Lexogoth

International Version: French - English

Manual

version 2.0, January 2025

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

1. Introduction

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

Lexogoth is a desktop application designed for the Win11-platform, primarily aimed at deepening and significantly expanding the lexical knowledge of French students who have already surpassed the beginner level or have a reasonably advanced understanding of the language. This is done using common and effectively used examples from everyday spoken and written language. The user of the program must understand that *Lexogoth* is **not a translation program but a textboo**k, with the ultimate goal of helping the language learner maintain and expand their lexical knowledge.

1.2 Who is Lexogoth intended for?

Lexogoth is intended for :

Intermediate to advanced students: They can use this program—especially the included text corpus—while completing communicative exercises or as support when studying vocabulary. Attention is consistently paid to language registers that are inevitably linked to the use of certain words or meanings, so that students gain a clear understanding of which meaning fits which context. Students are also encouraged to use the program as a “book”, regularly selecting certain entries and studying the accompanying chunks by reading.

Language teachers: They can compile a corpus of clear examples based on the program and make it available to their students. Additionally, they can use projection to demonstrate the use of certain words to the entire class. Finally, they can design individual remediation paths for grammar and vocabulary using the available lessons.

Beginning students: Many examples may be too difficult for students at the beginner level. However, the corpus also includes a substantial number of simple chunks or sentences that are highly suitable for beginners. Lexogoth also contains the full content of the A1-A2 level (CEFR) for grammar, language functions, and vocabulary. These lessons are easily accessible through a clearly structured menu—very practical for quickly reviewing something during a class or exercise.

1.3 Background to the design of the program

Words, their meanings, and their combinatorial possibilities constitute, in our view, the greatest challenges for a foreign language learner. For each word, a student must learn a whole range of additional elements that allow them to use the word correctly: pronunciation and spelling, syntactic possibilities, the language register to which the word belongs, the various meanings, and its possible combinations with other words. All these elements make vocabulary acquisition in a foreign language a highly demanding task.

Strangely enough, this complexity is not reflected in the way many textbooks introduce new vocabulary. All too often, we observe that many textbooks and methods restrict themselves to compiling lists of words with their translations, sometimes accompanied by a short example sentence. *Such an approach, in our opinion, leads to a very superficial and passive lexical knowledge in students, preventing them from using newly learned words correctly and fluently in communicative tasks*.

To address this gap in foreign language education, we created *Lexogoth*. Through this program, students who have already passed the beginner stage are given the opportunity to acquire new vocabulary quickly and effectively or to deepen their knowledge of previously learned vocabulary. *Lexogoth* allows students of French to search for user examples of French words (and combinations), based on authentic usage in everyday spoken and written language. For many words, numerous examples

are already available in *Lexogoth*, clearly illustrating the most frequent constructions and collocations. In this way, students can form a more complete and accurate understanding of word usage and more rapidly move on to communicative practice.

In the current version (july 2025; version 3.16), more than 40,000 clickable items are available, each leading to a number of sentences or sentence fragments, also referred to as *chunks*. These chunks may consist of expressions, fixed word groups, common (and less common) word combinations, or shorter, easily memorizable example sentences that illustrate both the meaning and the usage of the word.

The idea is that students of French, while doing productive exercises or studying vocabulary, systematically consult and read the chunks via the searched lemma (or lemmas). ***For many lemmas, multiple search results are available, and by frequently consulting these chunks, users will eventually assimilate them and build up an extensive (mental) network of word combinations that can be immediately activated in communicative exercises. Of course, the reader is confronted with a large number of examples – but that is precisely the purpose of Lexogoth: by repeatedly exposing students to frequent, recurring constructions (repetition...), they will eventually be able to assimilate them with ease.***

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 use. So, how can we quickly find a word or expression among all these examples? There are essentially **three search methods** in the program, the first two of which are accessed through the **search bar at the top left** [see photo 1]:

