



# Enfin! Une liste de vocabulaire française fondée sur les familles de mots

TOM COBB, CHRISTINA LINDQVIST ET MÅRTEN RAMNÄS

Diapo disponible sur: [lextutor.ca/acfas\\_lex\\_23/liste\\_fams.pdf](https://lextutor.ca/acfas_lex_23/liste_fams.pdf)

# Survol

- Listes de fréquence pour l'apprentissage, l'enseignement, et l'évaluation des mots
- Unités de comptage – types, lemmes, flexions, familles de mots
- Désavantages du lemme
- Avantages de la famille de mots
- Création et évaluation d'une liste de familles de mots

Il y a des listes de fréquence pour l'anglais. Pour le français, il y en a aussi, mais beaucoup moins.

BNC/COCA Word Family Lists (Nation)

De façon générale, on apprend d'abord les mots fréquents avant les mots moins fréquents

*A frequency dictionary of French – core vocabulary for learners*  
(Lonsdale & Le Bras, 2009)

1 – 25 000 lemmes ou familles de mots, divisées par tranches de 1 000 mots (1K, 2K, 3K, etc.)

Liste de fréquence aux universités en Suède, fondée sur les lemmes

Obligatoire, les mots de la liste sont testés lors d'un examen à la fin du semestre

# Le choix de l'unité de comptage est important :

- Pour fixer les mots à apprendre (combien de mots, quels mots), pour établir une liste de vocabulaire, pour savoir comment présenter les mots.
- Pour évaluer les connaissances lexicales (mesurer l'étendue du vocabulaire).
- Pour savoir quels mot (quelles formes) inclure dans un dictionnaire.
- Pour savoir le nombre de mots mots qui sont nécessaires pour comprendre une émission, un livre, p.ex.

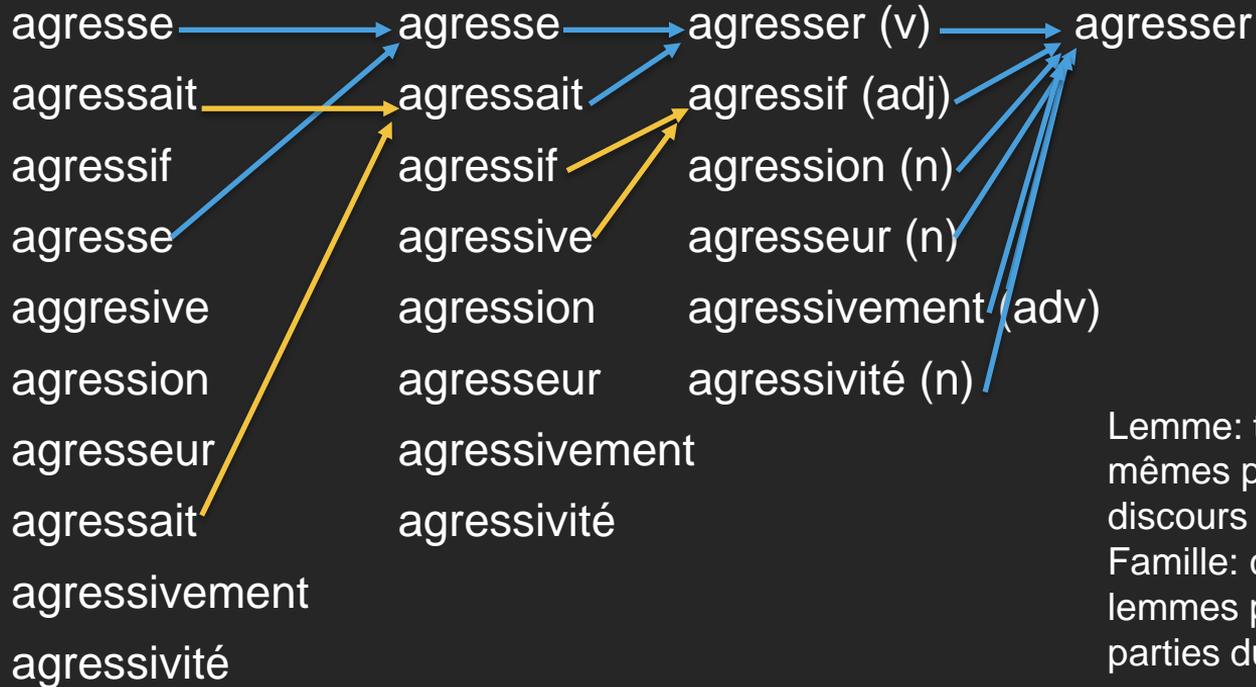
# Unités de comptage

**mot-occurrence**  
(9 tokens)

**mot-type**  
(7 types)

**(f)lemme**  
(5 lemmas)

**famille de mots**  
(1 word family)



Lemme: toujours les mêmes parties du discours  
Famille: comprend les lemmes plus d'autres parties du discours

