
Lexogoth



Handleiding

versie 4.9, februari 2026

Auteur: Frederik Meul

0. Workflow in Lexogoth: een beknopt overzicht

1. Inleiding

1.1 Wat is Lexogoth en wat beoogt het programma ?

1.2 Voor wie is Lexogoth bedoeld ?

1.3 Achtergrond bij het ontwerp van het programma

1.4 Een voorbeeld: 'La confiance'

2. Hoe werkt het programma ?

2.1 Het doorzoeken van de lexicale database

2.2 Grammatica, woordenschat en taalhandelingen

2.3 Werkwoorden vervoegen

2.4 Het menu *Zoeken*

2.5 Het menu *Help*

2.6 Het menu *Bestand*

2.7 Het scherm leegmaken (*refresh*-icoon)

2.8 De Lexicale Toolbox (^{nieuw})

3. Hoe gebruik je het programma als student of leerkracht ?

4. Methodologie van Lexogoth

4.1 Ontstaan van Lexogoth en primaire doelstelling

4.2 Samenstelling van het corpus: op basis van welke criteria werden woorden en structuren geselecteerd ?

4.2.1 Theoretische basis voor Lexogoth

4.2.2 Selectie en opbouw van het corpus

4.2.2.1 Een empirisch gebaseerde aanpak

4.2.2.2 Structuur van het corpus

4.2.2.3 Structuur van de *entries* in Lexogoth

4.2.3 Taalregisters, geografische varianten en taalvariëteiten

4.3 Hoe onderscheidt Lexogoth zich van andere taalapps en websites ?

4.4 Hoe studeren we nu best woordenschat en hoe kan Lexogoth hierbij helpen ?

4.5 Een lopend project

4.6 Geciteerde werken

5. FAQ

6. Hervormde spelling en belgicismen

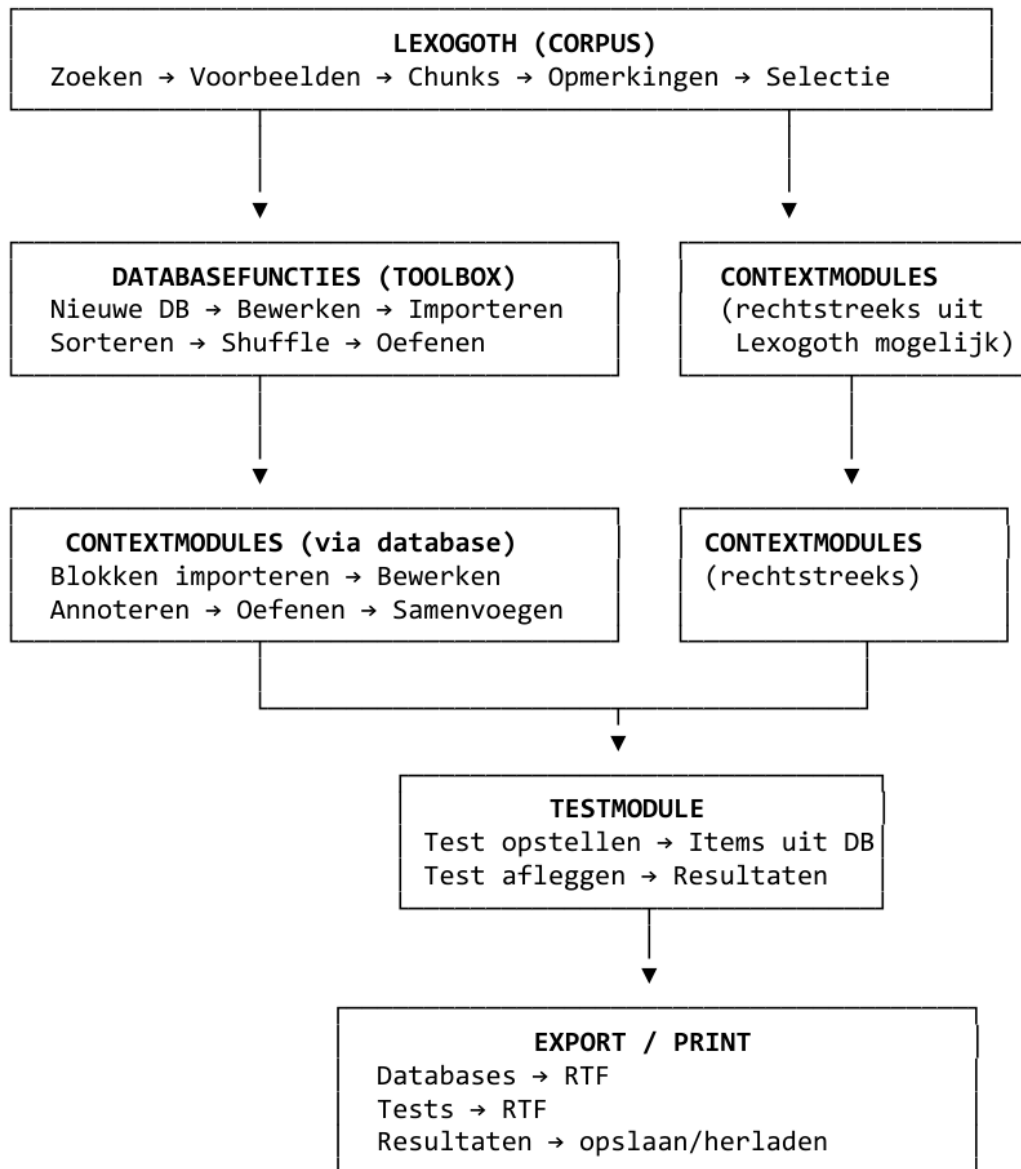
7. Variëteiten van het Frans: taalniveaus, belgicismen en anglicismen

8. *To Do* in volgende updates

9. Versies en updategeschiedenis

10. Technische zaken en copyright

0. Overzicht van de workflow in Lexogoth



1. Inleiding

1.1 Wat is Lexogoth en wat beoogt het programma ?

De technologische evolutie in het Vlaamse onderwijs, versneld door de *Digisprong*, heeft tot dusver de pedagogische verwachtingen niet volledig ingelost. Door een gebrek aan een diepgaande visie achter de implementatie, is het (secundair) onderwijs vaak overgeleverd aan digitale bordboeken, veredelde pdf's en logge cloud-systemen voor automatische testen. Er is een opvallend tekort aan leermiddelen die de computer ten volle benutten voor het creëren van nieuwe, waardevolle leermiddelen die voorheen onbereikbaar waren. Lexogoth wil deze leemte vullen door technologie niet in te zetten om analoge leermiddelen in een digitaal jasje te stoppen maar door een echt digitaal didactisch ecosysteem te creëren dat studenten en docenten de middelen geeft om het leerproces efficiënter en moderner te maken, en zo de grens te verleggen van wat binnen een traditioneel, analoog taalonderwijs technisch en didactisch mogelijk was. *Het biedt met andere woorden geen digitale ondersteuning van bestaande methoden maar een alternatief didactisch model voor woordenschatverwerving, dat gebaseerd is op context, corpusdata en authentiek gebruik.*

Lexogoth is een desktopprogramma voor Windows 11 waarmee studenten en docenten Frans gevorderde woordenschatkennis (vooral ERK-niveaus B en C) van het hedendaags Frans kunnen verwerven, inoefenen of onderwijzen. Het programma biedt een ruime selectie van veelgebruikte Franse woorden die telkens worden geïllustreerd met *duidelijke en contextuele voorbeeldzinnen* uit de gesproken en geschreven taal. Lexogoth is geen klassiek referentiecorpus en ook geen traditioneel woordenboek, maar een didactisch verrijkte lexicale kennisbank. De aangeboden taaldata zijn niet louter verzameld, maar bewust geselecteerd, gestructureerd en

geannoteerd met het oog op leerbaarheid, betekenisonderscheid en contextgevoelig gebruik. Betekenissen worden niet alleen getoond, maar ook gedifferentieerd, geïnterpreteerd en expliciet gemaakt, zodat de gebruiker inzicht verwerft in semantische nuances, gebruiksvoorwaarden en pragmatische beperkingen. Gebruikers moeten zich ervan bewust zijn dat Lexogoth ook geen vertaalprogramma is, maar eerder functioneert als een interactief leerboek en studie-instrument (en op termijn misschien referentiewerk) waarvan de uiteindelijke bedoeling is dat de taalleerder zijn lexicale kennis op een duurzame en efficiënte wijze onderhoudt, verdiept en uitbreidt. De focus ligt niet op snelle equivalenten, maar op begrip, context en correct gebruik. De kern van dit alternatieve leermodel is de **Lexicale Toolbox**: hij maakt selectie, annotaties, personalisering en actieve verwerking mogelijk en dat zijn net processen die in traditionele methoden ontbreken. Waar het corpus de bron van kennis is, fungeert de Toolbox als de persoonlijke werkplaats van de gebruiker. Hiermee kan de taalleerder een eigen, dynamische kennisbank opbouwen door relevante *chunks* en modellen uit het corpus te selecteren, te annoteren en te organiseren. Dit proces transformeert Lexogoth van een raadpleegbaar instrument naar een gepersonaliseerd leersysteem dat de brug slaat tussen ontdekking en eigen taalproductie.

1.2 Voor wie is Lexogoth bedoeld ?

Lexogoth is bestemd voor :

- *redelijk gevorderde studenten (ERK-niveaus B en C)*: zij kunnen dit programma, meer bepaald het aanwezige tekstcorpus, gebruiken tijdens het maken van communicatieve oefeningen of ter ondersteuning bij het studeren van woordenschat. We geven steeds de nodige aandacht aan de taalregisters die onvermijdelijk verbonden zijn met het gebruik van bepaalde woorden of betekenissen zodat de student een goed beeld krijgt

van welke betekenis in welke context kan gebruikt worden. Studenten dienen het programma ook te gebruiken als 'leesboek' en op regelmatige basis bepaalde lemma's te selecteren en de daarbij horende chunks **al lezend te studeren**. Via de **Lexicale Toolbox** kunnen zij bovendien hun eigen 'modeldatabanken' aanleggen, databases beheren en hun verworven kennis gericht inoefenen via de geïntegreerde testmodules en contextmodules.

- *taalleerkrachten* : Zij kunnen op basis van het programma een corpus van duidelijke voorbeelden opstellen en dit ter beschikking stellen van hun studenten. Verder kunnen zij via projectie het gebruik van bepaalde woorden tonen aan de hele groep. Dankzij de Toolbox-functionaliteiten kunnen docenten bovendien op maat gemaakte oefenmodules en woordenschatlijsten (databases) samenstellen, deze distribueren naar studenten en digitale testen genereren die direct aansluiten bij de behandelde lesstof. Ten slotte kunnen zij ook individuele remediërings-trajecten voor *grammatica* en *woordenschat* uittekenen aan de hand van de beschikbare lessen (ERK-niveaus A - B).

- *beginnende studenten (ERK-niveau A)* : Heel wat voorbeelden zijn wellicht te moeilijk voor studenten die op het beginniveau zitten. Het corpus bevat echter ook een flink aantal eenvoudige *chunks* of zinnestukjes die voor beginners uitermate geschikt zijn. We namen in Lexogoth ook de volledige leerstof van het traject A1-A2 (ERK) voor *grammatica*, taalhandelingen en woordenschat op. Deze lessen -ongeveer 90- zijn makkelijk raadpleegbaar via een duidelijk gestructureerd menu. Erg praktisch om tijdens de les of bij een oefening snel iets te herhalen.

1.3 Achtergrond bij het ontwerp van het programma [zie hoofdstuk 4: *methodologie voor een grondigere behandeling van de achtergrond*]

Woorden, hun betekenissen en hun combinatiemogelijkheden vormen volgens ons de grootste uitdagingen voor een

vreemdetaalleerder. Bij elk woord moet een student immers een heleboel andere elementen leren die hem in staat stellen om het woord correct te gebruiken: de uitspraak en schrijfwijze, de syntactische mogelijkheden, het taalregister waartoe het woord behoort, de verschillende betekenissen en de combinatiemogelijkheden met andere woorden. Al deze elementen zorgen ervoor dat het aanleren van woorden in een vreemde taal best een heel uitdagende opdracht is. Spijtig genoeg wordt de complexiteit van dit proces niet weerspiegeld in de manier waarop vele handboeken omgaan met het aanbrengen van nieuwe woordenschat. Maar al te vaak stellen we vast dat vele handboeken en methoden zich beperken tot het aanleggen van lijsten met woorden en hun vertalingen, al dan niet met een kort voorbeeldzinnetje. Een dergelijke aanpak leidt volgens ons tot een zeer oppervlakkige en passieve lexicale kennis bij studenten en laat hen niet toe om deze nieuw geleerde woorden correct en vlot te gebruiken in communicatieve oefeningen. Om in deze lacune van het vreemdetalenonderwijs te voorzien, creëerden we *Lexogoth*. Via dit programma bieden we studenten, die reeds het basisniveau voorbij zijn (ERK: B en C), de kans om op snelle en efficiënte wijze nieuwe woordenschat te verwerven of reeds gekende woordenschat uit te diepen. Het programma biedt aan studenten Frans immers de gelegenheid om gebruikersvoorbeelden op te zoeken van Franse woorden (en combinaties) die gebaseerd zijn op voorbeelden uit de alledaagse spreek- en schrijftaal. Voor vele woorden zijn er al talrijke voorbeelden aanwezig in *Lexogoth* die duidelijk de frequentste constructie- en combinatiemogelijkheden illustreren. **Op deze manier kan een student zich dan een vollediger en correcter beeld vormen over het gebruik van bepaalde woorden en sneller overgaan tot communicatieve oefeningen.** In de huidige versie zijn er al meer dan 45 000 aanklikbare items beschikbaar die elk leiden naar een aantal zinnen of stukken van zinnen, ook wel *chunks* genoemd. Deze chunks kunnen uitdrukkingen, vaste woordgroepen, courante (en minder courante) woordcombinaties of kortere maar makkelijk

leerbare voorbeeldzinnen zijn die de betekenissen én het gebruik van het woord illustreren.

Het idee is dus dat studenten Frans, zowel bij het maken van productieve oefeningen als bij het studeren van woordenschat, de beschikbare chunks via de gezochte lemma's systematisch raadplegen. Voor elk lemma vindt de gebruiker talrijke zoekresultaten die de taalleerder door frequente consultatie helpen om een ruim mentaal netwerk van woordcombinaties uit te bouwen.

Natuurlijk wordt de lezer hierbij geconfronteerd met een grote hoeveelheid data, maar dat is net de kracht van Lexogoth: door de student herhaaldelijk bloot te stellen aan courante, steeds terugkerende constructies, vindt een natuurlijke assimilatie plaats. Om dit proces van passieve herkenning naar actieve beheersing te versnellen, beschikt het programma over een krachtige **Lexicale Toolbox**. Deze module fungeert als een interactieve brug: studenten kunnen de meest relevante modellen en chunks uit het corpus selecteren en opslaan in persoonlijke databases. Door deze actieve verwerking – het selecteren, annoteren en systematisch inoefenen van deze zelfgekozen data – worden de woordcombinaties duurzaam verankerd en kunnen ze onmiddellijk en correct worden ingezet in communicatieve oefeningen.