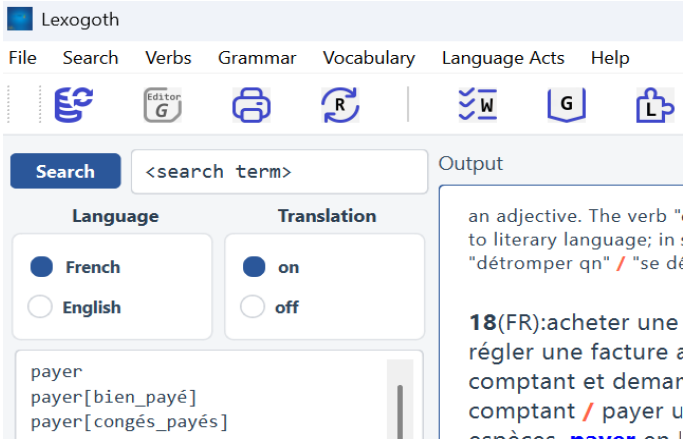


photo 1

*** Method 1 :**

You select the language in which you want to perform a search and indicate whether or not you want a translation of the sentences. Then, you type the word you're looking for. As soon as you type the first letters of your search term, the white box below will be filled with possible completions of your query (photo 2):

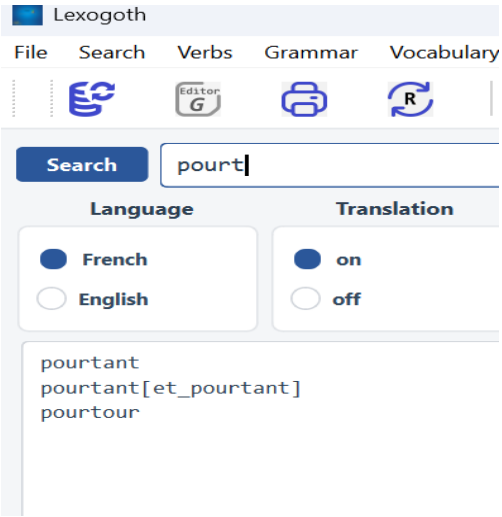
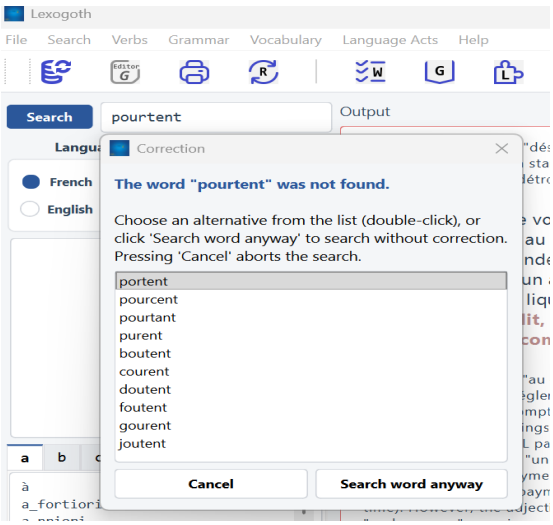


photo 2

If the word you're looking for appears in the list, you can double-click it, and the results will be shown in the main window. The searched word and its derived forms will be highlighted in blue (photo 3, next page).

If you enter a word that is not recognized, a popup will appear and propose other words which resemble the word you typed. If, for instance, you enter “probleme” (without an accent), a box will appear which allows you to correct your mistake by choosing a word from the proposed list. If you wish to ignore the suggestions (for instance when you searched for a proper noun), then just press “Search anyway”:



