations of the lemma with other words to give the language learner the necessary expressive freedom.

A combination of these features places the work neither with dictionaries nor with concordancers; the user can view and use *Lexogoth* as a source of information (examples of correct usage) or as a study book (selecting examples to study). This is far from all, because in some cases, we provide a bit more explanation about the content of the example sentences.

- *Why should I use this program instead of just using Deepl (Linguee), ChatGPT, or another AI tool? One click, and I get all the results.*

The aforementioned tools are very useful for someone who already has a good (or no) grasp of the language and wants to quickly create or translate documents without wasting too much time. Furthermore, students may benefit from these tools to save time or get help in drafting works. However, these tools are not helpful for people who are learning a language and need to actively use it. By systematically relying on these tools for exercises during the learning process, students disable their

own critical thinking about the language and fail to build knowledge of the language. A solution is provided, and it is adopted uncritically. Moreover, these programs do not explicitly differentiate between various language registers and do not provide explanations for the choices of certain constructions or words. Through *Lexogoth*, however, the language learner is required to critically assess all the examples presented and select the one that best matches their query. This approach offers numerous advantages: for instance, a learner will often read the same examples repeatedly, thereby expanding their lexical knowledge (significantly) just by reading. Furthermore, through the numerous comments, they also acquire the essential knowledge about how words are used, which is an absolute prerequisite to becoming a competent user of a foreign language.

- *What is the role of Lexogoth in FLE?*

Lexogoth acts as a bridge between corpus linguistics and the FLE classroom. Unlike traditional textbooks that often use artificial examples, Lexogoth provides students with *authentic, context-rich language data*.

Its role in FLE can be summarized in three key functions:

- **Data-Driven Learning:** It encourages students to act as "language detectives." Instead of just memorizing rules, they discover how French is actually used in various registers and contexts by exploring over 40,000 authentic entries.
- **Focus on Lexical Chunks:** Lexogoth emphasizes "prefabs" and collocations, that is the way words naturally stick together. This is crucial for FLE learners to achieve native-like fluency and move beyond literal translations.
- **Scaffolded Autonomy:** With the *Lexical Toolbox*, Lexogoth moves from a reference tool to an active learning environment. It allows teachers to create differentiated materials and enables students to build their own personalized learning paths based on the vocabulary they encounter.

- *Why does Lexogoth not use Spaced Repetition to promote retention?*

Spaced Repetition is a memorization technique where learning content is repeated at increasing time intervals. Traditionally, this is achieved through flashcards: a student creates cards with questions and answers, sorts them from easy to difficult, and reviews them over time. After each session, the student assesses which cards were easy and which were more

challenging. Easy cards are then moved to the back or shown less frequently, while difficult ones appear more often. This creates a system where difficult items return more quickly and certain items return more slowly. This form of spaced repetition has proven to be an efficient way to memorize data and is ubiquitous in many language learning programs, as almost all platforms offer some variant of this technique.

Lexogoth also supports repeated exposure and active processing of linguistic data, yet it does not use a Spaced Repetition System as found in numerous flashcard applications. The reasons why we avoid time intervals, scoring, or automatic position calculations are as follows:

(a) Lexogoth views language learning, and specifically the acquisition of usable lexical knowledge, as a contextual process. We believe it is pointless to study word lists, whether through traditional rote memorization or spaced repetition, as we have observed from our own experience as teachers that such an approach to vocabulary acquisition often leads to very superficial knowledge. Students learn nothing about the concrete use of words, potential combinations like collocations, or the appropriate register. These are all parameters that must be taken into account when studying. A student might master a thousand flashcards, but that will not truly enable them to produce correct sentences in the target language. It is, in a sense, a form of false knowledge. Lexogoth, by contrast, is entirely focused on active language proficiency by offering chunks embedded in rich contexts and accompanied by notes.

(b) In many language apps, the Spaced Repetition algorithm has become an end in itself. We observe that streaks, intervals, and gamification have become far more important than the learning process itself. Reducing language learning to a simple memory routine is a step we do not wish to take. This simplification does a serious injustice to the richness of language and does not treat the language learner with respect. Language is much more than sorting cards from difficult to easy.

All that aside, what does Lexogoth do instead? In Lexogoth, the user can and must also repeat and study, but in a way that aligns more closely with authentic language acquisition:

(a) Practicing databases: The user chooses which items to repeat and determines the principles by which a group of items is compiled, such as syntactic, semantic, or thematic criteria.

(b) Context modules: In Lexogoth, you do not practice with isolated words but with coherent chunks, which makes repetition automatically more meaningful.