1.4 Een voorbeeld: ‘*La confiance*’.

Een concreet voorbeeld om te verduidelijken hoe Lexogoth nuttig kan zijn voor studenten Frans : een *goede* student die de individuele woorden ‘*(une) confiance*’ en ‘*avoir*’ kent, kan misschien de zin ‘*Elle a de la confiance dans ses amies.*’ vormen. Hierop kan je eigenlijk niet veel zeggen want de student heeft immers met de lexicale (woordenlijsten!) en syntactische kennis die hij ter beschikking heeft een zin gemaakt die hem correct overkomt; spijtig genoeg zal je een dergelijke zin (bijna) niet tegenkomen en zullen de meeste taalsprekers deze als eerder onnatuurlijk ervaren. Via

Lexogoth kan een student nu verschillende gebruikersvoorbeelden terugvinden die hij kan studeren en die hij daarna met enkele minimale ingrepen onmiddellijk kan inzetten in communicatieve oefeningen. Voor ‘*la confiance*’ zouden we bijvoorbeeld de volgende chunks / zinnen kunnen aanreiken :

- (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)

Het is echter niet de bedoeling om een taalleerder te forceren om al deze voorbeelden, die het correcte gebruik van ‘*confiance*’ illustreren, plots uit het hoofd te laten leren ; dat zou een moeilijke en voor velen onhaalbare taak zijn. De student kan deze voorbeelden wel grondig doorlezen en een of twee modellen kiezen die het beste passen bij hetgeen hij nodig heeft. ***Een frequent gebruik en herlezen van dezelfde voorbeelden tijdens de hele opleidingsduur zal er waarschijnlijk toe leiden dat de student zich deze chunks langzamerhand eigen zal maken en zo systematisch een groot netwerk van bruikbare lexicale kennis zal uitbouwen.*** Het gaat immers in deze chunks niet alleen om het lemma ‘*confiance*’ maar ook bijvoorbeeld om het gebruik van ‘*manquer de*’. Een ander voorbeeld waarvoor *Lexogoth* nuttig kan zijn, is het probleem van de ‘verbindingswoorden’ (bijwoorden, voegwoorden) die gebruikt worden om zinnen of stukken tekst met elkaar (betekenisvol) te verbinden. Bij het instuderen van bijvoorbeeld de woorden ‘*toutefois, cependant, pourtant*’ zal menig student zich afvragen waar in de zin deze woorden zoal kunnen geplaatst worden. Bij gebrek aan voldoende zinvolle (en correcte) voorbeelden zal een scrupuleuze student zich

misschien beperken tot de aangereikte voorbeeldzin uit het handboek en het woord in kwestie systematisch op slechts één plaats zetten terwijl een meer avontuurlijke student misschien meent dat deze woorden dezelfde posities kunnen aannemen als in zijn eigen moedertaal, hetgeen een begrijpelijke maar potentieel foutieve redenering is. Via *Lexogoth* krijgt een student nu toegang tot meerdere correcte voorbeelden die ingestudeerd kunnen worden en kan hij zijn idee over het correct gebruik van een woord *makkelijk* en *zelfstandig* bijsturen.

Het programma omvat verder nog een volledig overzicht van de leerstof die je kan terugvinden in een typische cursus Frans (als vreemde taal) voor het volledige niveau A van het Europees Referentiekader voor Talen (ERK, Niveaus A1 – A2). We bieden een overzicht aan van de voornaamste grammaticatopics, woordenschat en taalhandelingen die behandeld worden op dit niveau. Voor de componenten *Woordenschat* en *Taalhandelingen* beschikken we ook over een hele reeks (communicatieve) oefeningen maar die hebben we niet opgenomen in het programma. We willen nogmaals benadrukken dat dit tweede luik vooral een ondersteunende functie heeft en dient om de student snel bepaalde stukken leerstof op zelfstandige en efficiënte manier te laten herhalen of leren.

2. Hoe werkt het programma ?

2.1 Het doorzoeken van de lexicale database

Het primaire doel van het programma is om woorden in hun gebruik te bestuderen. Hoe kunnen we nu een woord of uitdrukking snel terugvinden tussen al deze voorbeelden ? Er zijn in het programma essentieel 3 **zoekmethoden** waarvan de eerste twee methoden via het zoekvenster bovenaan links verlopen [zie foto 1] :

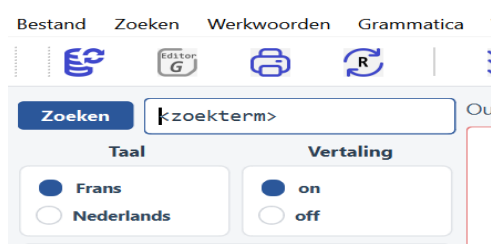


foto 1

* Methode 1 :

Je selecteert de taal waarin je een zoekopdracht wilt uitvoeren en je klikt aan of je een vertaling van de zinnen wenst of niet. Vervolgens typ je het woord in waarnaar je zoekt. Zodra je de eerste letters van je zoekterm intypt, wordt het wit vlak eronder gevuld door potentiële aanvullingen van je zoekterm (foto 2) :

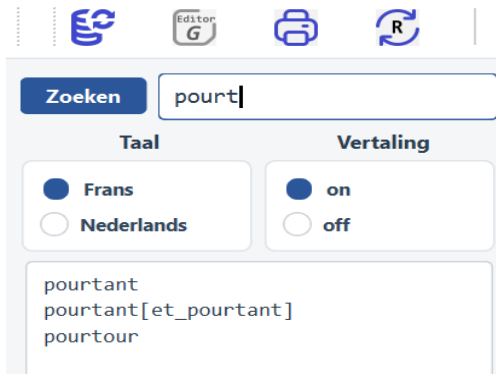
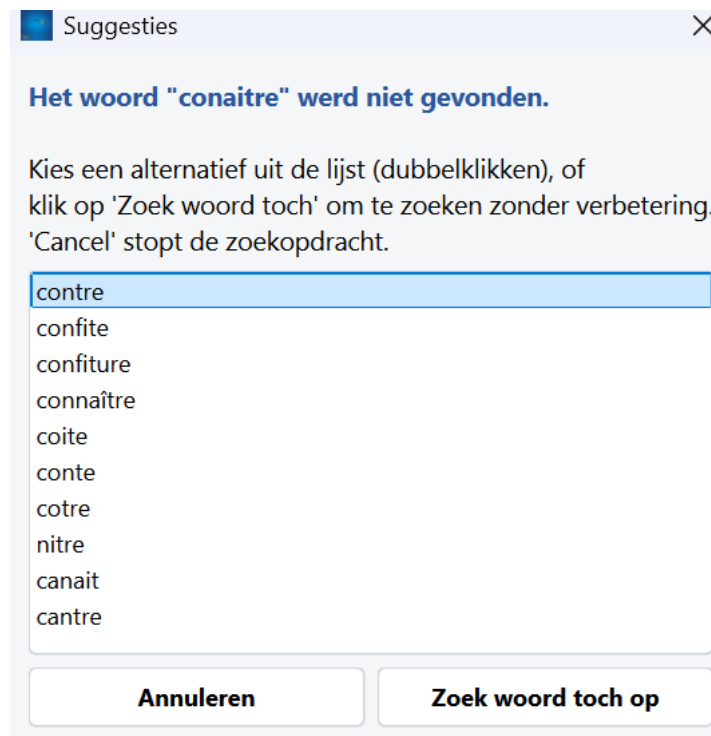


foto 2

Indien het woord dat je wil zoeken in de lijst verschijnt dan kan je hierop dubbelklikken en dan verschijnen de oplossingen in het hoofdvenster waarbij het gezochte woord en de daarvan afgeleide vormen in het blauw worden getoond.

Als je een woord invoert dat niet wordt herkend, verschijnt er een popup die andere woorden voorstelt die lijken op het getypte woord (berekend via Levenshtein-distance). Als je bijvoorbeeld “*probleme*” invoert (zonder accent), verschijnt er een vak waarmee je je fout kunt corrigeren door een woord uit de voorgestelde lijst te kiezen. Als je de suggesties wilt negeren (bijvoorbeeld wanneer je naar een eigenaam zocht), druk dan gewoon op “Search anyway”:



Let op als dit niet je eerste zoekopdracht is: sommige zoekopdrachten leveren soms tot meer dan 100 bladzijdenresultaten en als je als zo enkele zoekopdrachten hebt uitgevoerd dan kan het zijn dat de interface minder responsief wordt als je aan het resizen gaat. Maak daarom van tijd tot tijd het scherm leeg via de **refresh-knop** (icoon met letter R in cirkel) wanneer je een nieuwe zoekopdracht lanceert.

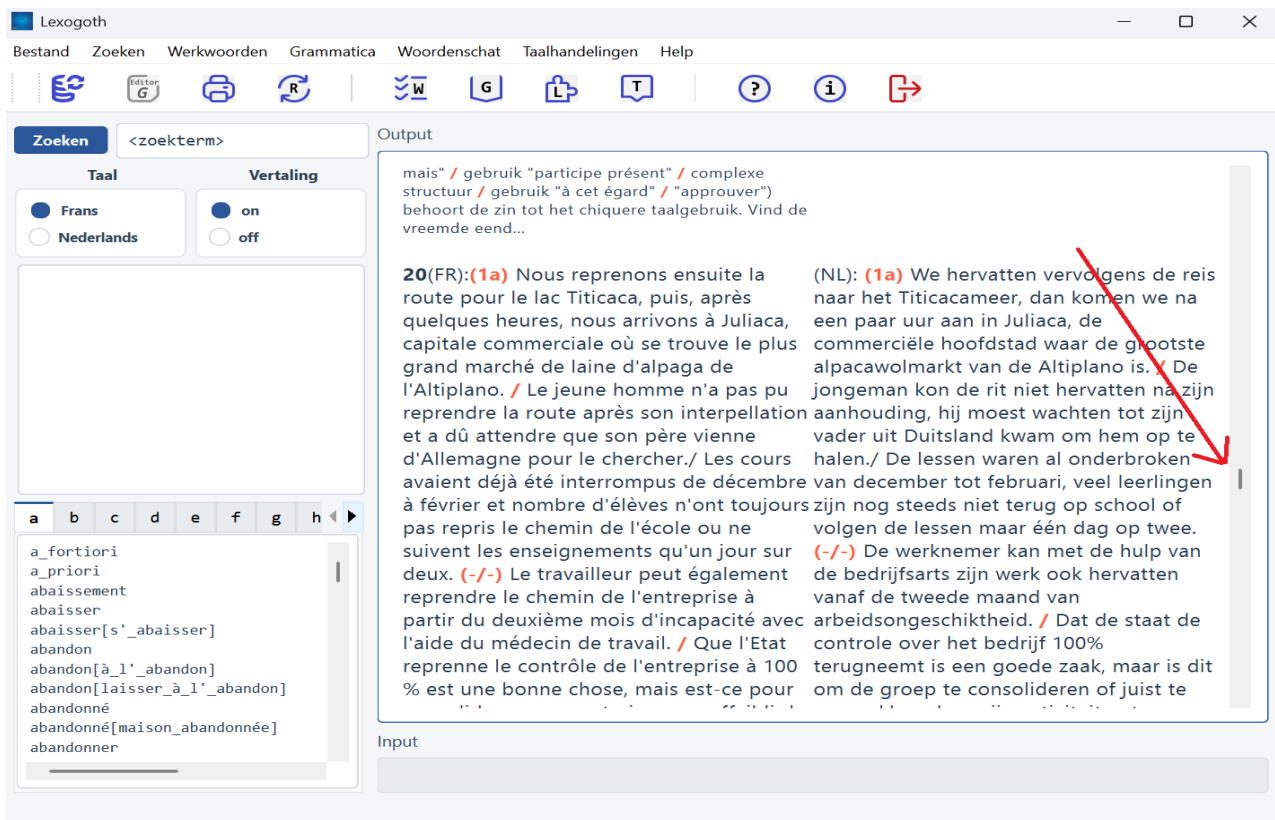


foto 3

* methode 2

Je kan er ook voor kiezen om de *autocomplete*-suggesties te negeren en je zoekopdracht te bevestigen door op de knop **Zoeken** te drukken of door gewoon de *enter*-toets in te drukken. Indien je op dergelijke wijze het corpus doorzoekt, zal **Lexigoth twee zoekresultaten** weergeven : Het eerste resultaat retourneert (onder **zoekresultaat 1**), indien het woord gekend is, het basislemma van de zoekterm en de woordcategorie. Deze zoekopdracht wordt uitgevoerd op het woordcorpus dat een lijst bevat (Frans :ongeveer 565 000 woordvormen ; Nederlands : ongeveer 35 000 vormen) van Franse woorden. Dit is handig als je bijvoorbeeld snel wil nakijken in welke 'vorm' een woord staat : voor 'vieille' geeft het programma onder **resultaat 1** weer dat het een *adjectief in het vrouwelijk enkelvoud* is dat afgeleid is van het lemma 'vieux'. Daarna wordt **zoekresultaat**

2 weergegeven; dat is het zoekresultaat van het tekstcorpus waarbij de vormen van het gezochte woord in het blauw worden getoond:

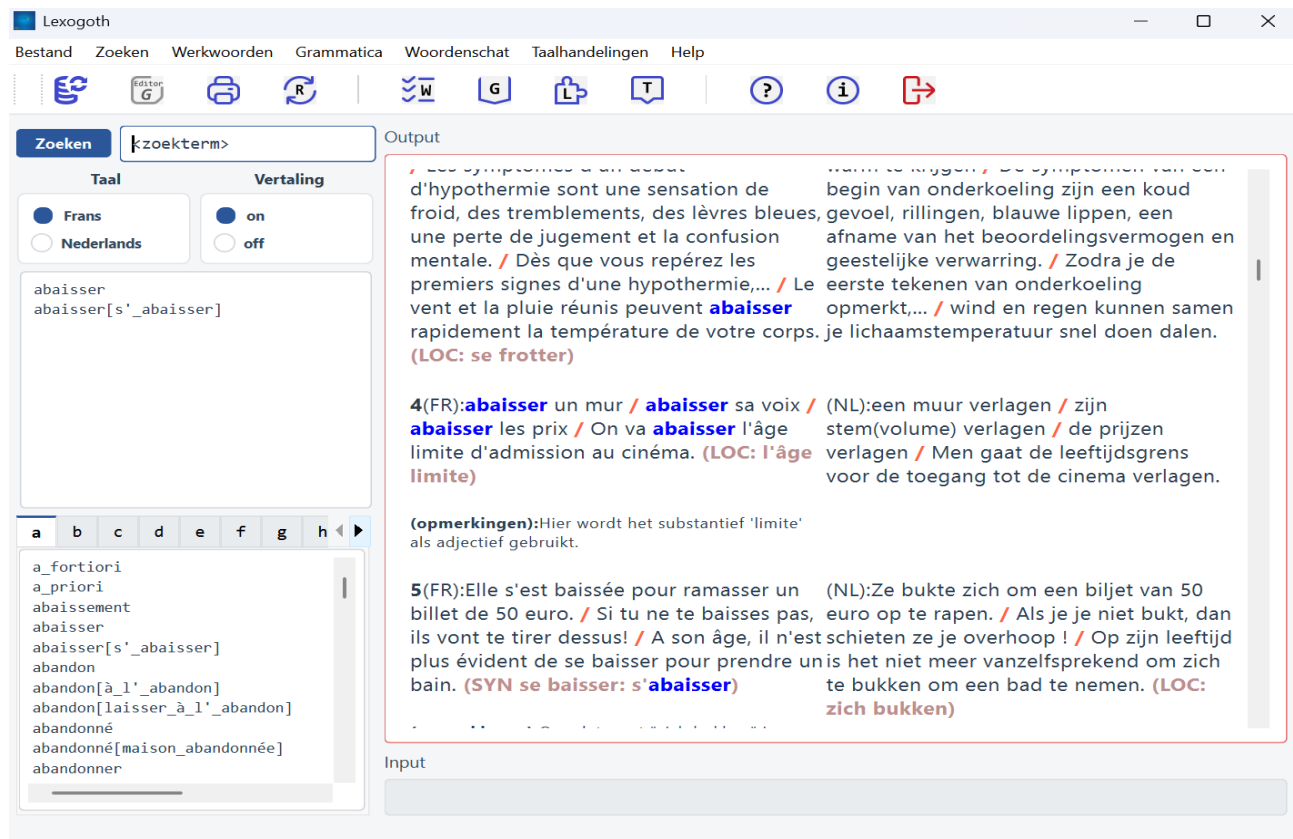


Foto 4

Je kan via deze methode ook zoeken naar een combinatie van meerdere termen tegelijk die samen in een zin verschijnen. Deze zoekwijze dient vooral om uitdrukkingen met meerdere woorden of bepaalde combinaties (bv. *film* en *cinéma*) terug te vinden. We geven twee voorbeelden die het gebruik illustreren.

Stel dat we willen weten wat 'een film op televisie uitzenden, geven,...' is. Dan kunnen we proberen of we ergens de combinatie van de woorden 'télévision' en 'film' kunnen terugvinden in het corpus. Je typt dan beide woorden in - gescheiden door een witruimte !- en drukt vervolgens op *zoeken*. Indien de combinatie ergens voorkomt in het corpus, dan wordt dit resultaat afgedrukt in het hoofdscherm en kan je kijken of je vindt wat je zoekt. Let op, het is schier

onmogelijk om alle potentiële combinaties te geven en de kans is dus groot dat je zoekopdracht niets oplevert. Indien dat het geval is, zal je een ruimere zoekopdracht moeten lanceren of een zoekopdracht met maar slechts één zoekterm (bijvoorbeeld *film*) om daarna alle voorbeelden te analyseren. Een voorbeeld voor de zoekopdracht met “film” en “cinéma”:

The screenshot shows the Lexigoth web interface. The search bar contains the text 'kzoekterm>'. The interface is in Dutch. The search results are displayed in a table with two columns: French (FR) and Dutch (NL). The results are numbered 12, 13, and 14. The results show various examples of the words 'film' and 'cinéma' used in different contexts, such as 'un film catastrophe', 'le nouveau film du réalisateur X', and 'le cinéma en France'. The interface also includes a sidebar with a search bar, a list of words, and a navigation menu.