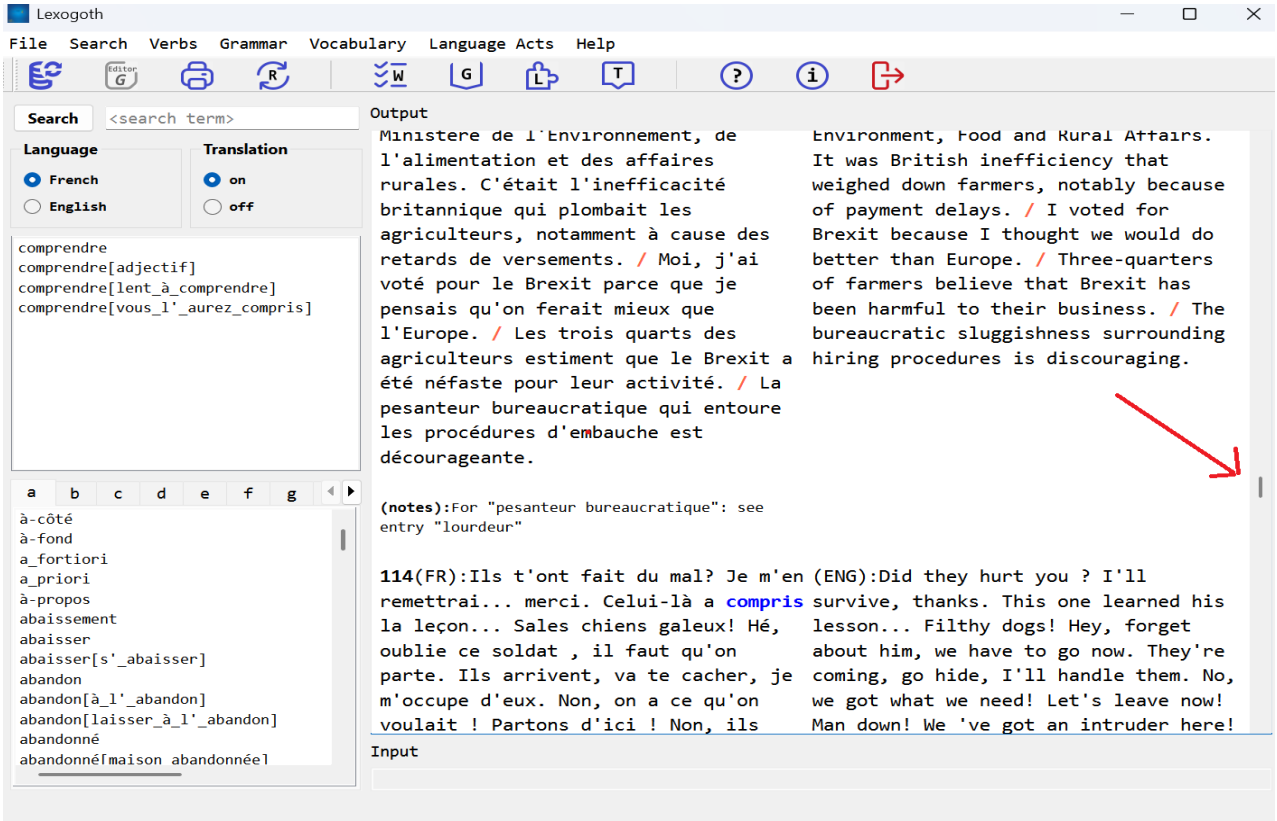


photo 3

Important: If this isn't your first search, be aware that the **scroll bar** (which determines the viewing position in the output window [**see red arrow in photo 3**]) will always reset itsel at the beginning of each new search result. Make sure to clear the screen using the refresh icon before starting a new query after a certain number of queries to ensure the interface remains responsive.

* method 2

You can choose to ignore the autocomplete suggestions and instead confirm your query by clicking the **Search** button or simply pressing **Enter**. When you search this way, Lexogoth will return two results:

- **Search result 1:** If the word is known, the program returns its **base lemma and word category**, using a form-based word corpus (approx. 350,000 French forms; 30,000 English forms). For example, if you enter *vieille*, the program will indicate that it is a **feminine singular adjective** derived from the lemma *vieux*.

- **Search result 2:** This is the result from the **text corpus**, where the occurrences of the searched word are highlighted in blue.