# Désavantages du lemme

- "Couverture": les recherches montrent que la compréhension d'un texte demande la compréhension de 98% des mots constituant le texte
- Dans les études du lexique français, le lemme a été utilisé
- Le nombre de lemmes nécessaire pour atteindre ce pourcentage est énorme...

- Pour l'anglais: 8 000 – 9 000 familles de mots pour comprendre un journal ou un texte académique (Nation 2006)
- Pour le français: les romans étudiés pendant le premier semestre de l'université en Suède: 7 000 – 9 000 lemmes (Ramnäs 2019)

# Camus, *L'Étranger*

Vocabulaire Nombre de lemmes	Nombre de mots- occurrence	Nombre de lemmes	Nombre de lemmes (nombre cumulé)	Couverture
1 000	29 776	711	711	86,0%
2 000	1 874	449	1160	91,5%
3 000	993	331	1491	94,3%
4 000	518	229	1720	95,8%
5 000	288	163	1883	96,7%
6 000	153	85	1968	97,1%
7 000	172	94	2062	97,6%
8 000	152	82	2144	98,0%
9 000	91	61	2205	98,3%
10 000	79	64	2269	98,5%
11 000	55	41	2460	98,7%
12 000	67	48	2553	98,9%
13 000	43	36	2589	99,1%
14 000	42	38	2643	99,2%
15 000	48	28	2460	99,3%
20 000	98	93	2553	99,7
25 000	44	36	2589	99,8
Off list	69	54	2643	100

- ...tandis que plus de la moitié de ces lemmes sont des mots probablement déjà connus, partiellement connus ou facilement inférables
- Dans des dictionnaires fondés sur les lemmes, comme Lonsdale & Le Bras (2009), les lemmes apparemment liés sont répartis dans différents niveaux de fréquence (5 000)
  - stable (3K)
  - stabilité (3K)
  - instable (5K)
  - instabilité (5K)
  - stabiliser (5K)
  - stabilisation (5K)

## Au-delà des 5 000 lemmes les plus fréquents, cela continue...

- stable (3K)
- stabilité (3K)
- instable (5K)
- instabilité (5K)
- stabiliser (5K)
- stabilisation (5K)
- déstabiliser (6K)
- déstabilisation (10K)
- stabilisateur (13K)
- déstabilisateur (20K)
- stablement (20K+)
- instablement (20K+)

# Avantages de la famille de mots

- En descendant la liste, l'apprenant rencontre de "vrais" nouveaux mots
- Permet à l'enseignant de concevoir un enseignement adapté et progressif du vocabulaire
- Permet de mieux combiner cours de grammaire et cours de vocabulaire
- Allège la charge d'apprentissage des étudiants (*learning burden*)
- Prise de conscience des séries dérivationnelles (conscience morphologique) → impact important sur la compréhension écrite
- Permet de mieux concevoir des textes simplifiés (*graded readers*)

Nous ne sommes pas les premiers à proposer l'utilisation de la famille de mots, mais les propositions précédentes n'ont pas été faites de manière conséquente, surtout en pédagogie

Le Nouveau Petit Robert (2015):

- adjectif et nom: *patient*
- participes: *périmé*, *handicapant*
- *licitement* dans le l'article de *licite*
- *glaciologue* et *glaciologique* dans l'article *glaciologie*
- **Besoin d'économie de place (entrées cachées)**
- **Sens compositionnel**

# Création d'une liste de familles de mots

- Pas d'outil automatique
- Travail manuel
- Lonsdale nous a fourni la liste complète de 25 000 lemmes (formes canoniques uniquement)
- Les lemmes ont été développés en ajoutant toutes les formes fléchies sur [lexique.org](http://lexique.org), e.g. *stabiliser*

stabiliser stabilisait stabilisant stabilise stabilisent stabilisera stabiliserait stabilisèrent stabilisé stabilisée stabilisées stabilisés ...