Het is ook mogelijk om deze zoekwijze te gebruiken om uitdrukkingen, vaste woordcombinaties en zegswijzen terug te vinden. Stel even dat je wil zoeken naar de uitdrukking ‘*ne pas y aller par quatre chemins*’, dan kan het al voldoende zijn om de termen ‘*aller*’ en ‘*quatre*’ in te typen om het gewenste resultaat te verkrijgen. Ook hier is een positief zoekresultaat niet gegarandeerd omdat elke *entry* in het corpus kan bestaan uit meerdere chunks en het is best mogelijk dat de ene chunk ‘*quatre*’ bevat en een andere het werkwoord ‘*aller*’ zonder dat beide gecombineerd voorkomen !

Let wel op dat je deze zoekmethode niet gebruikt om naar samengestelde woorden te zoeken zoals *parce que* of *pour que*. Daarvoor gebruik je best methode 3. [Zie ook verderop : [FAQ >](#)

Zoeken naar een exacte match voor uitdrukkingen met meerdere woorden].

* methode 3

Je kan ten slotte ook via een klassiek woordenboek zoeken. Aan de linkerzijde van de *interface* vind je onderaan verschillende *tabs* met de letters van het alfabet : selecteer de gewenste taal, klikt op de beginletter van je zoekterm en scroll daarna door de lijst totdat je bij het gewenste woord uitkomt. Deze methode is handig als je bijvoorbeeld naar uitdrukkingen zoekt die uit meerdere woorden bestaan. Voor de woordcombinatie 'faire face à' ga je naar het woord 'face' en scroll je door totdat je bij 'face[faire_face_à]' uitkomt. Daarna dubbelklik je op je gewenste zoekterm zodat het resultaat op je scherm verschijnt [foto 5] :

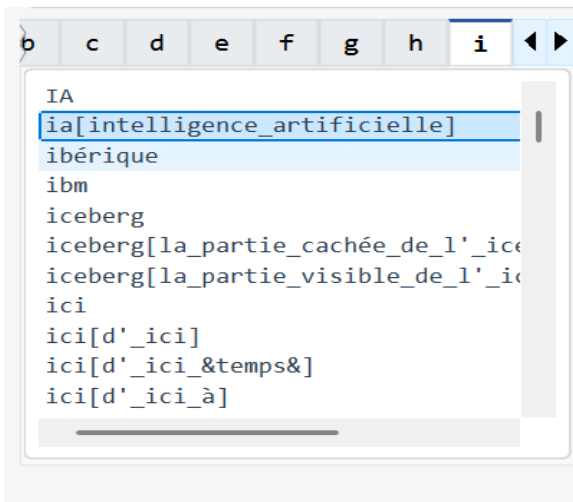


foto 5 : zoeken via woordenboek

2.2 Grammatica, woordenschat en taalhandelingen

In het programma kan je in de respectievelijke menu's *Grammatica*, *Woordenschat* en *Taalhandelingen* lessen terugvinden die de bijna alle topics van de ERK-niveaus A1 en A2 behandelen. Door het menu (foto 6) aan te klikken verschijnt er een submenu (foto 7) dat een overzicht biedt van de aangeboden lessen / uitleg. Selecteer een topic uit de lijst

en dubbelklik op het item om het te openen. De lessen worden geopend in de **externe** standaard pdf-viewer van het besturingssysteem waarop Lexogoth is geïnstalleerd. Mocht je PC deze bestanden niet vanzelf openen, dan moet je in de instellingen van je besturingssysteem een standaard programma selecteren als pdf-viewer. Voor de meesten zal dit de *Edge*-browser zijn.

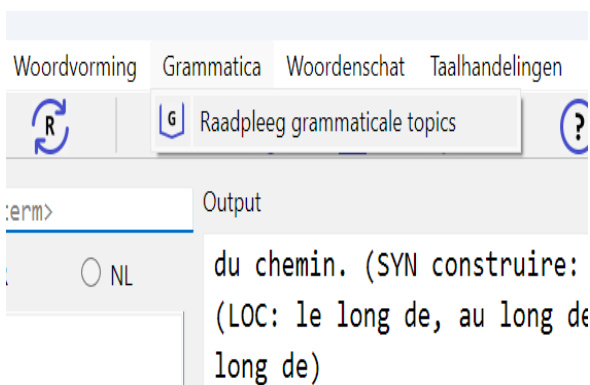


foto 6



foto 7

2.3 Werkwoorden vervoegen

Een pijnpunt voor menig student is de vervoeging van de Franse werkwoorden. Door op het menu *Werkwoorden* en het submenu *Toon de werkwoordvervoeging* te klikken kan je de volledige vervoeging van een Frans werkwoord laten weergeven. Let op, dit lukt alleen als het werkwoord gekend is in de database en je kan slechts de actieve vormen te zien krijgen. De *passé simple*, *passé antérieur*, *subjonctif imparfait* werden niet toegevoegd. De *passé simple* voegen we nog toe in een latere update. Let ook op de opmerkingen die worden gegeven onderaan de vervoeging van het gevraagde werkwoord ![foto's 8 en 9]

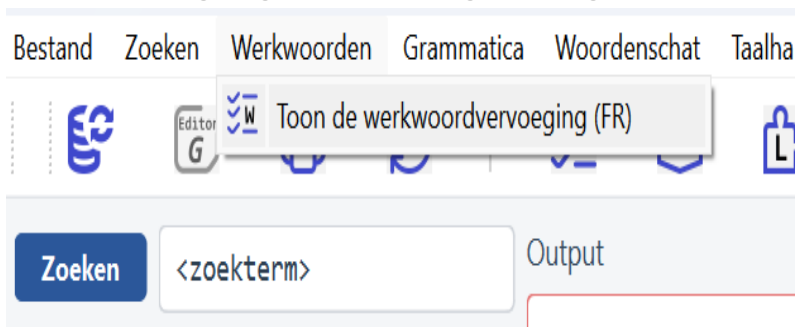


Foto 8: werkwoordvervoeging

Output

Infinitif :	<i>comprendre</i>
INDICATIF	
Indicatif présent :	<i>je comprends, tu comprends, il comprend, nous comprenons, vous comprenez, ils comprennent</i>
Indicatif imparfait :	<i>je comprenais, tu comprenais, il comprenait, nous comprenions, vous compreniez, ils comprenaient</i>
Indicatif futur simple :	<i>je comprendrai, tu comprendras, il comprendra, nous comprendrons, vous comprendrez, ils comprendront</i>
Indicatif passé composé :	<i>j'ai compris, tu as compris, il a compris, nous avons compris, vous avez compris, ils ont compris</i>
Indicatif pl-q-pf :	<i>j'avais compris, tu avais compris, il avait compris, nous avions compris, vous aviez compris, ils avaient compris</i>
Indicatif futur antérieur :	<i>j'aurai compris, tu auras compris, il aura compris, nous aurons compris, vous aurez compris, ils auront compris</i>
CONDITIONNEL	
Conditionnel présent :	<i>je comprendrais, tu comprendrais, il comprendrait, nous comprendrions, vous comprendriez, ils comprendraient</i>
Conditionnel passé :	<i>j'aurais compris, tu aurais compris, il aurait compris,</i>

foto 9 : resultaat vervoeging “comprendre”

Als je een werkwoord intypt en het wordt niet herkend, dan zal het zoekscherm verdwijnen en verschijnt de melding dat het werkwoord niet werd gevonden.

2.4 Het menu *Zoeken*: geavanceerde zoekmethoden

In het menu *zoeken* (foto 10) kan je verscheidene soorten van zoekopdrachten lanceren.

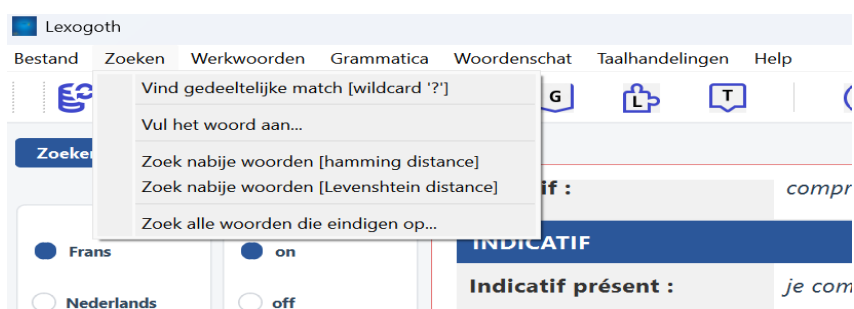


foto 10 : het menu *Zoeken*

Met ***Vind gedeeltelijke match*** kan je met *wildcards* (jokertekens ; hier is de *wildcard* het vraagteken) zoeken welke woorden voldoen aan de voorwaarden. Als je bijvoorbeeld zoekt naar ‘*compl??*’ dan krijg je het resultaat : *complet, complot, complus, complut*.

De opdracht **Vul het woord aan** spreekt voor zich : je typt het begin van een woord in (*prefix*) en het programma zoekt naar alle woorden die hieraan voldoen. Typ je bijvoorbeeld ‘complet’ in, dan krijg je als resultaat : *complet, complet-veston, complets...*

De opdracht **Zoek alle woorden die eindigen op** spreekt ook voor zich : hiermee kan je alle woorden die eindigen op een bepaalde lettercombinatie terugvinden (*suffix*). Nemen we bijvoorbeeld ‘ète’, dan krijgen we : *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* [+ nog meer]

De opdracht **Zoek nabije woorden [Hamming distance]** is een experimentele functie. Ze vindt woorden van dezelfde lengte die op een specifiek aantal letters na van elkaar verschillen. Door de afstand op 1 in te stellen, vind je woorden waarin slechts één letter is veranderd. Als je bijvoorbeeld zoekt naar ‘pars’ met een afstand van 1, krijg je resultaten als *bars, cars, gars, mars, pacs,....* Deze tool helpt je om gelijkaardige woordenschat te ontdekken op basis van het aantal vereiste letteraanpassingen. Houd er rekening mee dat deze berekening alleen van toepassing is op woorden van gelijke lengte. Het gebruik van deze functie maakt geavanceerde patroonherkenning binnen de database mogelijk.

De opdracht **Zoek alle nabije woorden [Levenshtein distance]** is eveneens een experimentele functie. Het vindt woorden van dezelfde lengte die op een specifiek aantal letters na van elkaar verschillen. Door de afstand op 1 in te stellen, vind je woorden waarin slechts één letter is veranderd. Als je bijvoorbeeld zoekt naar ‘pars’ met een afstand van 1, krijg je resultaten als *bars, cars, gars, mars, pacs,....* Deze tool helpt je om gelijkaardige woordenschat te ontdekken op basis van het aantal vereiste letteraanpassingen. Houd er rekening

mee dat deze berekening alleen van toepassing is op woorden van gelijke lengte. Het gebruik van deze functie maakt geavanceerde patroonherkenning binnen de database mogelijk.

2.5 Het menu *Help*

Hier vind je de gebruikelijke *help*-topics in de submenu's : *tutorial* (het document dat je nu leest), *Links* (verwijzingen naar online woordenboeken en grammatica's) en *About*.

2.6 Het menu *Bestand*

Dit menu bevat de mogelijkheid om het programma af te sluiten (*Sluit programma*), het output-scherf af te drukken (*Print huidige output*). Er was in de vorige versies ook een minimale tekst-editor *Golgoth Editor* voorzien. Deze editor werd op inactief gezet en is niet meer toegankelijk vanaf versie 4.6 (huidige) omdat de meeste Windows-computers al over geavanceerde editors beschikken (zie Notepad en het te downloaden Notepad++) en de Golgoth-editor hier geen meerwaarde meer biedt.

2.7 Een klein maar belangrijk icoon : het *refresh*-icoon.

Zodra je enkele zoekopdrachten hebt uitgevoerd kunnen de resultaten snel oplopen tot in de honderden. Daarom raden we de gebruikers aan om regelmatig het outputscherf leeg te maken door het *refresh*-icoon [foto 11] aan te klikken. Als je dit regelmatig doet, kan je veel comfortabeler leren, het programma makkelijk *resizen* en bovendien druk je niet ongewild honderden pagina's af, als je iets wenst af te drukken.

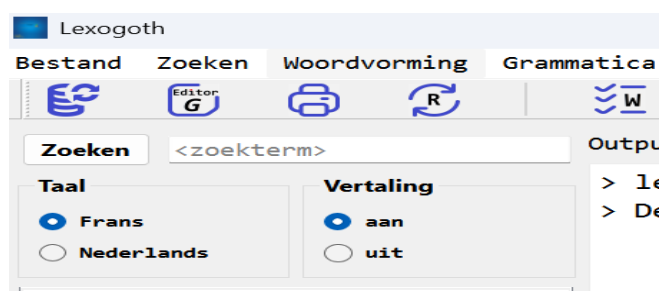


Foto 11 : (iconen van links naar rechts : open *Lexicale Toolbox*, open *Golgoth editor*, afdrukken output-scherm, **scherm verversen (Leegmaken)** , vervoeg werkwoord)

2.8 De Lexicale Toolbox van Lexogoth

In illustratie 11 (hierboven) zal een klik op het eerste icoon de Lexicale Toolbox openen. Deze module is nieuw en wordt vanaf Lexogoth v4.5 uitgerold. De *Lexicale Toolbox* van Lexogoth is een module voor studenten en docenten Frans, ontworpen om woordenschat en lexicale patronen te oefenen met behulp van authentieke corpusdata. Het programma is opgebouwd rond drie hoofdmodules:

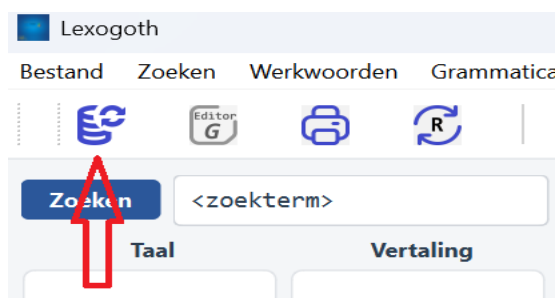
(1) Databasefuncties: Hiermee kunnen gebruikers lexicale databases creëren , bestaande databases bewerken, en de databases inoefenen door de Nederlandse of Franse vertaling te verbergen. Hierbij komen ook enkele belangrijke corpus-tools zoals een n-gram analyse, geavanceerde zoekopties, zoeken naar woordcombinaties,...

(2) Importeren / exporteren: Deze module vormt het hart van de Toolbox en staat gebruikers toe om voorbeelden (entries) rechtstreeks uit Lexogoth te importeren en te filteren met geavanceerde zoekopties (*AND/OR/Exact string match/Proximity Search*). Het maakt ook exporteren mogelijk van databases, testen en testresultaten naar een bewerkbaar RTF-tekstbestand.