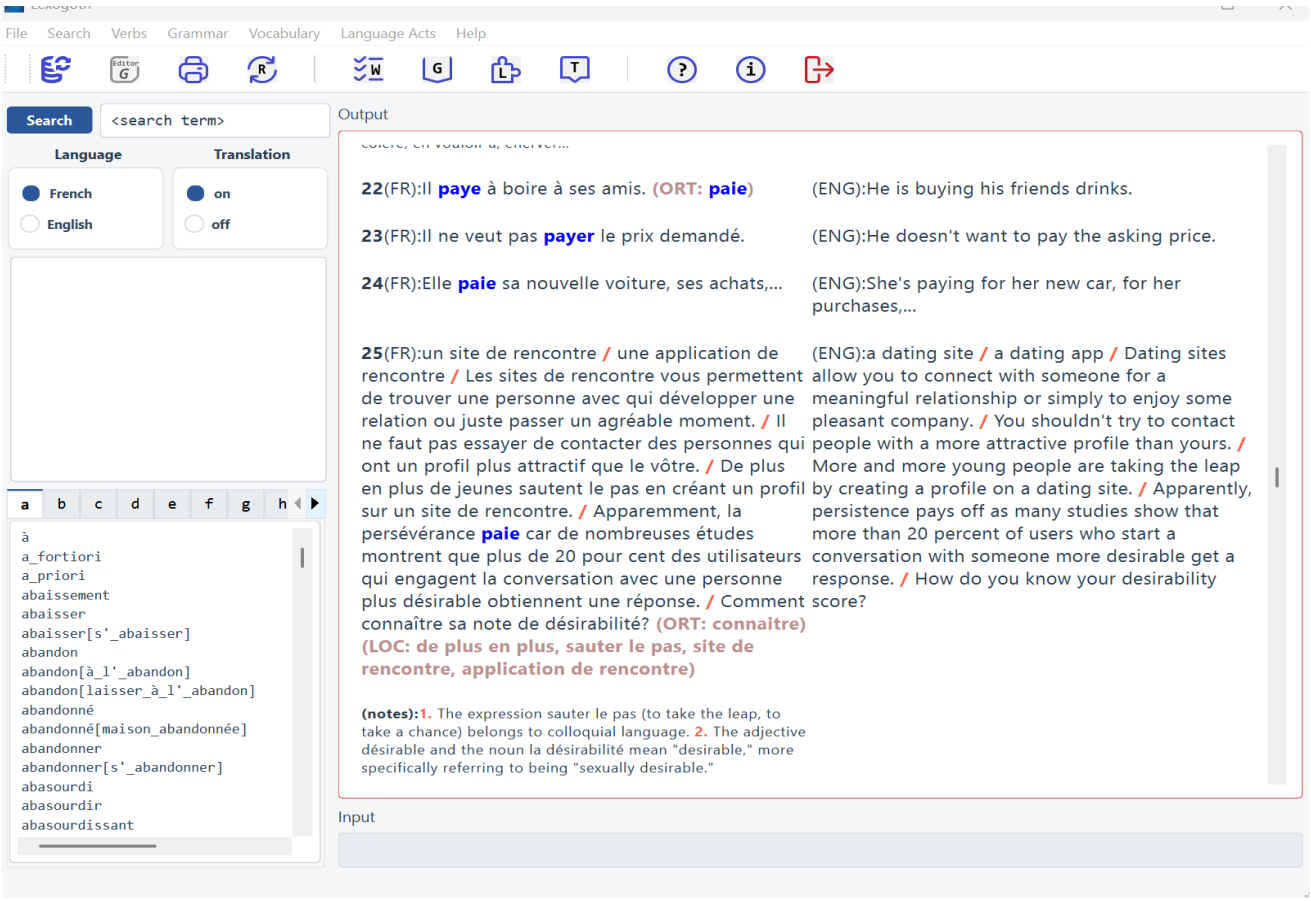
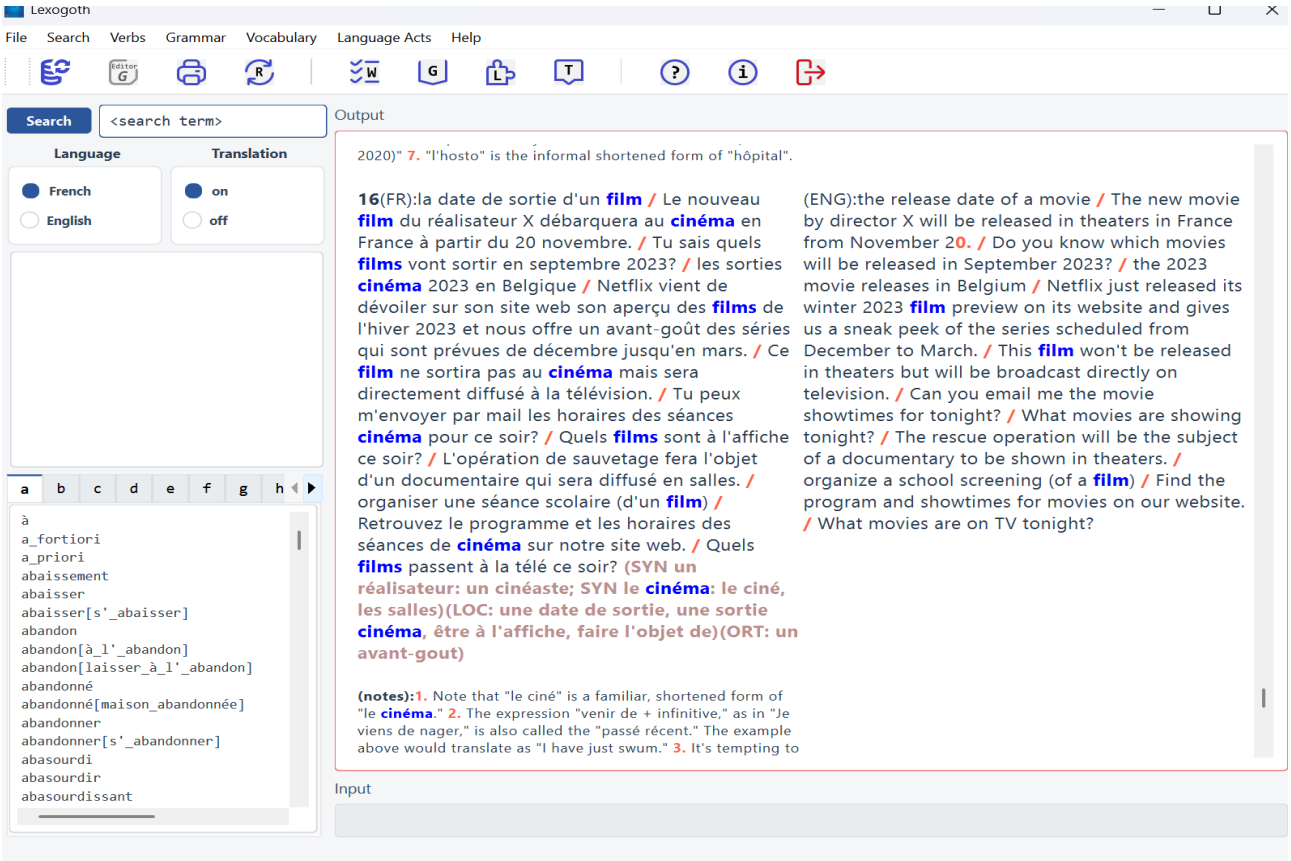