Nous avons fait des recherches manuelles dans la liste des 25 000 lemmes, afin d'identifier et joindre les formes en majuscules

Les critères d'identification étaient plusieurs, regroupés sous la question centrale : un apprenant est-il capable de voir un lien entre les formes?

700+ pages

Les lemmes ont été réunis en familles!

```
1 1
2 abord abords ABORDER ABORDA ABORDAI ABORDAIENT ABORDAIS ABORDAIT
  ABORDANT ABORDAS ABORDE ABORDENT ABORDERA ABORDERAI
  ABORDERAIENT ABORDERAIS ABORDERAIT ABORDERIONS ABORDERONS
  ABORDERONT ABORDES ABORDEZ ABORDIEZ ABORDIONS ABORDONS
  ABORDÂMES ABORDÈRENT ABORDÉ ABORDÉE ABORDÉES ABORDÉS ABORDABLE
  ABORDABLES ABORDAGE ABORDAGES INABORDABLE INABORDABLES
3 accepter accepta acceptai acceptaient acceptais acceptait acceptant acceptassent
  acceptasses accepte acceptent acceptera accepterai accepteraient accepterais
  accepterait accepteras accepterez accepteriez accepterions accepterons accepteront
  acceptes acceptez acceptiez acceptions acceptons acceptâmes acceptât acceptèrent
  accepté acceptée acceptées acceptés ACCEPTABLE ACCEPTABLES INACCEPTABLE
  INACCEPTABLES ACCEPTATION ACCEPTATIONS ACCEPTIONS ACCEPTABILITÉ
  ACCEPTABILITÉS
4 accord accords ACCORDER ACCORDA ACCORDAI ACCORDAIENT ACCORDAIS ACCORDAIT
  ACCORDAIT ACCORDANT ACCORDAS ACCORDASSENT ACCORDE ACCORDENT
  ACCORDERA ACCORDERAI ACCORDERAIENT ACCORDERAIS ACCORDERAIT
  ACCORDERAS ACCORDEREZ ACCORDERIEZ ACCORDERONS ACCORDERONT
  ACCORDES ACCORDEZ ACCORDIEZ ACCORDIONS ACCORDONS ACCORDÂT
  ACCORDÈRENT ACCORDÉ ACCORDÉE ACCORDÉES ACCORDÉS DÉSAccORD
  DÉSAccORDS
5 accuser accusa accusai accusaient accusais accusait accusant accuse accusent
  accusera accuseraient accuserais accuserait accuseriez accuserions accuseront
  accuses accusez accusiez accusions accusons accusât accusèrent accusé accusée
  accusées accusés ACCUSATION ACCUSATIONS ACCUSATEUR ACCUSATEURS
  ACCUSATRICE ACCUSATRICES
```

# Problème

- 1000 familles de mots = 26 639 formes
- La plupart de ces formes sont rarement utilisées
- Comment réduire le nombre de formes, pour ne garder que les formes les plus utilisées?
- Utilisées où?
- Nous avons besoin de tester nos familles sur un corpus de la langue de tous les jours
- Faute d'en trouver un, nous avons créé notre propre corpus

# Corpus de comparaison

2,7 millions de mots (oral, écrit, journalistique, académique, québécois, français)

## Cross Corpus

1 This is 2.7 million words of combined corpora draw from the following full corpora (described in detail below)

- WRITTEN ACAD academic 1 million+
- WRITTEN JOURNALISM 3 newspapers from Chambers 450,000 words (may yet get a Montreal paper into the mix) + part of LeMonde
- WRITTEN CHILDREN Petit Nicolas 130,000 words
- SPOKEN Paris 488,000
- SPOKEN Quebec 703,000

# Nuclear List Builder v.2.6

Reduce families to just the members present in a corpus

NEW IN '22-'23 - FRENCH FAMILIES

- +: headword override
- +: all/no-inflects override
- +: display derived option
- +: French fams (in prog.)
- NEW: Freq-for-complete-fams list
- NEW: Percent derived/fam **Apr 23**

Most family lists (e.g. BNC/Coca) are based on large corpora and complete as possible to classify every word of any text (e.g. in VP). But there is no reason for learners to know every form of every word. They may never meet them, and anyway different text genres (engineering, medical) employ signature subsets of family members. This program "crosses" family lists against word frequencies in a smaller standard corpus (1-4 million word) or user corpus (up to  $\cong$  800,000 words) to obtain just the members it contains.