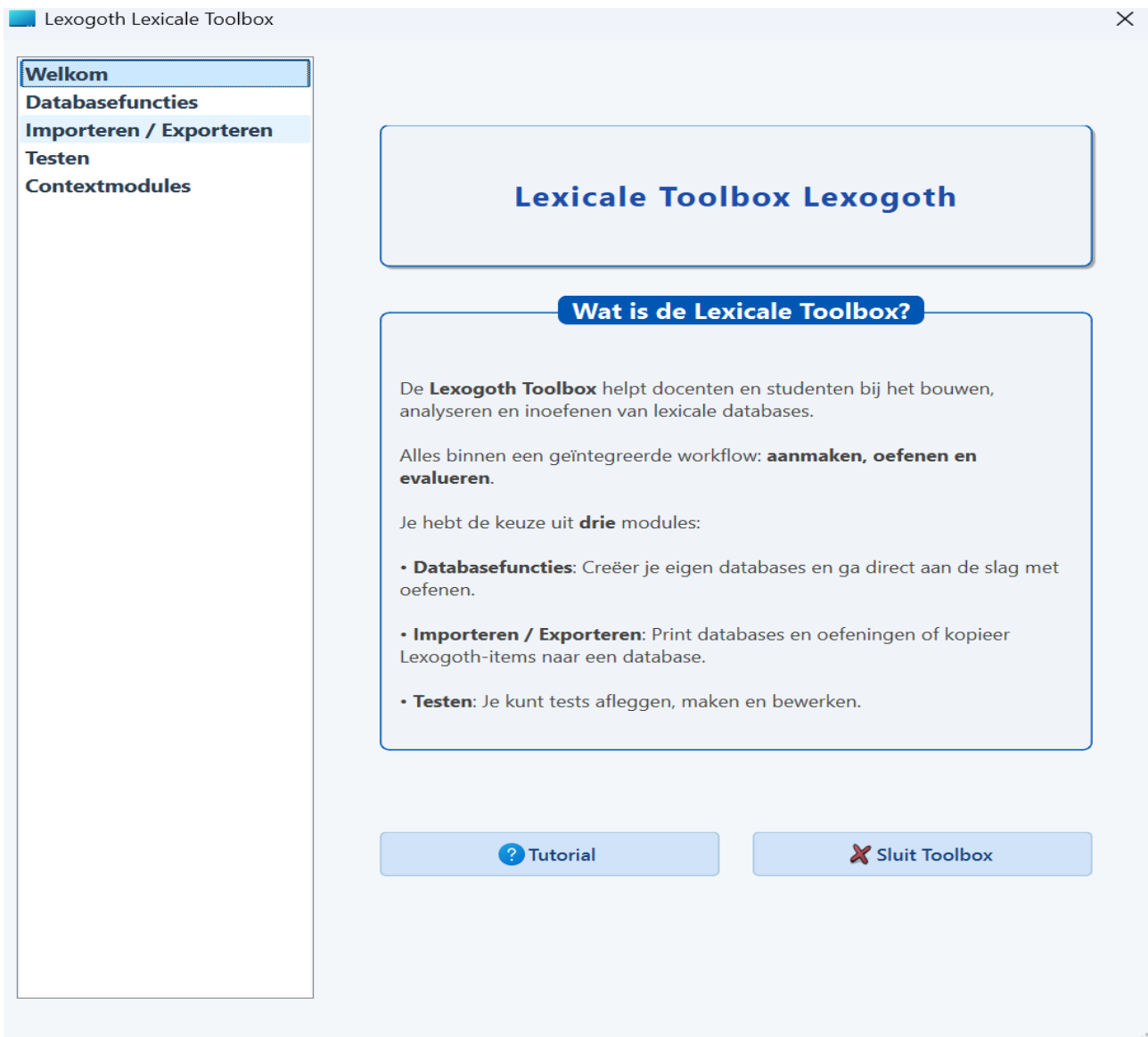
(3) Testen: Gebruikers kunnen diverse soorten testen opstellen en afleggen, waaronder *multiple choice*, *matching*, *exacte invulvragen* en *dropdown*-vragen.

Voor meer informatie verwijzen we de gebruiker naar de handleiding die hij kan raadplegen in het opstartscherm van de *Toolbox*.

Je opent de Toolbox door op het eerste icoon te klikken (boven rode pijl):



Een illustratie van het openingsscherm:



Je kan de uitgebreide handleiding terugvinden op de website <https://lexogoth.be>, onder het menu *Lexicale Toolbox*.

3. Hoe gebruik je het programma als student of leerkracht ?

-Je kan *als student* Lexogoth op de volgende manieren gebruiken:

(1) indien je een spreek- of schrijfoefening moet voorbereiden rond een bepaald thema, dan kan je ter ondersteuning enkele zoekopdrachten lanceren met kernwoorden van het thema waarrond je werkt. Met de geretourneerde zinnen kan je dan vervolgens aan de slag gaan om je teksten op te bouwen. De resultaten bieden je zowel taalkundige als inhoudelijke hulp. Taalkundig, daarmee bedoelen we dat je een overzicht krijgt van de mogelijke constructies van de woorden die je kan gebruiken; de thematische ondersteuning vind je terug in de voorbeeldzinnen die het lexicaal veld waarbinnen je werkt uitbreiden. Stel dat je rond "réchauffement climatique" wil werken, dan kan je met de resultaten voor o.m. de zoektermen "réchauffement (climatique)", "pollution", "température", "santé", "problème" al een heel eind komen. Voor een handigere oplossing hiervoor, zie *Lexicale Toolbox*.

(2) Je kan het programma ook gebruiken bij het studeren van woordenschat: je kan ter ondersteuning bij het studeren van moeilijke woorden (zoals connectoren en werkwoorden) zelf zinnen gaan opzoeken die het correcte gebruik van dit woord weergeven. Een dergelijke manier van studeren zal leiden tot een veel diepgaandere kennis van de taal.

(3) Bovendien is het ook mogelijk om Lexogoth als een soort lees- en leerboek te gebruiken. De taalleerder die gedurende een langere periode zijn kennis van het Frans wil uitbreiden, kan elke dag enkele momenten uittrekken om enkele woorden te verkennen. Doordat dezelfde woorden regelmatig terugkeren, zal hij zich die al lezend eigen maken. De lezer zal dus gaandeweg steeds meer affiniteit krijgen met de Franse taal en ook zijn uitdrukkingsmogelijkheden zullen steeds uitgebreider worden.

(4) Je kan je kennis van bepaalde taaltopics, die in het ERK-niveau A (en een stuk B) aan bod komen, op zelfstandige basis opfrissen door de talrijke lessen 'grammatica, woordenschat en taalhandelingen' die aanwezig zijn in het programma.

(5) In het programma is ook een module ingebouwd die de vervoegingen van de meeste Franse werkwoorden weergeeft; bij twijfel, kan je steeds met een druk op de knop de vervoeging van het gewenste werkwoord oproepen.

(6) Bij het werken aan korte taken in de les kan je ook de Lexogoth-editor gebruiken met de autocomplete-functie voor het Frans; het gebruik hiervan kan je helpen bij het juist schrijven van moeilijke woorden.

(7) Je kan zelfs databases of contextmodules samenstellen van de woordenschat die je moet studeren via de **Toolbox**. Verder kan je ook toetsen opstellen met verschillende soorten vragen om je kennis up-to-date te houden of te testen.

(8) Interactieve toetsen afleggen: Je kunt toetsen maken met verschillende vraagtypen, zoals meerkeuzevragen, exact invullen, koppelen en vervolgkeuzemenu's.

(9) Je leermateriaal verfijnen: Gebruik tijdens het importeren zoekfilters om je te richten op specifieke woorden of grammaticale patronen om ervoor te zorgen dat je studiedatabases relevant zijn.

(10) Exporteren voor offline studie: Je kunt je databases, toetsen en toetsresultaten converteren naar leesvriendelijke RTF-tekstbestanden om ze te bewerken of af te drukken met standaard tekstverwerkingsprogramma's.

- **Als Leerkracht** kan je als volgt aan de slag met Lexogoth :

(1) Je kan het programma tijdens de les inzetten als hulpmiddel om aan studenten bepaalde lexicale problemen uit te leggen; zo kan je bijvoorbeeld het verschil tussen "*prêter*" en "*emprunter*" makkelijk aantonen met de talrijke voorbeeldzinnen die aanwezig zijn. Het programma is ook erg interessant om de moeilijkere constructies van sommige woorden bloot te leggen aan de hand van enkele voorbeelden.

(2) De in het programma aanwezige voorbeeldzinnen kunnen ook gebruikt worden bij het opstellen van toetsen of bij het

voorbereiden van lessen. Zo kan de leerkracht, ter voorbereiding van spreek- of schrijf oefeningen, op voorhand een lijst van sleutelwoorden en representatieve voorbeelden voor zijn studenten maken zodat ze de oefeningen makkelijker tot een goed einde kunnen brengen. (zie voorbeeld "*réchauffement*" hierboven).

(3) Lexogoth kan ook dienen als een makkelijke en snelle tool om taalproblemen op te zoeken (synoniemen, gebruik sommige complexe constructies, nuttige voorbeelden, betekenissen, vertalingen, taalregisters,...) waarvoor een snel antwoord nodig is tijdens een les.

(4) De leerkracht kan Lexogoth ook gebruiken om remediërings-trajecten uit te tekenen voor studenten die wat achterstand hebben opgelopen: de voornaamste grammatica- en woordenschat topics en talrijke taalhandelingen voor het niveau A van het Europees Referentiekader werden in het programma opgenomen en indien een student problemen heeft met een van deze topics, dan kan de leerkracht de student alvast op goede weg helpen door hem de relevante hoofdstukken "grammatica, woordenschat of taalhandelingen" te laten herhalen via het programma.

(5) De leerkracht kan voor z'n studenten thematische databases of meer grammatica-geïnspireerde databases samenstellen met aandachtspunten. Verder kan hij ook (gescoorde) toetsen opstellen op basis van de in Lexogoth gevonden voorbeelden.

(6) Content cureren met geavanceerd zoeken: Gebruik de gespecialiseerde zoekfilters "EN", "OF", "Exacte overeenkomst", "Pattern Search" (morfologie) of "Syntactische patronen" om snel specifieke grammaticale patronen of thematische woordenschat uit grote geïmporteerde tekstblokken te extraheren.

(7) Verschillende testformaten ontwerpen: Je kunt tests maken en bewerken met meerdere vraagtypen, waaronder meerkeuzevragen, exact invullen, koppelen en vervolgkeuzemenu's.

(8) Materialen beheren en verfijnen: Voeg eenvoudig verschillende databases samen, verwijder dubbele items en rangschik items op lengte om inhoud van eenvoudig naar complex te organiseren.

(9) Printbaar materiaal genereren: Exporteer databases, tests en resultaten naar bewerkbare RTF-bestanden (Rich Text Format) om hand-outs of gedrukte examens te maken.

(10) Prestaties van leerlingen monitoren en beoordelen: Laad opgeslagen testresultaten (*.lexresult-bestanden) om een gedetailleerd tabeloverzicht te bekijken van scores, gegeven antwoorden en juiste antwoorden voor elke leerling.

4. Methodologie

In dit hoofdstuk leggen we uit hoe Lexogoth is ontwikkeld, welke taalkundige principes eraan ten grondslag liggen en hoe we een rijk corpus van authentiek Frans hebben samengesteld. Deze tekst biedt taaldocenten, specialisten in vreemdetalenonderwijs en nieuwsgierige gebruikers een dieper inzicht in de visie en aanpak achter Lexogoth.

4.1 Ontstaan van Lexogoth en primaire doelstelling

Lexogoth is ontstaan uit een frustratie die elke vreemdetalenaar al ondervonden heeft wanneer hij woordenschat moet studeren: het vinden (zoeken naar...) de juiste woordcombinaties en de vaststelling dat het leren van woordenlijsten onvoldoende is om vlot aan communicatieve oefeningen deel te nemen.

Deze moeizame overgang is volgens ons te wijten aan een gebrekkige studiemethode en inoefening voor woordenschat: Studenten vreemde talen beginnen bijna allemaal met het studeren van woordenschat via lijsten aan het begin van hun leertraject om op snelle en efficiënte wijze een kleine basiswoordenschat op te bouwen. Deze aanpak wordt in vele handboeken (Frans) gewoon verdergezet in de vervolgjaren van het leertraject, maar door deze piste te blijven bewandelen belast men de student met een serieuze handicap wanneer hij in een later stadium zelf

betekenisvolle taalhandelingen zal moeten stellen en zal vaststellen dat een combinatie van grammaticale regels en losse woorden onvoldoende zijn om taalhandelingen op een redelijke wijze uit te voeren.

Deze aanpak via woordenlijsten verhindert studenten immers om verbanden te leggen tussen de verschillende geleerde woorden en uitdrukkingen. Studenten leren met andere woorden een groot geheel van *gescheiden* lexicale eenheden zonder dat er een semantisch netwerk gecreëerd wordt dat al die lexicale elementen met elkaar verbindt. Dit gebrek aan interne samenhang leidt ook tot een oppervlakkige memorisering waardoor studenten de nieuw geleerde woorden sneller zullen vergeten dan wanneer ze stevig ingebed zijn in een netwerk. Woordenlijsten gaan ook vaak voorbij aan de complexiteit van een woord. Een woord is immers meer dan een definitie of een vertaling: om een woord goed te kunnen gebruiken moet men ook weten met welke complementen het wordt gebruikt, welke betekenis frequent is en welke niet, tot welk taalregister het behoort, in welke context het kan gebruikt worden, welke de synoniemen, antoniemen en afgeleide woorden zijn,... Daarbij komt nog dat de inoefening van de woordenschat in vele handboeken Frans ook te wensen overlaat: vaak worden er niet meer dan kruiswoordraadsels, invuloefeningen of aan te vullen lijsten aangeboden. Ruim onvoldoende om tot een gedegen lexicale kennis te komen... Door de verstarde aanpak van woordenschatonderwijs én het gebrek aan ruime, betekenisvolle input die de betekenis én het gebruik van nieuw geleerde woorden moeten illustreren, zullen studenten slechts moeizaam vorderen in hun kennis van de vreemde taal, veel moeizamer dan het zou moeten zijn. Voor enkele interessante werken over woordenschatverwerving en de hierboven aangekaarte problemen, zie: *Nation* (2001), *Thornbury* (2002), *Laufer* (2022), *Bentolila* (2010).

Volgens ons moet woordenschat steeds in context onderwezen worden en moeten studenten voldoende blootstelling krijgen aan talrijke voorbeelden om het gebruik van de nieuw geleerde woorden duidelijk te illustreren en om sneller een robuust

semantisch netwerk uit te bouwen. Het is vanuit deze visie dat we Lexogoth ontwikkelden: we willen studenten Frans een representatief corpus bieden met veelvoorkomende Franse woorden, uitdrukkingen en woordcombinaties in verschillende contexten. Dat stelt hen in staat om deze lexicale items efficiënter en sneller te onthouden en in te oefenen. We geven hieronder meer uitleg over de gebruikte methodologie en de totstandkoming van het corpus.

4.2 Samenstelling van het corpus: op basis van welke criteria werden woorden en structuren geselecteerd ?

4.2.1 Theoretische basis voor Lexogoth

De ontwikkeling van Lexogoth houdt nauw verband met drie evoluties binnen het vreemdetalenonderwijs die een serieuze opgang maakten in de jaren 1990:

- de *Lexical Approach* van Lewis (1993, “*The Lexical approach: The state of ELT and a way forward*”)
- het gebruik van corpora met authentiek materiaal ter ondersteuning van het leerproces (*corpuslinguïstiek*)
- *data-driven Learning* (vaak afgekort tot *DDL*): een didactische benadering waarbij studenten de rol van onderzoeker aannemen en zelf patronen en regelmaat ontdekken in grote hoeveelheden **authentieke** taaldata.

Laten we beginnen met de *Lexical approach* van Lewis. Een van de belangrijkste principes van deze benadering was dat, in de context van het vreemdetalenonderwijs, taal niet meer mag opgevat worden als een verzameling van zinnen die opgebouwd zijn uit grammaticale structuren waarin je naar believen lexicale items moet plaatsen (de traditionele grammatica-aanpak in het talenonderwijs), maar eerder moet gezien worden als een geheel van *ready-made* uitdrukkingen, samengestelde woorden en regelmatige woordcombinaties die in het geheugen van de

taalleerder opgeslagen worden en als dusdanig rechtstreeks én snel kunnen ingezet worden, weliswaar met minimale aanpassingen naargelang de gebruikssituatie.

Eenvoudiger gezegd vormt volgens Lewis dus niet de grammatica maar wel het netwerk van lexicale eenheden de basis van succesvol vreemdetalenonderwijs. Lexicale eenheden worden in deze context ook wel *chunks* genoemd en kan je definiëren als het geheel van individuele en samengestelde woorden, frequente woordcombinaties, uitdrukkingen en collocaties. Het idee van Lewis is dat docenten van vreemde talen vertrekken vanuit lexicale eenheden als fundament voor taalonderricht en de studenten voorzien van de nodige rijke en gevarieerde input door middel van authentieke teksten en gesprekken, en intensief gebruik van concordancers en woordenboeken.