photo 4

With this method, you can also search for combinations of **multiple terms** that appear together in a sentence. For example, if you want to find out how to say "broadcast a movie on television," you might try searching for the combination *télévision* and *film*. Type both words, separated by a space, and hit search. If the combination appears anywhere in the corpus, the result will be shown in the main window. Keep in mind, though, that not all combinations will yield results. If nothing appears, you may need to broaden your search or search using only one term (e.g., *film*) and manually analyze the examples.



This method is also useful for finding **expressions, fixed phrases**, and idioms. Suppose you want to look for the expression *ne pas y aller par quatre chemins*, then typing *aller* and *quatre* might be enough to get the desired result. Again, a successful result isn't guaranteed, as each entry can contain multiple chunks, and it's possible that *quatre* appears in one and *aller* in another.

Note: This method is **not** recommended for searching compound expressions such as *parce que* or *pour que*. For that, use **Method 3**. (See also: *FAQ > Searching for exact matches with multi-word expressions*.)

* method 3

Lastly, you can search via a **classic dictionary interface**. On the left side of the interface, you'll find tabs with letters of the alphabet: select the desired language, click on the **first letter** of your search term, and scroll through the list until you find the word you're looking for. This method is particularly helpful when looking for **multi-word expressions**. For example, to search for *faire face à*, go to the word

face, scroll until you reach *face[faire_face_à]*, and double-click to display the result on your screen [photo 5].

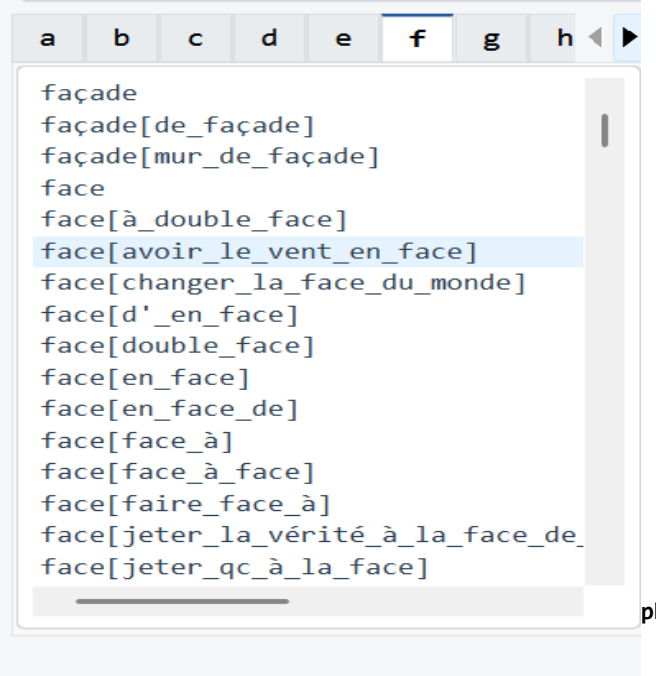


photo 5 : dictionary search

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**.