(c) Lexogoth features a vast corpus of authentic chunks and sentences: the user can consult this corpus at any time and reread specific sentences when looking something up. Frequent exposure to the same patterns leads to a natural form of spaced repetition.

(d) In Lexogoth, repetition does not arise from an algorithm but from meaningful interaction with authentic language data. By regularly reading, searching, selecting, and reorganizing, the user naturally creates a sustainable and genuine form of repetition that is much closer to real language acquisition than an externally imposed interval schedule.

- If I click on "fait" in the word list, I also see all search results for the verb "faire." Why is that the case, and how can I avoid it?

The Lexogoth program is based on approximately 2,500 pages of text, combining English and French. When the program starts, it processes the texts and the program's 'tokenizer-lemmatizer' assigns all possible labels to words. For example, the word 'fait' will be linked to both 'substantif' (*un fait*) and 'verbe' (*faire*). For this reason, when you click on 'fait' in the word list, you will see results for both 'faire' and 'un fait' (and also 'en fait', 'au fait', etc.). For words that are not as frequent, it's not a problem to read a few examples to find the desired form, but for very frequent words, this can be an issue. To help the reader, we have manually added labels so that they can more easily find the desired category of a word with multiple labels. For example, for 'fait'/'faire', we added the following labels: 'fait[substantif], faire[verbe], fait[en_fait],...'. This way, the reader can more easily navigate through all the examples. Be sure to use the visual search methods to take advantage of this feature. **If the user navigates to the Toolbox module, they gain access to more advanced search routines, allowing for queries via wildcards, distance, combinations, POS-tags, and more.**

- Why don't I find all compound words like 'pomme de terre' or 'film d'horreur' right away?

In French, compound words are most often created with prepositions, meaning that one 'word' often consists of several words separated by a space. For a text program, this is, of course, a nightmare. Because how do we distinguish between a true word (like *pomme de terre*) and a random combination of words? It's not always clear where to draw the line to recognize a letter sequence (or sequence of sounds) as a word: in English, we say *potato* (a single sequence of letters), while in French,

more than three 'words' are needed to refer to the same object (the potato), namely *pomme de terre*. Now, suppose we recognize *pomme de terre* as a single word (because, for example, speakers of the language no longer analyze it as the sum of three different words—namely, an apple from the ground—but as a whole). What do we do with phrases like *film d'action* / *film d'aventures*... where things are less clear? Should we also assign them the status of a 'word' because they have the same form as the *aardappel* example, even though they can still be analyzed based on their components? Intuitions about this will likely vary, and that is also reflected in the search results. For example, we had to decide which word groups formed a unit and which did not: in the example above, we chose to treat *film d'action* as three separate words, while we consider *pomme de terre* as one word. **If the user navigates to the Toolbox module, they gain access to more advanced search routines, allowing for queries via wildcards, distance, combinations, POS-tags, and more.**

- How do I search exactly for expressions that consist of multiple words? / What is the meaning of the "underscore" in words? And why is "parce que" listed as "parce_que" in the word list, but "pour que" as "pour[pour_que]"?

The underscore was added internally to indicate that the words belong together and form a unit. So, "parce_que" is considered as one word, while the computer treats "parce que" as two separate words. However, if you look in the dictionary, you will find the sequence "parce_que" without problems, but not "pour_que". Why is that? To avoid making the dictionary cluttered and hard to read by systematically adding every fixed expression as a single string (e.g., "à_bride_abattue", along with all expressions starting with the preposition "à"), we chose to add compound expressions and words based on the meaningful words that make up the expression. In the example above, you will find the expression under the lemma "bride" as "bride[à_bride_abattue]" and "abattre[à_bride_abattue]", but not under the meaningless preposition "à". But why don't we find "parce que" under "parce[parce que]"? Because "parce" is not a valid and existing lemma in the French language, hence its direct addition. This note also applies to anglicisms. In summary: for compound words, it's best to search visually (in the dictionary or autocomplete) with the key words that make up the expression or word.

- What are the used language registers? What is their importance?

Sometimes, following the example of consulted works, we indicate which language register or level a particular word belongs to. The distinguished language registers are *familiar language*, *more refined language*, *standard language*, *vulgar language*, and *colloquial language*. Most words and expressions included in our lexicon belong to the common standard language (*le registre courant*), meaning that this word is used in daily language in virtually all situations. The familiar language (*registre familier*) is used in casual conversations, typically not in more formal, serious situations. In English, this is sometimes referred to as informal language. The more refined language (*le registre littéraire / soutenu*) or formal language is characterized by sophisticated language and more complex, refined expressions. Then, we have the colloquial language level (*le registre populaire*), often synonymous with language used in socially disadvantaged environments. Vulgar language (*le registre vulgaire*) needs no explanation—it has very limited applicability and should be used cautiously. Some interlocutors are more sensitive than others and may take offense more easily at the use of vulgar terms. We mention the registers to give language learners some guidance when processing and using words and structures in a foreign language. When learning a foreign language, you have to do so without the norms of your mother tongue, which can sometimes lead to mistakes.