De *Lexical approach*-aanpak met chunks biedt heel wat voordelen; hier zijn er enkele:

- Het gebruik van chunks verbetert de schrijfvaardigheid van vreemdetaallearners (2022,Albaqami,*The Role of Lexical Chunks in Promoting English Writing Competence among Foreign Language Learners in Saudi Arabia*).
- Native-speakers berusten ook voor een belangrijk deel van hun communicatie op vooraf geformuleerde chunks of zinnen (Pawley en Syder,1983, *Two puzzles for linguistic theory: Nativelike selection and nativelike fluency*).
- Er zijn aanwijzingen die aantonen dat studenten die chunks leren een grotere vlotheid aan de dag leggen en beter scoren op spreekvaardigheid terwijl de studenten die individuele woorden leren het moeilijker hebben om *on the spot* zinnen te vormen. (2006, Boers, Eyckmans, Kappel, Stengers, *Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test*)

Natuurlijk wekte Lewis' voorstel om taalonderwijs vooral te enten op een hoofdzakelijk lexicale basis ook heel wat kritiek op; een overzicht van enkele belangrijke punten:

- Op welke basis selecteert men woordcombinaties en uitdrukkingen en hoe kan men op basis van dergelijke elementen een cursus opstellen? En aangenomen dat men de meest frequente woorden selecteert, hoe kan men op basis daarvan een heel jaarprogramma opstellen waarin al het (authentiek) materiaal rond de aan te leren combinaties moet draaien? (Alison Wray, 2002, *Formulaic Language and the Lexicon*)

- Het is relatief makkelijk en overzichtelijk om een groot aantal woordenlijsten te leren en deze in het langetermijngeheugen te stoppen maar wat doet men met woordcombinaties waarbij studenten geacht worden om voor één woord tientallen courante combinaties te studeren? Dat lijkt ons praktisch en menselijk niet haalbaar omdat het buitensporig veel tijd en energie zou vragen van studenten. (Alison Wray, 2002)

- Wat doen we met de grammatica? Grammatica, meer bepaald de syntaxis, bepaalt hoe we correcte zinnen bouwen. Waarom moet dat naar de achtergrond verdwijnen terwijl het toch een van de belangrijkste mechanismen biedt om *nieuwe* zinnen te creëren? (Rod Ellis, 2003, *Task-Based Language Learning and Teaching*)

Een tweede belangrijke ontwikkeling is die van de corpuslinguïstiek, een taalkundige discipline die ontstond in de jaren 1960 en die (meestal) via computer systematisch grote hoeveelheden taaldata verzamelt en onderzoekt om tendenzen te vinden. Enkele bekende (Engelstalige) corpora die je online kan vinden, zijn onder meer *The Corpus of Contemporary American English (COCA)* en *The British National Corpus (BNC)*. Door de steeds betere toegankelijkheid van computers in de jaren 1990 kregen meer mensen, onder wie ook taaldidactici, toegang tot de verschillende corpora die voorheen alleen maar binnen onderzoeksinstellingen beschikbaar waren. Op die manier kon men volop beginnen na te denken over een eventuele inzet van tekstcorpora in het vreemdetalenonderwijs. Vanaf de jaren 2000 volgden verschillende initiatieven en pogingen om authentieke tekstcorpora tot bij de studenten te brengen en te integreren in het lesgebeuren door de creatie van onder meer elektronische

woordenboeken die worden aangevuld met corpusmateriaal, concordancers om woordgebruik / syntactische constructies te bestuderen, syllabi met uitdrukkingen en woordcombinaties die gebaseerd zijn op hun frequentie,...

Een derde belangrijke ontwikkeling waarop Lexogoth steunt is ten slotte de didactische stroming *data-driven Learning (DDL)*. In hun paper (2024) waarin ze een stand-van-zaken opmaken over *DDL* omschrijven Boulton en Pascual Pérez-Paredes *DDL* als een vorm van inductief leren waarbij de student optreedt als een taalonderzoeker. In plaats van passief regels te consumeren, analyseert de leerling authentieke corpusdata om patronen en registerverschillen zelfstandig te 'ontdekken'. De auteurs benadrukken echter dat puur corpusonderzoek voor studenten vaak te complex is door de enorme hoeveelheid ruwe data waarbinnen een vreemdetaalleerder maar erg moeilijk zelf de 'regel' zal kunnen ontdekken. Je kan immers niet verwachten van een student dat hij op basis van tientallen of honderden zinnen zelf de verschillende betekenissen van een complex polysemisch woord als "reprendre" louter en alleen op basis van voorbeeldzinnen kan vinden. Dat zou tijdrovend, frustrerend en cognitief erg belastend zijn en... het leerrendement zou waarschijnlijk verwaarloosbaar zijn. Net hier slaat Lexogoth de brug tussen theorie en praktijk door gebruik te maken van *pedagogically-mediated DDL*: de data zijn authentiek maar didactisch bewerkt én voorzien van uitgebreide verklaringen, waardoor de technische drempel wordt verlaagd en de focus volledig op de lexicale verwerving komt te liggen, hetgeen aansluit bij de recente ontwikkelingen in de *DDL*-wereld waar reeds lang vraag is naar tools en corpora die makkelijk te gebruiken zijn voor (vreemde)taalleerders die geen onderzoekers of studenten taalkunde zijn. Dit is volgens ons dé manier om grote tekstcorpora met authentieke data toegankelijk en nuttig te maken voor taalleerders van verschillende niveaus.

Deze drie ontwikkelingen, de *Lexical Approach*, *corpuslinguïstiek* en *data-driven Learning* zijn de methodologische en theoretische basis geweest waarop we ons geïnspireerd hebben.

We vertrekken vanuit het idee dat, zodra een student een zeker taalniveau heeft bereikt, taalverwerving, en meer bepaald verwerving van een vreemde taal, vooral moet draaien om het leren van authentieke en contextgebonden lexicale eenheden (frequente woordcombinaties, zinsstukken, vaste uitdrukkingen,...) die we zorgvuldig met behulp van tools uit de corpuslinguïstiek hebben samengesteld. Lexogoth combineert in één programma de vele voordelen van een concordancer, een woordenboek en een leerboek. Het programma verstrekt ook nuttig advies met betrekking tot het gebruik van woorden en woordcombinaties (frequentie, verheldering betekenissen, grammaticale problemen, register, uitspraak,...). De student moet door het gebruik van Lexogoth niet langer rekening houden met de nadelen die onherroepelijk verbonden zijn aan het individueel gebruik van deze tools:

- Concordancers geven zoekresultaten met voorbeeldzinnen die tot in de duizenden kunnen oplopen en waarin alle betekenissen van het gezochte lemma door elkaar staan. Als een student die een vreemde taal leert al die voorbeelden zelf moet doorlopen en op basis van woorddefinities uit woordenboeken al de verschillende gebruiken uit elkaar moet houden, zal het leerproces erg traag verlopen. In Lexogoth hebben we het zware werk voor de studenten al opgelost en de verschillende betekenissen duidelijk onderscheiden en voorzien van representatieve voorbeelden.

- De definities van online woordenboeken zijn in vele gevallen te abstract voor iemand die de taal leert en bieden weinig garantie op een correct gebruik van het te leren woord door een gebrek aan voldoende duidelijke voorbeelden. In Lexogoth bieden we duidelijke voorbeelden aan en voor vele woorden leggen we de definities uit in een aparte rubriek in het Nederlands zodat elke taalleerder zeker mee is.

- Concordancers zijn soms erg gespecialiseerde programma's waarvan het gebruik niet makkelijk kan aangeleerd worden. De gebruikersinterface in Lexogoth is erg eenvoudig en intuïtief.

- De 'oneindige' lijsten voorbeelden die door corpora worden afgeleverd zijn niet geschikt om als rechtstreeks studiemateriaal te dienen: te veel voorbeelden, geen melding van register, geen voorstellen voor synoniemen, geen parafrases, geen socioculturele uitleg bij bepaalde fenomenen (zoals *verlan*),...

We willen tot slot nog benadrukken dat Lexogoth **geen** lesmethode is zoals Lewis oorspronkelijk voor ogen had met de *Lexical Approach*, maar een **hulpmiddel** om studenten sneller frequente woordcombinaties en woorden in context aan te leren en een geheel van authentiek materiaal te geven waarmee ze systematisch en autonoom hun lexicale kennis kunnen inoefenen en herhalen.

4.2.2 Selectie en opbouw van het corpus

4.2.2.1 Een empirisch gebaseerde aanpak

Zoals we hierboven al vermeld hebben zijn verklarende (of vertalende) woordenboeken geen ideale tool voor vreemdetalende omdat er volgens ons te veel betekeniscomponenten van lemma's, die noodzakelijk zijn voor een goed gebruik, niet expliciet genoeg worden gesteld. We denken hierbij bijvoorbeeld aan de selectie van complementen door een werkwoord of het échte gebruik in de courante taal en niet het 'virtuele' gebruik uit woordenboeken of schoolboeken. Larousse geeft bijvoorbeeld voor de werkwoorden "frôler" en "friser" (*Le Robert en Ligne*) de volgende definities:

(a) [frôler] 1. Toucher légèrement et 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.

Afgaande op bovenstaande definities zal iemand die *frôler* en *friser* wil gebruiken, stellen dat beide woorden grotendeels als synoniemen kunnen beschouwd worden. Na verificatie in de gebruikte corpora stelden we vast dat *friser* en *frôler* in een figuurlijke betekenis (abstracte complementen) als synoniemen kunnen beschouwd worden maar dat ze dat in de letterlijke betekenis (dwz gevolgd door concrete complementen) dan weer niet zijn omdat we nauwelijks gebruiksvoorbeelden konden vinden waarbij *friser* wordt gebruikt. Hoewel de definities grotendeels identiek zijn mag je dus niet stellen dat hun gebruik dat in realiteit ook is. Zo zal je nauwelijks of niet zinnen als *un motard alcoolisé qui frisait dangereusement les passants dans le centre-ville de...* tegenkomen.

Om taalleerders voor dergelijke problemen te behoeden en vlotter woordbetekenissen te laten leren hebben we er dus voor gekozen om de woorden en structuren, en hun verschillende betekenissen, die we in ons corpus opnemen steeds te controleren op frequentie. We deden dit aan de hand van twee middelen:

(a) raadpleging van het corpus *French Web 2023* van *SketchEngine* (alle taalregisters)

(b) raadpleging van *RTBF*-corpus (Franstalig-Belgische televisie- en nieuwszender; standaardtaal), het corpus van *Le Monde* (Franstalige Franse krant; standaardtaal en verzorgd taalgebruik)

Daarna raadpleegden we de volgende referentiewerken om te verifiëren of het lemma of de woordcombinatie opgenomen was of niet en om de voornaamste betekenissen van het lemma in kaart te brengen:

- Le Robert (dictionnaire en ligne)

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

Na deze stap gingen we op zoek naar bruikbare voorbeelden voor de betekenissen die in de huidige Franse taal (ongeacht het register) worden gebruikt. We probeerden zo ruim mogelijk te zoeken naar bruikbare voorbeelden en raadpleegden de volgende middelen:

- het corpus *French Web 2023* van *SketchEngine*
- de website *www.Languefrançaise.net*, meer bepaald het algemene woordenboek en *Bob, l'autre trésor de La Langue* voor familiair of volks taalgebruik
- Franstalige Belgische media: *RTBF* (publieke omroep), Belgische kranten: *Le Soir*, *La Libre Belgique*, *La Dernière Heure*, Franse kranten: *Le Monde*, *Charlie Hebdo*, *Le Figaro*, *Le Parisien*, *Libération*, Zwitserse kranten: *24 heures*, *Le Matin*, *La Tribune de Genève*, Franstalige media uit Madagaskar, Canada, Congo (RDC), Haïti, Marokko en Algerije.
- Overheidswebsites van België en Frankrijk, in mindere mate Canada en Zwitserland; websites van bedrijven uit de Franstalige wereld; websites van internationale instellingen: Europese Unie, Verenigde Naties, NAVO, IMF,...
- Talrijke internetfora (vooral voor de familiale taal en spreektaal); enkele van de voornaamste bronnen: *JeuxVideo.com* (vooral jongerentaal), *Forum Santé - Doctissimo*, Franstalige *Reddit*-kanalen, *AlloCiné* (filmkritiek), *SensCritique* (filmkritiek). Hierbij horen ook de talrijke blogs gaande van kapselarrangementen en technologie tot internationale politiek (alle taalregisters).
- Franstalige ondertiteling van recente films en reeksen; we hebben erop gelet dat de ondertiteling niet gegenereerd werd met

AI maar door een Franstalige moedertaalspreker om de authenticiteit te garanderen.

- websites met taaladvies: de *Banque de dépannage Linguistique* (Canada), de rubriek *Dire et ne pas dire* van de *Académie*, de opmerkingen van *Larousse* (dictionnaire en ligne), *Projet Voltaire*, het forum van *Wordreference.com*, de contrastieve grammatica Nederlands-Frans *Grammar Warrior*

Vervolgens selecteerden we voorbeelden die regelmatig terugkwamen op verschillende plaatsen en namen we deze gedeeltelijk (als chunks) of volledig (als zinnen) op in Lexogoth. We zorgden ervoor dat de voorbeelden duidelijk de bestaande (abstracte) betekenissen illustreren en dat er bijvoorbeeld in het geval van werkwoorden een duidelijke selectie van de verwachte complementen werd gebruikt. Hier komen we terug op het voorbeeld van “*frôler*” om dit te illustreren:

...*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. ...*

Omdat niet alles door middel van voorbeelden kan gezegd worden, voegen we regelmatig uitgebreide opmerkingen met betrekking tot het register, de compatibiliteit met bepaalde complementen, synoniemen,... toe om de taalleerder nog beter tot een correct gebruik te brengen. Voor “*frôler / friser*” voegden we bijvoorbeeld deze informatie toe:

*(Opmerkingen): 1. In de figuurlijke betekenis van "net ontsnappen aan" of "grenzen aan" kan je "friser" ook als synoniem gebruiken: "friser La mort, la faillite, le burn-out, les 50 pour cent, ...". 2. In de letterlijke betekenis (rakelings langs iets scheren, lichtjes aanraken), waarbij het object van "frôler" [zoals in "La balle lui a **frôlé** le nez"] iets concreets is, lijkt de equivalentie tussen "frôler" en "friser" niet meer op te gaan, hoewel Larousse en de Académie beide woorden ook als synoniemen beschouwen in deze betekenis. In deze betekenis wordt het werkwoord "friser" niet of nauwelijks gebruikt; je zal dus niet tegenkomen: "...frisait les passants...", "...frisé le visage..."*

Indien een betekenis niet voorkwam in de woordenboeken (recente anglicismen, vulgaire taal, neologismen,...) maar wel een degelijke frequentie kende, kozen we er toch voor om het woord in ons corpus op te nemen. Als een woord of betekenis daarentegen wel door een of meerdere woordenboeken werd erkend maar niet of erg weinig in de geconsulteerde corpora verscheen, dan kozen we ervoor om het niet op te nemen.

Soms zagen we ons ook genoodzaakt om hier en daar de definities van de woordenboeken te comprimeren omdat de verschillen tussen sommige definities te fijnmazig waren om duidelijk onderscheiden te worden.

4.2.2.2 Structuur van het corpus

Het corpus in Lexogoth (juli 2025) beslaat twee delen: eerst hebben we de woordenschat van de belangrijkste lexicale velden voor de niveaus A1 en A2 van het Europees Referentiekader voor Talen (ERK) opgenomen als korte, bruikbare chunks (stukken zin of woordcombinaties) die specifiek gericht zijn op functionele taalvaardigheid voor beginners. Dit deel omvat grosso modo 15 procent van het totale corpus. Vervolgens hebben we verdergewerkt aan een uitdieping van deze woorden door hun belangrijkste en meest courante betekenissen en gebruiken toe te voegen. Dit eerste niveau omvat ongeveer 15 procent van het totale aantal voorbeelden.

Het tweede en grootste deel van het corpus bestaat uit een grondige uitdieping van de bestaande lemma's en hun gebruik in de familiale en courante dagelijkse taal, wat eerder het niveau B van het ERK bestrijkt. Een voorbeeld voor het werkwoord "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.