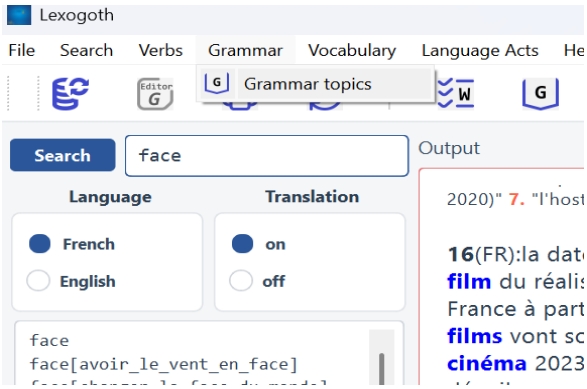


photo 6

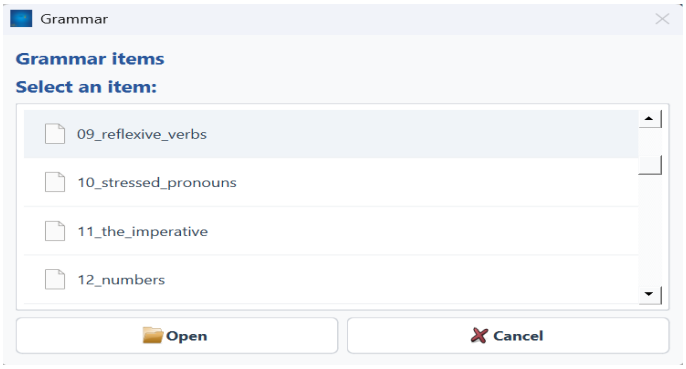
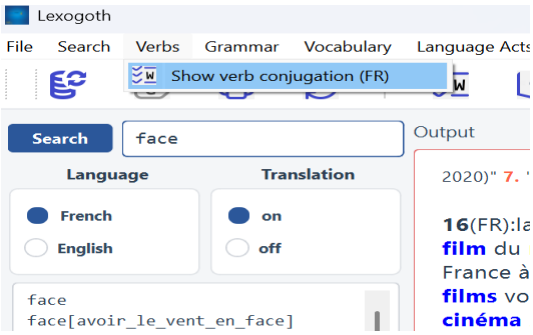


foto 7

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].



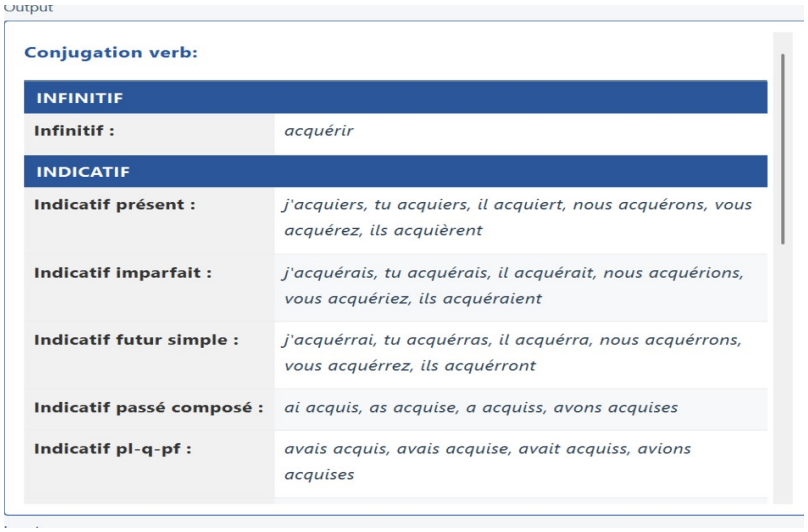


photo 8 : Verb conjugation
photo 9 : result: « acquérir » →

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.

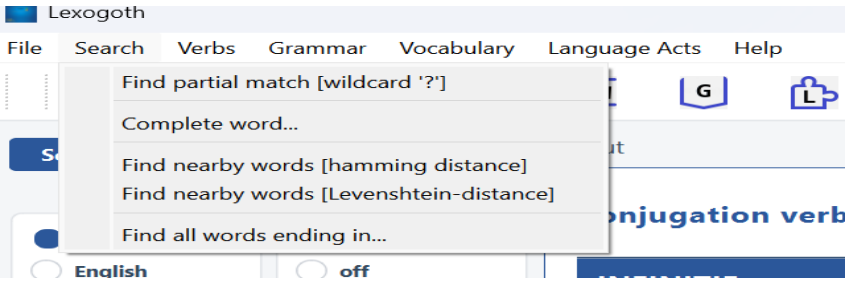


photo 10 : *Search menu*

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.