If we don't mention a language register for a word, you can assume that it belongs to the common language. We have paid particular attention to the familiar (informal) language usage when addressing different meanings. More information about language registers can be found in paragraph 6.

- What do the abbreviations SYN, LOC, EXP, CONT, ORT, PROV mean?

- **CONT(A: B):** This means that the antonym (opposite) of word A is word B. We have added this relationship only sporadically.
- **PROV:** This means that we are dealing with a saying or proverb. We will add these sayings in a later stage, as they have limited usefulness for language learners (and general users).
- **SYN(A: B, C, D,...):** This means that word A has the words B, C, D as synonyms. Note that sometimes this relationship applies only to one specific meaning of the word!
- **ORT(A, B, C):** This represents the reformed or alternative spelling of a word. For example, if you have a search result with the word 'connaît,' you should find later: (ORT: connait).

- **LOC()**: This abbreviation stands for "locution" (adjectival, adverbial, verbal, etc.) and indicates fixed word groups that often appear in the language. Phrases like 'à cause de,' 'pour que,' and 'demander pardon' fall under this somewhat vague definition. We also placed under LOC all reciprocal verbs or sometimes verbs that have an unexpected fixed preposition. For Dutch, LOC also includes non-separated forms of verbs (e.g., "doorlopen," "rondkomen"). For instance, if there is an analysis of the sentence "Jan speaks to Marie," the computer (and perhaps the student as well) might not automatically recognize the word 'afspreken' (to arrange or meet). Therefore, we add this under LOC.
- **EXP()**: This abbreviation stands for "expressions." Undoubtedly, the question arises about the difference between these and the fixed word groups we referred to earlier as "locutions." We consider a phrase an expression if, at first glance, it is not clear how the sum of the individual meanings leads to the overall meaning of the expression. The criteria for distinguishing between locutions and expressions are somewhat more complex, but for practical purposes, the program makes a rudimentary distinction. It doesn't matter much to the language learner whether something is a locution or an expression. In future updates, we will try to distinguish between the two more systematically. If you want to search for all fixed expressions with the word 'main,' for certainty, you should first search for 'LOC main' and then for 'EXP main,' so you can combine the results.

-

- Why doesn't this program offer spell checking and why doesn't it correct sentences?

This program doesn't have the necessary modules (a parser with rules) to syntactically analyze a sentence. To achieve this, we would need to add a parser that can determine whether a sentence is correctly formed or not. Unfortunately, it doesn't stop there, because since we are in a "learning context," this parse module would also need to be able to pinpoint exactly where the error lies (a notoriously difficult and unresolved problem) and suggest ways to improve the sentence. The latter is a significant challenge and cannot be quickly and easily implemented. Moreover, the program is not a spell checker, although it can be used for this purpose with some tweaks to the search queries. A spell checker performs a superficial parse of the text and suggests possible errors to the user based on the immediate proximity of surrounding words. For

example, if you want to know whether you can say "une vieux femme," you can enter the sequence "vieux femme", and the program will return the labels for both words. You will quickly notice that these labels are incompatible: "masculine plural, singular" for "vieux" does not match "feminine singular" for "femme"!

- Can this program replace a teacher?

As mentioned above, this program is purely supportive and not designed as a tool for fully autonomous language acquisition. The role of the teacher is to design the learning path and guide students through it by interacting with them and providing the necessary feedback on their progress. This feedback (on language correctness, etc.) is crucial when learning a foreign language and cannot simply be replaced by software. However, if the student has already gained a certain level of autonomy in the new language, they can largely map out their own learning path (for lexical knowledge) through the program. Interaction will, however, always remain essential.

- Where can you find the program? / Do you need the internet to use the program? / Are there additional costs for updates? / What about privacy?

You can purchase the program via the *Microsoft Store*, and once the download and installation are complete, you can use the program indefinitely under your personal account. Unlike many textbooks and language packages, Lexogoth does not require an internet connection to function, and all (quarterly) updates are offered for free via the *Store*. Therefore, there will **never** be **additional costs** once you have purchased and installed the program. Lexogoth does not collect any personal data from users. If you wish to purchase the program for a group, you can no longer do so through the *Microsoft Store*, as this functionality was discontinued a year ago; you will need to contact the author directly at info.Lexogoth@gmail.com for this.