(Opmerkingen):1. "descendre l'escalier" of zonder complement "descendre" is "naar beneden komen via de trap". Deze constructie is de meest frequente. Merk op dat dit een transitief gebruik is van "descendre" en dat het hulpwerkwoord "avoir" is: "J'ai **descendu** l'**escalier**". Als "descendre" geen COD bij zich heeft, dan gebruik je "être": "Je suis **descendu**. / Elle est **descendue** par l'**escalier**. /....". Hetzelfde geldt voor "monter". 2. "descendre de..." wordt vooral gebruikt in de betekenis van "naar beneden komen van (een object) waarop je zit, staat of geklommen ben" zoals in "descendre d'un arbre, d'une échelle" of "uit, van een vervoersmiddel komen" zoals in "descendre de son vélo, du train,...". Het hulpwerkwoord is hier ook "être". 3. "descendre par l'**escalier**,...": als men specificeert hoe (via welke weg) men naar beneden is gekomen, dan is dat vaak om dit in contrast met een andere wijze te plaatsen: "on **descend** par l'**escalier** ou par l'ascenseur?". 4. "une mamie" of "mamy" betekenen "grootmoeder", "papy" of "papi" verwijzen naar een "grootvader". Voor het meervoud kan je aan de woorden een "s" toevoegen: des "mamies, mamys, papis, papys". 5. "quatre à quatre" betekent "halsoverkop, vliegensvlug", vaak gebruikt met werkwoorden die een beweging uitdrukken zoals "grimper, monter, **descendre**, franchir,...".

Voor dit tweede deel voegen we systematisch ook de betekenissen en het gebruik toe dat we tot het gevorderde niveau kunnen

rekenen (niveau C van het ERK); daarbij voorzien we regelmatig opmerkingen die de taalleerder moet helpen bij het juiste gebruik en instuderen van de nieuwe structuren. Nog een voorbeeld bij “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.

(Opmerkingen): ... 2. "en flèche" kan je als adjectiefgroep gebruiken ("une montée en flèche") of als bijwoordgroep ("monter en flèche") 3. "en flèche" wordt het vaakst gebruikt met werkwoorden van "stijgen, verhogen" zoals "monter, augmenter, grimper,..." en betekent dan "pijlsnel". Ook redelijk vaak gebruikt bij werkwoorden van de betekeniscluster "partir, démarrer,...". De combinatie wordt ten slotte ook gebruikt bij werkwoorden die een daling uitdrukken, zoals "baisser, descendre,...", weliswaar in veel mindere mate. Zie volgende opmerking echter. 4. Soms wordt "descendre en flèche" ten onrechte gebruikt in plaats van de familiaire uitdrukking "descendre en flammes", wat betekent "iemand hevig bekritisieren, afmaken, afbreken". Zo had men hierboven beter geschreven: "...de me faire descendre en flammes..." en "...va se faire descendre en flammes".

In het tweede deel brengen we ook systematisch woorden en structuren aan die frequent opduiken in de media, op internetfora en op overheidswebsites. Door blootstelling aan authentiek en actueel materiaal kunnen taalstudenten media en literatuur over belangrijke maatschappelijke kwesties begrijpen en krijgen ze ook de middelen om zich hierover zelf uit te drukken.

4.2.2.3 Structuur van de entries in Lexogoth

De manier waarop de entries in Lexogoth werden opgebouwd vormt de kern van onze innovatieve benadering: we combineren de functionaliteit van een woordenboek, concordancer én leerboek. Het was een doelbewuste keuze om lexicale precisie, authentiek gebruik en didactische inslag bij elkaar te brengen zodat studenten Frans efficiënter en sneller frequente lexicale combinaties (chunks) kunnen verwerven, door middel van slimme

herhaling of gerichte consultatie. Om die redenen zal de gebruiker *verschillende* soorten entries terugvinden in het corpus:

(a) Entries met exhaustieve behandeling van één lemma: deze entries sluiten het dichtst aan bij de woordenboekfunctie en illustreren de voornaamste gebruiken en betekenissen van een woord, vaak op overzichtelijke wijze geordend

(b) Entries rond een bepaald thema zoals fraude, klimaatopwarming, elektrisch rijden, ... Deze entries vormen een goede basis om een lexicaal netwerk aan te leggen en uit te breiden

(c) Entries rond een taalprobleem op grammaticaal, lexicaal of pragmatisch vlak en waarmee studenten Frans geconfronteerd worden bij het gebruik van Frans

(d) Eenvoudige entries die een opsomming van woorden geven of slechts gewijd zijn aan één uitdrukking of combinatie. Deze entries zijn er vooral op gericht om een basiswoordenschat uit te bouwen.

Dankzij de gevarieerde opbouw van de entries en corpus zal de gebruiker op een doeltreffende en flexibele manier nieuwe woordenschat en frequente lexicale patronen kunnen verwerven.

4.2.3 Taalregisters, geografische varianten en taalvariëteiten

We trachten in Lexogoth de diversiteit van het Frans zo breed mogelijk te belichten. Zo besteden we veel aandacht aan de geografische varianten van het Frans, meer bepaald aan de belgicismen. Belgicismen zijn Franse woorden, structuren,... die eigen zijn aan het Frans dat in België wordt gebruikt maar daarbuiten weinig verbreiding kent; aan de andere taalvariëteiten zoals het Frans van Canada, Zwitserland, Luxemburg, Congo, Algerije,... besteden we minder aandacht in ons werk. Anglicismen -dat zijn verfranste Engelstalige (vooral Amerikaanse) woorden en structuren- krijgen ook een ruime behandeling in ons werk en waar mogelijk geven we steeds synoniemen uit de standaardtaal. Hoewel de aandacht in Lexogoth

vooral uitgaat naar de standaardtaal besteden we veel ruimte aan de familiale taal, dat is de taal zoals Franssprekenden die in hun alledaagse omgang gebruiken; we vermelden steeds bij de opmerkingen of een uitdrukking tot de familiale taal behoort of niet en ook hier trachten we synoniemen of alternatieve formuleringen uit de standaardtaal te geven. Ook het vulgaire taalgebruik en het typisch Franse fenomeen van het *verlan* komen uitgebreid aan bod.

4.3 Hoe onderscheidt Lexogoth zich van andere taalapps en websites ?

Veel taalapps en taalwebsites richten zich op het snel en gemakkelijk aanleren van elementaire woordenschat door middel van de klassieke aanpak die gebaseerd is op geïsoleerde woorden, voorzien van een tekening, foto of vertaling. Hoewel deze aanpak nuttig kan zijn in de beginfase van het aanleren van een vreemde taal of bij het leren van gespecialiseerd vakjargon, schieten ze vaak tekort in het ontwikkelen van de actieve, productieve woordenschat die nodig is voor correct en vloeiend taalgebruik. Lexogoth is ontworpen om juist die kloof te overbruggen en onderscheidt zich op de volgende cruciale punten.

- **Authentieke context:** Vele populaire taalapps (Duolingo, Drops, Babbel,...) baseren hun woordenschatonderricht op het herhalen van losse woorden of erg eenvoudige zinnen, die vrij vaak onnatuurlijk aanvoelen (*I can fly* → ...). Het probleem hiervan is dat men aan de taalleerders niet toont hoe het woord in zijn natuurlijke omgeving, dwz in de échte geschreven en/of gesproken taal, wordt gebruikt: welke zijn de complementen die vaak samen verschijnen met het woord, wat is het taalregister waarin het wordt gebruikt, zijn er socioculturele elementen die het gebruik ervan beperken, welke betekenissen van een zijn frequent en welke niet,... Lexogoth gaat uit van het principe dat een woord pas gekend is wanneer een gebruiker weet hoe het in context kan gebruikt worden en tracht dit aan te leren door uitsluitend authentieke tekstfragmenten aan de gebruikers voor te leggen

waarbij we extra aandacht geven aan frequente collocaties en idiomatische uitdrukkingen.

- **relevante woorden en structuren:** Daar waar apps vaak werken rond thematische lijsten (*wonen, eten en drinken, zich verplaatsen, het dagelijkse leven,...*) die studenten moeten memoriseren, bestaat Lexogoth uit een zorgvuldig, met de hand samengesteld corpus dat woorden en woordcombinaties weergeeft die actueel en relevant zijn en die met een (relatief) grote frequentie gebruikt worden. Om dat te garanderen, consulteerden we bij de opname van lemma's steeds recente tekstcorpora.

- **diepe verankering van lexicale kennis:** Traditionele taalapps proberen hun gebruikers bij de les te houden door kortstondige sessies aan te bieden (Drops) waarbij het leren en oefenen van woordenschat als een spel wordt gepresenteerd (*streaks, badges, achievements,...*) door middel van eenvoudige invul-, vertaal- of uitspraak oefeningen. Lexogoth pakt het anders aan: de app bevat geen oefeningen in de klassieke zin van het woord. Het is de bedoeling dat gebruikers door herhaaldelijk en aandachtig lezen van de voorbeelden de patronen en correcte woordcombinaties diep verankeren. Dit proces, dat overigens sterk lijkt op de manier waarop moedertaalsprekers hun taal gebruiken (zie *Lexical Approach*, hierboven), zal ervoor zorgen dat de verworven kennis makkelijk opgehaald en gebruikt worden tijdens communicatieve situaties. Studenten zullen met andere woorden minder vaak moeten terugvallen op de tijdrovende en cognitief belastende woord-voor-woord zinsopbouw en zullen sneller en correcter kunnen reageren in communicatieve situaties.

- **aandacht voor diversiteit in taal: taalregisters, geografische varianten en socioculturele context.** De meeste apps schenken weinig tot geen aandacht aan de verschillende taalregisters die verbonden zijn aan het gebruik van woorden of specifieke betekenissen. Een gebrek aan aandacht voor deze nuances leidt tot ongepast of onnatuurlijk taalgebruik. Dezelfde opmerking geldt ook voor de geografische varianten van het Frans en de socioculturele componenten die aan een bepaald woordgebruik

verbonden zijn, zoals het geval is bij het verlan-taalgebruik. Lexogoth besteed aan deze diversiteit de nodige aandacht zodat de gebruiker niet alleen de correcte betekenissen van een woord gebruikt maar ook het woord in een passende context kan inzetten.

- geen AI-gegenereerd materiaal maar empirisch gevalideerde data: Talrijke taalleerapps en -methoden gebruiken generatieve AI zodat ze snel en goedkoop leermateriaal kunnen genereren. Dit is een slimme aanpak maar de kwaliteit van de gegenereerde zinnen is niet altijd even betrouwbaar omdat er regelmatig zinnen worden gecreëerd die onnatuurlijk zijn of die woordcombinaties gebruiken die weinig tot niet voorkomen. In Lexogoth zijn de voorbeelden niet gegenereerd door AI maar zorgvuldig, handmatig geselecteerd uit relevante corpora met recent taalgebruik van moedertaalsprekers. We moeten wel toevoegen dat we soms DeepL, ChatGPT en Gemini gebruikten om het vertaalproces naar het Nederlands te versnellen. De vertalingen werden echter handmatig gecontroleerd en bijgewerkt indien dat nodig bleek zodat we toch een hoge nauwkeurigheid voor de gebruiker kunnen garanderen.

4.4 Hoe studeren we nu best woordenschat en hoe kan Lexogoth hierbij helpen ?

Woordenschatverwerving is een lastig en soms pijnlijk proces voor taalleerders: het is immers niet voldoende om de vertaling van een woord te kennen maar je moet er ook nog een heleboel andere zaken bij leren zoals het type complement dat volgt op een werkwoord, het juiste voorzetsel na een substantief, het taalregister waartoe een betekenis behoort, de uitspraak,...

Dit leerproces wordt volgens ons door Lexogoth nu enigszins verlicht en verkort: het corpus voorziet voor vele courante Franse woorden en combinaties meerdere, duidelijke gebruikscontexten die de gebruiker veel tijd kunnen besparen (opzoeken in woordenboeken, doorgronden van moeilijke definitives,...) en die hem nu de middelen geeft om op

betrouwbare wijze systematisch woordenschat te leren én te herhalen.

De gebruiker mag wel niet in de val van het gemakzucht vervallen want efficiënte woordenschatverwerving moet ook gepaard gaan met kwalitatief onderricht en zeer gerichte productieve oefeningen (spreken, interactie, schrijven). Maar dat is natuurlijk de taak van de docent...

4.5 Een lopend project

Ondertussen (januari 2026) bevat Lexogoth al meer dan 40 000 aanklikbare items en we zijn nu bezig met het opvullen van de lacunes die (zeer zeker) nog aanwezig zijn in het werk. Dit proces zal zeker meerdere jaren in beslag nemen en we voorzien om de drie maanden een uitgebreide update.

4.6 Geciteerde werken (methodologie)

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.

Boulton, A. & Pascual Pérez-Paredes. (2024) *Data-driven Learning: Pedagogy and technology*. In R. Hampel & U. Stickler (Dir.), *Bloomsbury handbook of language learning and technology*. Bloomsbury Publishing.

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

- Kan ik als Franstalige ook Nederlands Leren met dit programma ?

Het is mogelijk om je Nederlands bij te spijkeren via dit programma, maar we raden dit niet aan. Het programma *Lexogoth* werd ontworpen voor Nederlandstaligen die Frans leren en niet andersom. Deze keuze impliceert dat bijna alle aandacht uitgaat naar Franse lexicale problemen en dat moeilijkheden uit de Nederlandse taal (voor bijvoorbeeld Franstaligen) niet of nauwelijks aan bod komen. Het Nederlands speelt hier een louter ondersteunende rol en wordt gebruikt om een redelijk deel van het Franse lexicon zo efficiënt en makkelijk mogelijk te laten verwerven door studenten; dat verklaart ook waarom de Nederlandse lijsten beduidend minder uitgebreid zijn dan de Franse, zo voorzagen we de Nederlandse lemma's van veel minder synoniemen en maakten we ook niet systematisch het onderscheid tussen de verschillende mogelijke etiketten voor dezelfde woorden ('drinken' als substantief en als werkwoord, bijvoorbeeld). Bovendien behoren het Nederlands en het Frans tot verschillende taalgroepen en zou een evenwaardige behandeling van beide talen in een programma een andere aanpak vereisen.

Toch kan een al redelijk gevorderde student Nederlands ook baat hebben bij het programma : zo voegden we de mogelijkheid toe om (alleen voor het Nederlands) de database te bevragen via de gescheiden vormen van scheidbare werkwoorden. Indien het programma het woord herkent, retourneert het de basisvorm en de voorbeelden waarin het gebruikt wordt. Als je bijvoorbeeld wil zoeken naar voorbeelden met het woord « afspreken » dan doe je dit natuurlijk in de eerste plaats door de infinitief (basislemma) in te geven maar je kan dit woord ook opzoeken via de termen « *spreek##af* », « *spraken##af* » [werkwoord + rest]. Je verbindt het werkwoord en de rest gewoon door twee *hashtags* (#) zodat het programma weet dat het om een samengesteld werkwoord gaat. Vervolgens retourneert het programma - indien de sequentie gekend is - de basisvorm van het werkwoord en de voorbeeldzinnen. Let op deze zoekoptie geldt alleen voor het Nederlands en lang niet alle scheidbare werkwoorden uit

het corpus werden al toegevoegd in hun gescheiden vorm. Deze zoekmethode kan nuttig zijn voor anderstaligen die Nederlands leren en wensen te weten of een werkwoord al dan niet scheidbaar is.

- *Als ik in de woordenlijst 'fait' aanklik, krijg ik ook alle zoekresultaten voor het werkwoord 'faire' te zien. Waarom is dat het geval en hoe kan ik dat vermijden?*

Het programma Lexogoth is gebaseerd op ongeveer 2500 bladzijden tekst, Nederlands en Frans gecombineerd. Bij het opstarten van het programma worden de teksten door het programma doorlopen en kent de 'tokenizer-lemmatizer' van het programma *alle mogelijke etiketten* toe die op een woord van toepassing zijn. Zo zal het woord 'fait' gekoppeld worden aan zowel 'substantif' (*un fait*) als aan 'verbe' (*faire*). Om deze reden zal je bij het aanklikken van 'fait' in de woordenlijst zowel de resultaten voor 'faire' als voor 'un fait' (en ook 'en fait', 'au fait',...) te zien krijgen. Voor woorden die niet zo frequent zijn, is het geen probleem om enkele voorbeelden te lezen om zo bij de gewenste vorm uit te komen maar voor erg frequente woorden is dit wel een probleem. Om de lezer hierin tegemoet te komen, hebben we manueel etiketten toegevoegd zodat hij makkelijker de gewenste categorie van een woord met meerdere etiketten kan terugvinden. Bijvoorbeeld, voor 'fait'/'faire' voegden we de volgende etiketten toe: 'fait[substantif], faire[verbe], fait[en_fait],...'. Op deze manier zal de lezer makkelijker zijn weg vinden tussen alle voorbeelden. Gebruik wel de visuele zoekmethoden om deze mogelijkheid te benutten.