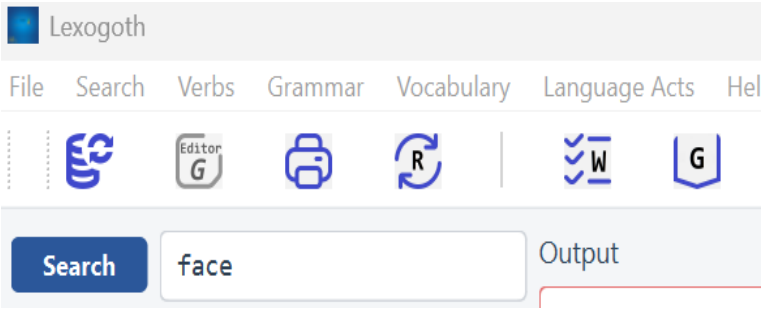
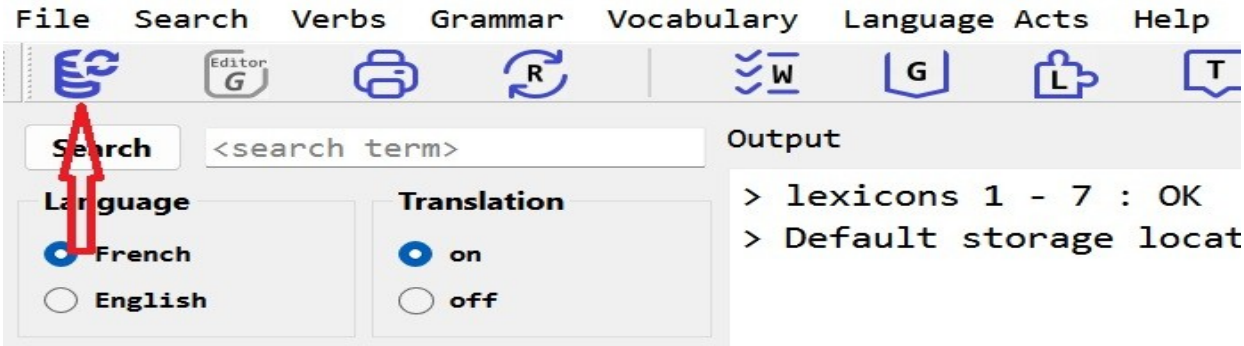


Photo 11 :(Icons from left to right: Open Lexical Toolbox, open Golgoth editor [disabled in International version], print output screen, **refresh screen (clear screen)**, conjugate verb, Vocabulary lessons)