- How do you update the program?

As mentioned above, you can only obtain *Lexogoth* via the *Microsoft Store*. We commit to releasing updates for the program every three months. These updates mainly focus on expanding the text corpus and improving any errors. The update process is as follows: Open the *Microsoft Store* app on your computer, then select *Library*. After that, you will see the option to *Update (Apps)*. When you click that, all apps will be updated. Keep in mind that updating *Lexogoth* must always be done manually through the *Microsoft Store* app and will not happen automatically via regular Windows updates. Organizations that have made a group purchase will receive an update via email.

- Comments regarding the installation of Lexogoth and Loading times

* **Important:** Starting from version 3.8 of Lexogoth (current version is v.3.15), the user no longer needs to install the *Microsoft Visual C++ redistributable*, at least version 14.40.x, architecture x64 first; we have already included the necessary files in the installer. If error messages appear when opening Lexogoth regarding missing files, visit the following link, download and install the Microsoft Visual C++ redistributable (architecture x64):
<https://learn.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-170> [under the title Latest Microsoft Visual C++ Redistributable Version]. After that, try reinstalling Lexogoth.

* If your Windows operating system is running in S mode (rare), the program may not function correctly. A solution could be to disable S mode or download and install the program on another computer. This possible limitation has to do with permissions that are not granted to programs by an operating system in S mode.

* If you notice that resizing the program is lagging, there may already be thousands of example sentences in the search results window, without you being aware of it. The easiest solution is to refresh your screen (refresh icon), and resizing will proceed smoothly.

6. The Reformed Spelling

In our lexicon, we have tried to take the 1990 reformed spelling into account as much as possible and to systematically add it to the corpus.

For most words, we use the traditional spelling and then indicate the updated spelling at the end of the sentence. You can recognize this by the addition **ORT**.

In principle, both spellings of a word are included in the search terms; however, there are still quite a few words for which we haven't yet been able to do this. Therefore, we recommend users to search for words with a **circumflex accent** both **with and without the accent** to ensure all relevant search results are returned. For example, if you want to search for "**croître**", you should search for both "**croître**" and "**croitre**".

We haven't yet applied this reformed spelling systematically. Do note that the pre-reform spelling should **not** be considered outdated or incorrect—this spelling remains valid **indefinitely** and is fully equivalent to the updated spelling. Just to make things a bit more complicated.

7. Varieties of French: Language Registers, Belgicisms, and Anglicisms

Lexogoth pays close attention to the different varieties of the French language. We usually indicate which language register a word or construction belongs to. *Lexogoth* distinguishes four broad registers for the vocabulary and constructions covered:

- **Standard or everyday language:** This is the language used in daily life, such as at work, school, and in the media.
- **Formal, refined, or literary language:** Vocabulary and constructions that require a more advanced lexical knowledge and are mainly found in written French.
- **Familiar language:** Informal language used between people who know each other well, mostly in spoken communication.
- **Popular (colloquial) register:** Extremely informal language, almost exclusively oral, characterized by casual pronunciation, limited lexical and syntactic variation, simple constructions, the use of coarse (vulgar) words, *verlan* (French slang inversion), and Anglicisms. For these reasons, this is considered the lowest register. It generally has limited communicative value for most language learners. Users are advised to handle this register with caution and avoid using it without a clear understanding of the communicative context.

For further explanation of the characteristics of these registers, users can consult the document in the grammar section of *Lexogoth*: **les registres de langue**.

Lexogoth also always notifies users when a word or expression is a **Belgicism**, a term primarily used by Belgian French speakers and rarely found outside Belgium. Such expressions are not considered part of standard French. Many of these terms originate from regional Belgian French, Dutch (especially Flemish or Standard Dutch), or Walloon, a closely related Romance language historically spoken in southern Belgium and largely replaced by French after World War I. Well-known examples include *nonante*, *septante*, and the excessive use of *savoir*. It is up to the user to decide, based on their communicative goal, whether or not to use such expressions. Most lexical Belgicisms can be easily found in Lexogoth by searching for the term “**belgicisme**”. Note that *Lexogoth* focuses almost exclusively on **French** Belgicisms.

French is also not immune to English influence, and *Lexogoth* pays due attention to **Anglicisms**—words or constructions borrowed directly from (mostly American) English. Users can easily find them by entering “**anglicisme**” in the search field (with language set to French).