- *Waarom vind ik niet zomaar alle samengestelde woorden terug zoals 'pomme de terre' of 'film d'horreur' ?*

In het Nederlands geldt de norm dat woorden aan elkaar geschreven worden en dus typografisch als een woord kunnen worden beschouwd. In het Frans worden samengestelde woorden het vaakst aangemaakt met voorzetsels, dat wil zegen dat één 'woord' vaak uit meerdere door een witruimte gescheiden woorden bestaat. Voor een tekstprogramma is dat natuurlijk een nachtmerrie. Want hoe maken we nu het onderscheid tussen een echt woord (zoals bijvoorbeeld *pomme de terre*) en een toevallige samenstelling? Het is niet altijd duidelijk waar je de grens moet trekken om een lettersequentie (of sequentie van klanken) te erkennen als woord: in het Nederlands zeggen we bv *aardappel* (1 aaneengesloten sequentie van letters) terwijl het Frans meer dan 3 'woorden' nodig heeft om hetzelfde object (de aardappel) te benoemen, namelijk *pomme de terre*. Stel dan even dat we *pomme de terre* erkennen als één woord (omdat bijvoorbeeld de sprekers van de taal dit woord niet meer analyseren als de som van 3

verschillende woorden, namelijk een *appel uit / van de grond*, maar als geheel), wat doen we dan met woorden als *film d'action / film d'aventures...* waarvoor de zaken al iets minder duidelijk zijn ? Moeten we die dan ook het statuut van 'woord' toekennen omdat ze net dezelfde vorm hebben als het aardappel-voorbeeld terwijl zij wel nog kunnen geanalyseerd worden op basis van hun componenten? De intuïties hierover zullen waarschijnlijk uiteenlopen en dat weerspiegelt zich natuurlijk ook in de zoekresultaten. Zo moesten we bijvoorbeeld beslissen welke woordgroepen een eenheid vormden en welke niet : in het voorbeeld hierboven verkozen we om *film d'action* als 3 losse woorden te zien terwijl onze *pomme de terre* als 1 woord wordt gezien.

- *Hoe zoek ik exact naar uitdrukkingen die uit meerdere woorden bestaan ? / Wat is de betekenis van het « liggend streepje » in woorden ? En waarom staat « parce que » wel als « parce_que » in de woordenlijst maar « pour que » als « pour[pour_que] » ?*

Het liggend streepje voegden we intern toe om aan te duiden dat de woorden bij elkaar horen en een eenheid vormen. Zo wordt 'parce_que' als één woord beschouwd terwijl de computer 'parce que' als twee woorden beschouwt. Als je echter in het woordenboek gaat kijken zal je daarin de sequentie 'parce_que' zonder problemen terugvinden maar niet 'pour_que'. Waarom is dat ? Om te vermijden dat het woordenboek onoverzichtelijk en slecht leesbaar zou worden door systematisch elke vaste uitdrukking als één string toe te voegen (bv : *à_bride_abattue*, en nog alle uitdrukkingen die met het voorzetsel 'à' beginnen) kozen we ervoor om de samengestelde uitdrukkingen en woorden toe te voegen op basis van de betekenisrijke woorden die deel uitmaken van de uitdrukking. In het voorbeeld hierboven zal je de uitdrukking dus terugvinden bij de lemma's 'bride' onder 'bride[à_bride_abattue]' en 'abattre[à_bride_abattue]' maar niet onder het betekenisloze voorzetsel 'à'. Maar waarom vinden we dan niet 'parce que' terug onder 'parce[parce que]' ? Omdat 'parce' geen geldig en bestaand lemma is in de Franse taal, vandaar zijn rechtstreekse toevoeging. Deze opmerking geldt vaak ook voor anglicismen. Samengevat : voor **samengestelde woorden zoek je best visueel** (woordenboek of autocomplete) met de kernwoorden waaruit de uitdrukking of het woord bestaat.

- *Welke zijn de gebruikte taalregisters ? Wat is het belang ervan ?*

Soms geven we, naar het voorbeeld van de geconsulteerde werken, mee tot welk taalregister of taalniveau een bepaald woord behoort. De

onderscheiden taalregisters zijn *familiaire taal*, *chique(re) taal*, *standaardtaal*, *vulgaire taal*, *volkse taal*. De meeste woorden en uitdrukkingen die we in ons lexicon opnamen, behoren tot de courante standaardtaal (*le registre courant*), dat wil zeggen dat dit woord gebruikt wordt in dagelijkse taal in vrijwel alle situaties. De familiale taal (*registre familier*) behoort tot de losse omgangstaal, dat is een taal die je meestal niet hanteert in formelere, serieuzere situaties. In het Nederlands spreekt men vaak over informele taal. Het chiquere taalgebruik (*le registre littéraire / soutenu*) of het formele taalgebruik kenmerkt zich door een gesofisticeerde taal en complexere, geraffineerdere uitdrukkingen. Dan hebben we nog het volkse taalniveau (*le registre populaire*), vaak een synoniem voor taal die wordt gebruikt in sociaal achtergestelde milieus. Het vulgaire taalgebruik (*le registre vulgaire*), wel ja, daar moeten we geen uitleg bij geven. Dit soort taal heeft slechts een erg beperkte inzetbaarheid en moet je voorzichtig hanteren. Sommige gesprekspartners zijn al wat gevoeliger dan anderen en zullen sneller aanstoot nemen aan het gebruik van vulgaire termen. We vermelden de registers om de taalleerder van een vreemde taal wat houvast te geven bij het verwerken en inzetten van de woorden en structuren in een vreemde taal ; bij het leren van de vreemde taal moet je het immers doen zonder de aangeleerde normen van je moedertaal en dan kan je al eens kemels schieten. Wees nu eerlijk, wat zouden jullie ervan denken mocht iemand aan jullie in gebroken Nederlands op de trein vragen of « hij zich naast jou mag neervlijen » ? Indien we geen taalregisters vermelden bij een woord kan je aannemen dat het woord tot de courante taal behoort. We hebben bij het behandelen van de verschillende betekenissen een bijzondere aandacht gegeven aan het familiale (informele) taalgebruik. Meer informatie over de taalregisters vind je terug in paragraaf 6.

- Wat betekenen de afkortingen SYN, LOC, EXP, CONT, ORT, PROV ?

- **CONT(A : B)** : dit betekent dat het antoniem (tegenovergestelde) van woord A het woord B is. Deze relatie voegden we slechts sporadisch toe.

- **PROV** : Dit betekent dat we te maken hebben met een zegswijze of spreekwoord. De zegswijzen zullen we pas in een later stadium toevoegen ; ze hebben immers ook maar een beperkte bruikbaarheid voor taalleerders (en gebruikers in het algemeen).

- **SYN(A : B, C, D,...)** : dit betekent dat woord A de woorden B,C,D als synoniem heeft. Let op, soms geldt de relatie slechts voor een betekenis van het woord !

- **ORT(A,B,C)** : Dit geeft de hervormde spelling of een alternatieve spelling van een woord weer. Als je bijvoorbeeld een zoekresultaat hebt met het woord 'connait', dan zou er wat verderop moeten staan : (ORT : connait).
- **LOC()** : Deze afkorting staat voor *locution (adjectivale, adverbiale, verbale,...)* en geeft de 'vastgeroeste' woordgroepen weer die in de taal vaak voorkomen. Woordgroepen zoals 'à cause de, pour que, demander pardon' vallen onder deze ietwat vage definitie. We plaatsten onder *LOC* ook *alle wederkerige werkwoorden* of *soms* werkwoorden die een onverwacht vast voorzetsel hebben. Voor het Nederlands bevat *LOC* ook de niet gescheiden vormen van de werkwoorden (bv : *doorlopen, rondkomen*). Als er bijvoorbeeld een analyse is van de zin 'Jan spreekt af met Marie', dan herkent de computer (en misschien de student ook niet) niet automatisch het woord 'afspreken', vandaar dat we dit dan toevoegen onder *LOC*.
- **EXP()** : Deze afkorting staat voor *les expressions*, uitdrukkingen. Ongetwijfeld rijst de vraag wat het verschil nu is met de vaste woordgroepen die we hierboven *les locutions* noemden. We beschouwen een woordgroep als *expression* indien het op het eerste gezicht niet duidelijk is hoe we van de som van de betekenissen komen tot de gehele betekenis van het de uitdrukking. De criteria om *locutions* en *expressions* van elkaar te scheiden zijn best wel wat complexer, maar omwille van de praktische aard van het programma kozen we ervoor om tussen beide slechts een rudimentair onderscheid te maken ; het maakt immers voor een taalleerder niet zoveel uit of iets nu een *locution* of een *expression* is. Bij updates zullen we beide systematischer proberen te onderscheiden. Als je met andere woorden een zoekopdracht wil lanceren naar alle vaste uitdrukkingen met het woord 'main' dan zoek je voor alle zekerheid eerst naar '*LOC main*' en vervolgens naar '*EXP main*' zodat je beide resultaten vervolgens kan samenleggen.

- *Waarom biedt dit programma geen spellingcorrectie aan en waarom verbetert het ook geen zinnen ?*

Dit programma beschikt niet over de nodige modules (*parser met regels*) om een zin syntactisch te analyseren. Willen we dit bereiken dan moeten we een parser toevoegen die kan oordelen of een zin correct gevormd is of niet. Spijtig genoeg houdt het daar niet bij op want omdat we in een « leercontext » zitten, zou deze parse-module ook moeten kunnen zeggen waar de fout juist zit (een notoir moeilijk en onopgelost probleem) en suggesties doen om de zin te verbeteren. Dit laatste is een uitdaging vanjewelste en kan niet snel en makkelijk geïmplementeerd worden.

Bovendien is het programma ook geen spellingcorrector, al kan het daar wel via enkele tweaks in de zoekopdrachten voor ingezet worden. Een spellingcorrector doet een oppervlakkige *parse* van de tekst en suggereert eventuele fouten aan de gebruiker op basis van de directe nabijheid van de omringende woorden. Als je wil weten of je kan zeggen *'une vieux femme'* dan kan je de sequentie *'vieux femme'* wel ingeven en dan zal het programma de etiketten retourneren van beide woorden. Al snel zal je opmerken dat deze etiketten niet verenigbaar zijn: 'mannelijk meervoud, enkelvoud' voor 'vieux' past niet bij 'vrouwelijk enkelvoud' voor 'femme' ! Indien de gebruiker twijfelt over de schrijfwijze van Franse woorden, dan hij hij het woord ook steeds de zoekfunctie gebruiken: zodra hij een onbestaand woord intypt, dan zal er een popup met suggesties verschijnen; deze suggesties werden berekend op basis van de *Levenshtein-distance* (met aanpassingen).

- *Kan dit programma een Leerkracht vervangen ?*

Zoals hierboven gezegd is dit programma louter ondersteunend en niet opgevat als middel om Frans op volledig autonome wijze te verwerven. De functie van de leerkracht bestaat er immers in om het leertraject uit te tekenen en de studenten hierin te begeleiden door met hen in interactie te treden en door de nodige feedback te geven op hun progressie. Deze feedback (over de taalcorrectheid...) is van erg groot belang bij het aanleren van een vreemde taal en kan niet zomaar vervangen worden door software. Indien de student echter een zekere autonomie heeft verworven in de nieuwe taal, kan hij wel al grotendeels een eigen leertraject (voor lexicale kennis) uittekenen door het programma. Interactie zal echter steeds een *must* blijven.

- *Waarom zou ik dit programma gebruiken en niet gewoonweg DeepL (Linguee), ChatGPT of een andere AI-tool gebruiken ? Een druk op de knop en ik krijg immers alle resultaten.*

Bovengenoemde tools zijn erg nuttig voor iemand die de taal reeds goed (of totaal niet) beheerst en snel documenten in een andere taal wenst op te stellen of te vertalen zonder al te veel tijd te verliezen. Verder kunnen studenten er ook baat bij hebben om bijvoorbeeld tijdwinst te maken of hulp te krijgen bij het opstellen van werken. Deze tools zijn echter **niet zo zinvol** voor mensen die een taal leren en die taal actief moeten kunnen gebruiken. **Door deze tools systematisch te gebruiken bij oefeningen tijdens het leertraject schakelen studenten het eigen nadenken**

over de taal uit en bouwen ze ook helemaal geen kennis over de taal uit. Er wordt een oplossing geboden en die wordt onkritisch overgenomen. Bovendien wordt er ook geen expliciet onderscheid gemaakt in deze programma's tussen de verschillende taalregisters en krijgt de gebruiker ook geen uitleg over het waarom van sommige constructies of woordkeuzes. Via *Lexogoth* wordt de taalleerder echter verplicht om kritisch naar **alle** aangereikte voorbeelden te kijken en datgene te selecteren wat het dichtst aansluit bij zijn vraag. Deze aanpak biedt talrijke voordelen : zo zal een leerder vaak dezelfde voorbeelden steeds opnieuw doorlezen en op die manier al lezend zijn lexicale kennis (aanzienlijk) uitbreiden. Bovendien verwerft hij door de talrijke opmerkingen ook de broodnodige kennis over het gebruik van woorden, een absolute voorwaarde om een competent taalgebruiker van een vreemde taal te worden.

- Wat is de rol van Lexogoth in het vreemdetalenonderwijs voor Frans ?

Lexogoth fungeert als een brug tussen corpuslinguïstiek en het klaslokaal Frans als vreemde taal (FLE). In tegenstelling tot traditionele tekstboeken die vaak kunstmatige voorbeelden gebruiken, biedt *Lexogoth* leerlingen authentieke, contextrijke taalgegevens. De rol binnen FLE kan worden samengevat in drie kernfuncties:

Datagestuurd leren: Het moedigt leerlingen aan om zich op te stellen als "taaldetectives". In plaats van enkel regels uit het hoofd te leren, ontdekken ze hoe het Frans daadwerkelijk wordt gebruikt in verschillende registers en contexten door meer dan 40.000 authentieke items te verkennen.

Focus op lexicale brokken: *Lexogoth* legt de nadruk op "prefabs" en collocaties, oftewel de manier waarop woorden op een natuurlijke manier samenhangen. Dit is essentieel voor FLE-leerders om een natuurlijke vloeiendheid te bereiken en verder te gaan dan letterlijke vertalingen.

Ondersteunde autonomie: Met de Lexicale Toolbox evolueert *Lexogoth* van een naslagwerk naar een actieve leeromgeving. Het stelt docenten in staat om gedifferentieerd materiaal te maken en stelt leerlingen in staat om hun eigen gepersonaliseerde leertrajecten op te bouwen op basis van de woordenschat die ze tegenkomen.

- Waar vind je het programma ? / Heb je internet nodig om met het programma te werken ? / Zijn er bijkomende kosten voor updates ? / Wat met de privacy ?

Je kan het programma aankopen via de *Microsoft Store* (26 euro) en zodra de download en de installatie zijn afgerond, kan je het programma onbeperkt in de tijd blijven gebruiken onder je persoonlijke account. In tegenstelling tot vele handboeken en taalpakketten is er voor de werking van Lexogoth geen internetverbinding nodig en worden alle (driemaandelijke) updates via de *Store* gratis aangeboden. Er zullen dus nooit extra kosten opduiken eenmaal je het programma hebt aangekocht en geïnstalleerd. Lexogoth verzamelt overigens geen persoonlijke gegevens van de gebruikers. Indien je het programma wenst aan te kopen voor een groep, dan kan je dit niet meer via de *Microsoft Store* doen omdat deze functionaliteit een jaar geleden werd afgeschaft; je zal hiervoor steeds de auteur moeten contacteren op het adres info.Lexogoth@gmail.com.

- Hoe moet je het programma updaten?

Zoals hierboven gezegd kan je *Lexogoth* uitsluitend via de *Microsoft Store* verkrijgen. We verbinden ons ertoe om driemaandelijks updates voor het programma uit te brengen. Deze updates hebben vooral betrekking op de uitbreiding van het tekstcorpus en de verbetering van fouten. Het updateproces verloopt als volgt: Je opent de *Microsoft Store*-app op je computer, vervolgens selecteer je *Library* (Bibliotheek), daarna krijg je de optie *Update (Apps)* te zien. Als je die aanklikt, worden alle apps geüpdatet. Onthoud dat het **updaten** van Lexogoth steeds **manueel** via de *Microsoft Store*-app dient te gebeuren en dat dit niet automatisch gebeurt via de reguliere Windows-updates. Organisaties die een groepsaankoop hebben verricht, krijgen een update via e-mail.