2.8. The Lexical Toolbox



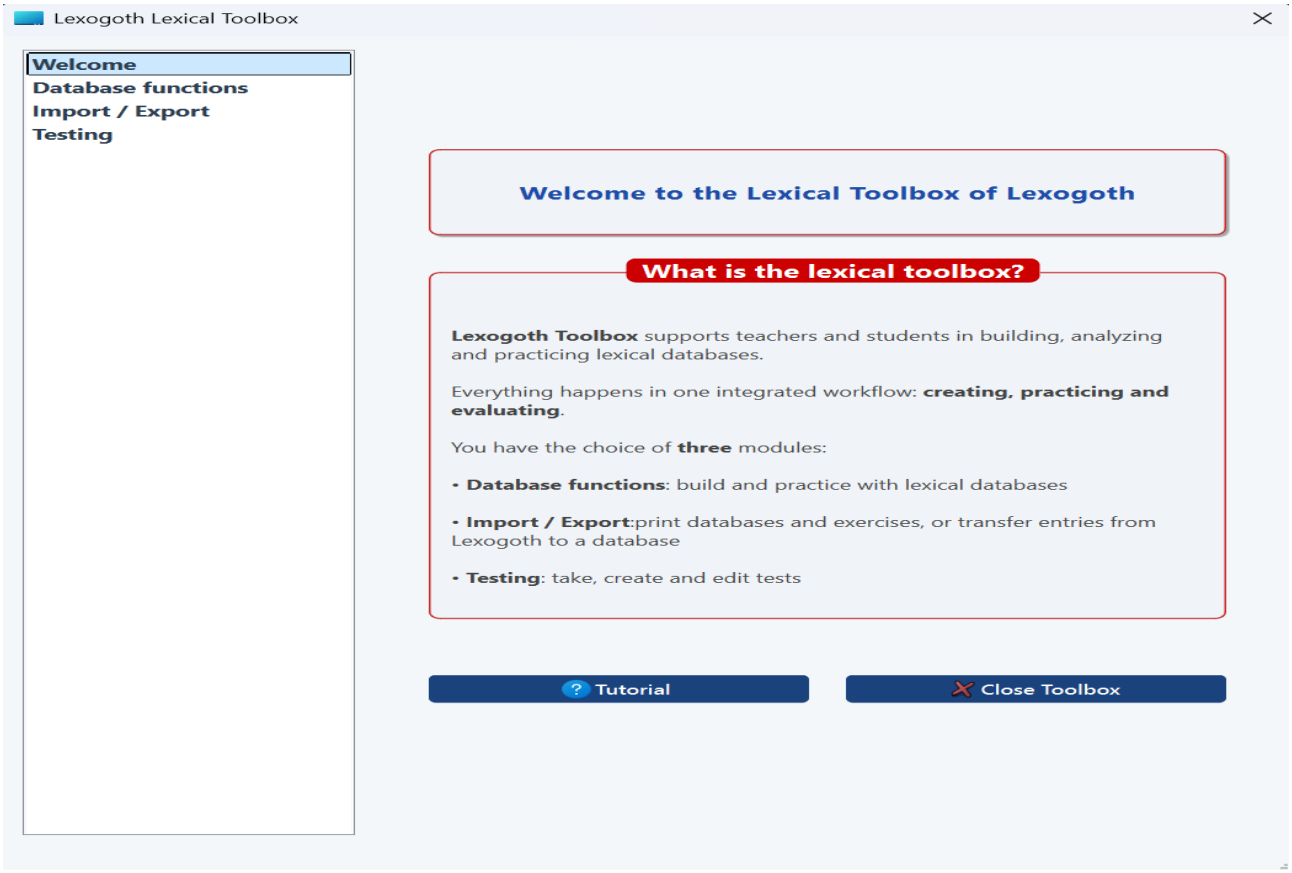
In the illustration above, clicking on the first icon will open the *Lexical Toolbox*. This module is new and is being rolled out from *Lexogoth International* v2. The *Lexical Toolbox* of *Lexogoth* is a module for students and teachers of French, designed to practice vocabulary and lexical patterns using authentic corpus data. The program is built around three main modules:

(1) Database functions: These allow users to create lexical databases, edit existing databases, and practice the databases by hiding the Dutch or French translation.

(2) Import / export: This module forms the heart of the Toolbox and allows users to import entries directly from Lexogoth and filter them with advanced search options (AND/OR/Exact string match). 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 questions and dropdown questions. For more information, we refer the user to the manual which he can consult in the startup screen of the Toolbox.

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.

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*)
- The use of corpora with authentic material to support the learning process (corpus linguistics)

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 is corpus linguistics, a linguistic discipline that emerged in the 1960s and typically uses computers to systematically collect and analyze large amounts of language data to identify patterns. Well-known English-language corpora available online include The Corpus of Contemporary American English (COCA) and The British National Corpus (BNC). Thanks to the growing accessibility of computers in the 1990s, more people—including language educators—gained access to corpora that had previously been limited to research institutions. This opened the way to exploring the use of text corpora in foreign language teaching. From the 2000s onwards, various initiatives sought to bring authentic text corpora to students and integrate them into teaching through electronic dictionaries supplemented with corpus data, concordancers for studying word usage and syntactic structures, and syllabi of expressions and word combinations based on frequency.

Both developments —the Lexical Approach and corpus linguistics—have been the methodological and theoretical foundation that inspired us. We start from the idea that, once a student has reached a certain level, language acquisition—especially in a foreign language—should focus mainly on learning authentic and context-bound lexical units (frequent word combinations, phrases, fixed expressions, etc.) that we have carefully compiled using corpus linguistics tools. Lexogoth combines in one program many of the advantages of a concordancer and a dictionary. It also provides useful advice on the use of words and word combinations (frequency, meaning clarification, grammatical issues, register, pronunciation, etc.).

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 Dutch 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.

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.languefrançaise.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 coeur / 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 Dutch. 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.

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.

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.

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.

– 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.

- 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.

- 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

We aim to release an update every four months that includes corrections and expansions of the corpus.

9. Versions and updates

- v2.0 (January 2026, current):

- 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 3.16) 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 International - Version 2.0 (January 2026)

Author : Frederik Meul, 2023

Contact : info.lexogoth@gmail.com

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