8. To Do in the next updates

- During the software development of Lexogoth, we exclusively used AI systems (Gemini and CoPilot, the respective AI engines of Google and Microsoft) to identify persistent errors, eliminate layout issues (within the Qt framework), and extensively test and debug the logic used (for searching). The concept, the didactic vision, and the concrete development (workflow, selection of algorithms, etc.) are entirely the work of the author. We also note that the heart of Lexogoth, the French corpus sentences, is completely authentic, which means they were spoken or written by native speakers and were not generated by AI. The notes were compiled by the author, naturally in consultation with reference works, but without intervention from AI. For the translations into English (purely to support comprehension), we did

regularly use translation engines (including Google Translate) to accelerate the work.

- We strive to release an update with corrections and expansions of the corpus every four months. Next update: end of October 2026.

9. Versions and updates

- v4.0 (July 2026, current):

Search Engine and Interface Enhancements: Standardized search logic across the entire application (including Toolbox), implementation of clear visual highlighting, refined parameter settings (words before/after), improved readability of text materials, streamlined general interface, addition of splash screen information, and enhanced overall user experience.

Module Integration and New Functionalities: Introduction of the Text Analysis module (POS-tagging, tense recognition, integrated concordancer), launch of the Collocations & Syntax module (syntactic patterns, co-occurrences, auto-correct), cross-module integration (searching corpus data within the database, test, and collocation modules), and expanded management tools (merging context and test modules, direct addition of corpus sentences to tests).

Advanced Data Processing and Technical Stability: Implementation of advanced database search options (batch deletion, N-gram analysis), improved validation logic for imports, security hardening against HTML injections, optimized memory management for better stability, faster startup times, increased search speeds for large corpora, and enhanced export capabilities (RTF).

Corpus Maintenance and Content Expansion: Ongoing systematic updates including corrections and expansions of the Lexogoth corpus, improved data accuracy, and the addition of functional data tools to transfer corpus notes directly into personal databases.

- v3.0 (May 2026): internal test version

- **v2.0 (January 2026):**
 - Addition of the Toolbox module
 - Massive expansion of entries and comprehensive corrections;
 - New autocorrect function for search terms
 - Complete revision of the layout and search logic.
- **v1.4 (september 2025): update: corrections, additions**
- **v1 (July, 2025):** First international (French – English) version of Lexogoth, based on the existing Dutch-based version.

10. Technical matters and copyright

Lexogoth was written in C++ and compiled with Visual Studio (version 17.12.3). The graphical user interface (GUI) was built using Qt under an LGPL license (version 6.4; <https://www.qt.io/>). We also used some library functions from Boost (version 1.78; <https://www.boost.org/>).

The works and websites consulted for the compilation of the corpus include; see also the chapter on *methodology*:

- dictionaries :

<https://www.larousse.fr/>
<https://dictionnaire.lerobert.com>
<https://www.dictionnaire-academie.fr/>
<https://www.vandale.nl/gratis-woordenboek/frans-nederlands>
<https://fr.wiktionary.org/wiki/Wiktionnaire>
<https://dictionnaire-commercial.com/>
<https://www.dictionnairedesfrancophones.org>
<https://www.collinsdictionary.com/dictionary/french-english>
<https://www.languefrancaise.net/Bob/>

- Linguistic recommendations :

<https://vitrinelinguistique.oqlf.gouv.qc.ca/>

<https://usito.usherbrooke.ca/>

<https://www.larousse.fr/>

<https://etienne-meul.be/>

<https://www.deepl.com/en/translator>

- Lexicons

French Lexicon LeFFF, published under LGPL. The original version used, 2.1, can be found here:

<https://www.labri.fr/perso/clement/lefff/telechargement.html>

English Lexicon: custom compilation based on a corpus.

- corpus

The text corpus (version 4.0) contains over 40,000 clickable items. These were compiled over a period of about ten years. The examples were mostly rewritten (with the aim of easy acquisition by language learners) and come from the following sources: blogs, internet forums, websites, dictionaries (see above), language advice websites, films and TV series, written and spoken press (mainly from Belgium and France), and government documents.

===== Lexogoth corpus & metadata =====

File: -

Validated records: 13 453

Totaal number of chunks: 40 361

French tokens (Source): 486 658 tokens

English tokens: 421 882 tokens

Notes: 370 674 tokens (incl. tags)

Information density: 0.76 (Ratio Notes / French tokens)

=====

Lexogoth International - Version 4.0 (July 2026)

Author : Frederik Meul, 2023

Contact : info.lexogoth@gmail.com

**Websites: <https://english.lexogoth.be> and
<https://sites.google.com/view/lexogothinternational/welcome>**