- *Opmerkingen in verband met de installatie van Lexogoth en de Laadtijden*

* **Belangrijk:** vanaf versie 3.8 van Lexogoth (huidige is v.3.15) moet de gebruiker niet meer eerst de *Microsoft Visual c++ redistributable*, **minstens versie 14.40.x**, architectuur x64 geïnstalleerd hebben; we voegden de nodige files zelf al toe aan het installatieprogramma. Mochten er bij het openen van Lexogoth foutmeldingen komen over

ontbrekende bestanden, dan surf je naar onderstaande link en download en installeer je de Microsoft Visual C++ redistributable (architectuur x64): <https://learn.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-170> [onder de titel *Latest Microsoft Visual C++ Redistributable Version*] Daarna probeer je Lexogoth opnieuw te installeren.

*Indien je Windows-besturingssysteem in S-modus draait (zeldzaam), kan het programma mogelijk niet goed werken. Een oplossing zou dan kunnen zijn dat je de S-modus uitschakelt of dat je het programma naar een andere computer downloadt en het daarop installeert. Deze mogelijke beperking heeft te maken met machtigingen die niet worden verleend aan programma's door een besturingssysteem in S-modus.

*Het duurt toch zo'n acht à tien seconden om het programma te laten opstarten ; deze laadtijd is te wijten aan de grote hoeveelheid gegevens die in het geheugen moet worden geladen (ongeveer een zestal 'lexicons'). Zodra het programma echter is opgestart, kan elke zoekopdracht in een fractie van een seconde worden voltooid.

*Indien je merkt dat het *resizen* van het programma haperend verloopt, dan zullen er in het venster van de zoekresultaten wellicht al duizenden voorbeeldzinnen staan, zonder dat je je daarvan bewust bent. De gemakkelijkste oplossing is dan om je scherm te verversen (refresh-icoon) en het resizen zal weer vlot kunnen verlopen.

6. De hervormde spelling

We hebben in ons lexicon getracht om zoveel mogelijk rekening te houden met de hervormde spelling van 1990 en die systematisch toe te voegen aan het corpus. Bij de meeste woorden hanteerden we de klassieke spelling en gaven we vervolgens de vernieuwde spelling op het einde van de zin. Deze herken je door de toevoeging **ORT**.

In principe zijn dus beide schrijfwijzen van een woord opgenomen bij de zoektermen ; er zijn echter nog wel wat woorden waarvoor we dit nog niet hebben kunnen doen en daarom raden we de gebruikers aan om zeker woorden met een 'accent circonflexe' (ja, dat *dakjes*accent) op te zoeken met én zonder accent om zeker alle relevante zoekresultaten te krijgen. Als

je bijvoorbeeld wil zoeken naar « croître » dan zoek je best zowel naar « croître » als naar « croitre ». We hebben deze hervormde spelling nog niet systematisch ingezet. Onthoud wel dat de spelling van voor de hervorming **niet** als verouderd of foutief mag worden beschouwd, deze spelling blijft voor « onbepaalde tijd » geldig en is volledig evenwaardig aan de hernieuwde spelling. Kwestie van de zaken te bemoeilijken.

7. Variëteiten van het Frans: taalniveaus, belgicismen en anglicismen

We besteden in Lexogoth de nodige aandacht aan de verschillende variëteiten van het Frans. Zo vermelden we meestal tot welk taalregister een woord of een constructie behoort. We onderscheiden in *Lexogoth* 4 ruime taalregisters voor de behandelde woorden en constructies :

- de *standaardtaal* of *courante taal*, dat is de taal zoals die in het dagelijkse leven wordt gebruikt op het werk, op school en in de media.
- de *chiquere, verzorgde en literaire taal*: woordenschat en constructies waarvoor toch wel aanzienlijke lexicale kennis verondersteld wordt en die vooral tot het geschreven taalgebruik behoren
- de *familiaire taal*: dit is het informele taalgebruik, gangbaar tussen sprekers die elkaar goed kennen; vooral mondelinge taal
- de *volkse taal (registre populaire)*: Dit is de zeer informele taal die bijna uitsluitend mondeling 'bestaat' en zich kenmerkt door een nonchalante uitspraak, weinig lexicale en syntactische variatie, makkelijke constructies, gebruik van grove (vulgaire) woorden, gebruik van verlan en anglicismen, ... Om deze redenen beschouwt men het volkse register als het laagste taalniveau. Deze variëteit heeft voor de meeste vreemdetaalleerders slechts een beperkte, communicatieve inzetbaarheid. We raden de gebruikers aan om zeer voorzichtig met dit register om te springen en om dit niet lichtzinnig te gebruiken zonder eerst goed de communicatieve context begrepen te hebben.

Voor meer uitleg over de kenmerken van de verschillende taalregisters verwijzen we naar het document dat via het menu

grammatica in Lexogoth kan geraadpleegd worden: *Uitbreiding 12: Les registres de Langue*

We signaleren aan de gebruiker ook steeds wanneer hij te maken heeft met een belgicisme. Een belgicisme is een uitdrukking die vooral gebruikt wordt door Belgische Franstaligen en buiten België nauwelijks of geen verspreiding kent; belgicismen worden met andere woorden niet tot de algemene standaardtaal gerekend. Deze belgicismen komen in veel gevallen voort uit de regionale varianten van het Belgisch Frans, uit het Nederlands, meer bepaald het Vlaams of het Standaardnederlands, of uit het Waals, een zuster taal (met talrijke varianten) van het Frans die van oudsher in Zuid-België werd gesproken en na WOII definitief werd verdrongen door het Frans. De bekendste belgicismen zijn ongetwijfeld « *nonante* », « *septante* » en het overmatig gebruik van « *savoir* ». De lezer zal moeten oordelen, in functie van het communicatieve doel, of het aangewezen is om dit belgicisme al dan niet te gebruiken. Je kan makkelijk de meeste (lexicale) belgicismen in *Lexogoth* terugvinden door de zoekterm « belgicisme » in te geven. We besteden in *Lexogoth* bijna uitsluitend aandacht aan de Franse belgicismen en niet aan de Nederlandse belgicismen.

Ook het Frans is niet immuun voor de invloed van het Engels en we besteden dan ook de nodige aandacht aan de anglicismen. Anglicismen zijn constructies en woorden die rechtstreeks uit het Engels (meestal Amerikaans-Engels) werden overgenomen in het Frans. De lezer kan ze makkelijk terugvinden door de zoekterm “anglicisme” in te typen in het zoekveld (optie taal: Frans).

8. To Do in de volgende updates

- We trachten systematisch viermaandelijks een update uit te brengen met correcties en uitbreidingen van het corpus.
Volgende update: eind maart 2026

9. Versies en updates

- v4.9 (februari; 4.8, 4.7, v4.6: interne testversies)

- Toevoeging *Contextmodules* waarmee zelf blokken kunnen samengesteld worden.
- Verbetering van de zoeksnelheid voor grote corpora
- correcties en toevoegingen aan het corpus
- verbetering leesbaarheid tekstmateriaal
- snellere opstarttijden
- verbeterde technische robuustheid: Optimalisatie van het geheugenbeheer en algemene stabiliteitsverbeteringen.
- verbeterde layout
- verbeterde validatielogica voor imports
- verbetering van de zoeklogica
- Geavanceerde zoeklogica in de database: zoeken binnen de database met verschillende opties, batch-verwijdering van records, N-gram analyse van database, concordantiefunctie (combinaties woorden)[Toolbox]
- Verbetering gebruikersgemak
- Mogelijkheid om opmerkingen van het Lexogoth-corpus makkelijk over te zetten naar de eigen database
- Verbetering opslaglogica en ruimtebesparing bij grote bestanden
- Toevoeging gegevens over Lexogoth-database [zie onderaan]
- Mogelijkheid om contextmodules samen te voegen
- Mogelijkheid om testen samen te voegen
- Concordantie-analyse: instelbare parameters (woorden voor en na); visuele highlighting; previewvenster
- beveiliging tegen html-injecties
- verbeterde export-functies (rtf)

- v4.5 (januari 2026)

- correcties
- toevoegingen van ong. 50 bladzijden chunks
- toevoeging van de module *Lexicale Toolbox*: met deze module kunnen docenten en studenten lessen, oefeningen, databases samenstellen op basis van Lexogoth (corpus) en erg gericht te werk gaan tijdens het studeren of onderwijzen.

- **v4.4 (november 2025)**
 - correcties
 - toevoeging chunks (ong. 100 bladzijden; grondige analyse van enkele courante werkwoorden zoals (*se*) *reprendre*, *postuler*, *candidater*, *enquêter*...)
 - verbetering van de leesbaarheid van de zoekresultaten door gebruik van kleuren bij het afficheren van de resultaten: nu worden de verschillende chunks gescheiden door felrode “/” en “(-/-)” om de lezer makkelijker te oriënteren.

- **v4.3 (oktober 2025)**
 - correcties
 - toevoeging chunks (ong. 100 bladzijden rond o.m. spreektaal, interjecties, verkeer)

- **v4.2 (eind augustus 2025)**
 - uitgebreide uitleg over de gebruikte methodologie op de website (<https://lexogoth.be>)
 - grondige toevoeging van chunks (ong. 100 bladzijden)

- **v4.1 (juli 2025)**
 - revisie van het volledige werk: typfouten, controle van opgenomen idiomatische uitdrukkingen (uitbreiding met voorbeelden of verwijdering in geval van te lage frequentie)
 - Nederlandse woordenlijst werd gereviseerd en genormaliseerd (o.m. eliminatie van niet-zelfstandig bestaande stammen zoals “vaardigen” in “uitvaardigen”)
 - forse uitbreiding voorbeelden (ongeveer 100 bladzijden)
 - verbeteringen en aanpassingen lessen grammatica, meer bepaald de les over de betrekkelijke voornaamwoorden.

- **v3.15 (april 2025)**
 - correcties en aanvullingen

- **v3.14 (interne testversie, april 2025)**

- **v3.13 (maart 2025)**
 - correcties en aanvullingen
- **v3.12 (maart 2025)**
 - correcties
 - toevoegingen letter “e” [ongeveer 90 bladzijden]
- **v3.11 (januari 2025):**
 - correcties
 - toevoeging anglicismen, courante verlan-woorden
 - toevoeging les over taalregisters
- **v3.10 (december 2024):**
 - correcties in het corpus, talrijke toevoegingen
 - aanpassing van de tutorial
- **v3.9 (december 2024):**
 - Het installatieproces voor Lexogoth werd vereenvoudigd door de benodigde *dll*-files aan het programma zelf te koppelen zodat gebruikers de Visual C++ Redistributable niet meer manueel hoeven te installeren.
- **v3.8 (december 2024):**
 - verduidelijking via andere kleur consequent ingevoerd voor beide talen.
- **v3.7 (december 2024):**
 - programma-update om de gebruiksvriendelijkheid te verhogen: voor het Frans worden alle zoektermen nu in het blauw weergegeven in de voorbeelden zodat de gebruiker snel de voorbeelden kan terugvinden zonder verplicht te worden de hele tekstblokken te lezen. Voor het Nederlands is deze functionaliteit al deels beschikbaar.
 - lexicale toevoegingen en correcties
- **v3.6.55 (november 2024):**
 - correcties; toevoegingen informele taal

- **v3.6.54 (november 2024):**
 - correcties; toevoegingen informele taal
- **v3.6.53 (november 2024):**
 - Correcties
 - Toevoegingen grammatica: *Le gérondif / L'accord du participe passé*
 - toevoeging informele uitdrukkingen
 - toevoeging deel 1 van letter "e" (ongeveer 100 bladzijden)
 - aanpassingen handleiding
- **v3.6.52:**
 - testversie
- **v3.6.51 (augustus 2024):**
 - Volledige revisie van de Franse en Nederlandse lemma's
 - Verbeterde zoekmogelijkheden Franse lemma's door onderverdeling lemma's per categorie.
 - Toevoeging voorbeelden familiair taalgebruik
 - Toevoeging belgicismen (ongeveer 164)
 - Toevoeging korte gebruiksaanwijzing
- **v3.6.49 (juni 2024):**
 - correcties
 - afwerking van de letter "d"
- **v3.6.48 (maart 2024):**
 - correcties en enkele toevoegingen
- **v3.6.47 (maart 2024):**
 - toevoeging van ongeveer 1500 chunks; forse uitbreiding van de letter "d" en toevoeging van talrijke familiale lemma's en constructies.
 - Verandering functionaliteiten: de optie *zoeken* is uit het menu verdwenen omdat ze overbodig was geworden en een toevoeging onder het zoekveld laat de gebruiker

kiezen of de vertaling al dan niet getoond wordt.

- v3.6.46: toevoeging van ongeveer 1000 chunks; begin systematische opvulling lacunes (letter “d” tot en met “d mo-”); toevoeging belgicismen; verbeteringen
- v3.6.43: toevoeging van ongeveer 1000 chunks
- v3.6.41-42: correcties; toevoeging talrijke belgicismen
- v3.6.40:
 - uitleg over gebruik voegwoorden *quoique,   condition que* (+ gebruik *  condition de, de peur de,...*)
 - toevoegingen terminologie * lektrische wagens*
- v3.6.39: veranderingen aan de vorige versie (3.6.35):
 - correctie spelfouten en verduidelijkingen
 - toevoeging van ongeveer 50 bladzijden bijkomende voorbeelden

10. Technische zaken en copyright

-*Lexogoth* werd geschreven in c++ en gecompileerd met Visual Studio (versie 17.12.3). De gebruikersinterface (GUI) werd opgebouwd met Qt onder een LGPL-licentie (versie 6.4; <https://www.qt.io/>). We gebruikten enkele bibliotheekfuncties van *Boost* (versie 1.78 ; <https://www.boost.org/>).

De geraadpleegde werken en websites voor de samenstelling van het corpus :

-woordenboeken :

<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/>

-taaladvies :

<https://vitrinelinguistique.oqlf.gouv.qc.ca/>
<https://usito.usherbrooke.ca/>
<https://www.larousse.fr/>
<https://www.vlaanderen.be/taaladvies/>
<https://etienne-meul.be/>
<https://www.deepl.com/en/translator>

- Lexicons

Frans lexicon LeFFF, uitgegeven onder LGPL. De originele gebruikte versie 2.1 kan hier teruggevonden worden :
<https://www.labri.fr/perso/clement/lefff/telechargement.html>
Nederlands lexicon : eigen compilatie op basis van corpus

-samenstelling corpus

Het tekstcorpus (versie 4.8) bevat ruim 45 000 aanklikbare items. Deze werden over een periode van een tiental jaren samengesteld. De voorbeelden werden meestal herschreven (met het oog op een makkelijke verwerving door de taalleerders) en zijn afkomstig uit de volgende bronnen : blogs, internetfora, websites, woordenboeken (zie hierboven), websites voor taaladvies, films en series, geschreven en gesproken pers (vooral België en Frankrijk) en overheidsdocumenten.

- Gegevens corpus:

=====

FINALE CORPUS & METADATA ANALYSE

=====

Gevalideerde records: 13405

Totaal aantal chunks: 40215

Woorden Frans: 455994 tokens

Woorden Nederlands: 428558 tokens

Woorden Opmerkingen (+): 338039 tokens (incl. tags)

INFORMATIEDICHTHEID: 0.74

VERDELING CATEGORIEEN (TAGS):

- Q : 27343 keer
- LOC : 7097 keer
- EXP : 1331 keer
- SYN : 532 keer
- ORT : 452 keer
- CONT : 65 keer

=====

Opmerking: De bovenstaande cijfers tonen de omvang en de diepgang van de Lexogoth-database aan. Met een informatiedichtheid van 0.74 en een sterke focus op contextuele annotaties (Q en LOC) functioneert dit corpus als een geavanceerd **didactisch** instrument. Dit getal betekent dat er voor 100 Franse woorden 74 woorden uitleg wordt gegeven, een indicator dat het Lexogoth corpus sterk geannoteerd is en kan dienen als didactisch instrument.

Lexogoth Versie 4.9 (februari 2026)

Auteur : Frederik Meul, 2023

Contact : info.lexogoth@gmail.com

Websites:

<https://lexogoth.be>,

<https://sites.google.com/view/lexogothinternational/> [Engels-Franse versie]