Read a [research paper](#) or [summary](#) about this

**(1) Choose full BNC/Coca, AWL, or French (Fr-5) families list**

Fr\_fams\_1k ▾

**(2) Upload Own Cross-Corpus/Text**

(850k wds max; must be ~.txt)

Choose File **No file chosen**

or

**Choose Cross-Corpus**

French Combo (1.42m, last 5) ▾

**(3) View output as...**

**FIRST**  
Complete frequencies and %'s (To explore cutoffs ? or get Fam-Freq list ■ )

**THEN**  
Options and outcomes

**(4) Choose cutoffs ↓**

**(5) Click**

Submit >>

LINGO: French

LIST: fr\_fams\_heads\_k01

ORIGINAL FAMILIES: 1001

ORIGINAL WORD TYPES: 26639

**CROSSED x CORPUS LIST:**

Fr\_combo\_freq.txt

SHOWING AS: Freqs\_and\_percents

WORD FAMILIES >0%: 1001

WORD TYPES >0%: 13233

**LEGEND:**

Corpus-freq, Word, Fam%

[Select list](#)

1. **de**

140,971 TOTAL

99442 de 70.54% -

23878 des 16.94% -

17651 du 12.52% -

2. **le**

137,818 TOTAL

66943 le 48.57% -

43246 la 31.38% -

27629 les 20.05% -

3. **à**

41,047 TOTAL

28182 à 68.66% -

9163 au 22.32% -

3702 aux 9.02% -

4. **être**

35,814 TOTAL

17506 est 48.88% -

4146 sont 11.58% -

3275 été 9.14% -

2776 était 7.75% -

2656 être 7.42% -

803 soit 2.24% -

703 sera 1.96% -

643 suis 1.80% -

591 étaient 1.65% -

516 serait 1.44% -

360 fut 1.01% -

306 sommes 0.85% -

278 seront 0.78% -

247 étais 0.69% -

185 étant 0.52% -

163 êtes 0.46% -

159 soient 0.44% -

124 seraient 0.35% -

73 es 0.20% -

66 furent 0.18% -

55 étions 0.15% -

32 êtres 0.09% -

31 fût 0.09% -

28 serais 0.08% -

20 étiez 0.06% -

15 serai 0.04% -

11 serons 0.03% -

11 sois 0.03% -

8 serez 0.02% -

5 soyez 0.01% -

109. **grouper**

1,288 TOTAL

932 groupe 72.36% -

214 groupes 16.61% -

33 regroupe 2.56% -

20 regroupant 1.55% -

17 regroupement 1.32% -

16 groupement 1.24% -

10 regrouper 0.78% -

10 regroupés 0.78% -

7 regroupé 0.54% -

6 regroupements 0.47% -

4 groupements 0.31% -

4 regroupées 0.31% -

3 groupés 0.23% -

3 regroupaient 0.23% -

3 regroupent 0.23% -

2 regroupera 0.16% -

1 groupé 0.08% -

1 groupée 0.08% -

1 groupées 0.08% -

1 regroupée 0.08% -

0 grouper 0% -

0 groupa 0% -

0 groupaient 0% -

0 groupais 0% -

0 groupait 0% -

0 groupant 0% -

0 groupent 0% -

0 groupera 0% -

0 grouperaient 0% -

0 grouperait 0% -

0 grouperons 0% -

- En éliminant les formes qui constituent moins de 1 % de la famille, la liste est réduite:
  - 1K: 26 639 > 7 429
  - 2K: 18 383 > 7 083
  - 3K: 11 479 > 4 327
- Réduction totale: 56 501 > 18 839
- Réduction de 2/3
- 3 000 familles de mots, environ 6 formes par famille
- De plus: 8 000 lemmes=80,000 formes différentes, alors que 3 000 familles =1 800 formes différentes > charge d'apprentissage allégé!

109. grouper

1,288 TOTAL  
932 groupe 72.36% -  
214 groupes 16.61% -  
33 regroupe 2.56% -  
20 regroupant 1.55% -  
17 regroupement 1.32% -  
16 groupement 1.24% -  
10 regrouper 0.78% -  
10 regroupés 0.78% -  
7 regroupé 0.54% -  
6 regroupements 0.47% -  
4 groupements 0.31% -  
4 regroupées 0.31% -  
3 groupés 0.23% -  
3 regroupaient 0.23% -  
3 regroupent 0.23% -  
2 regroupera 0.16% -  
1 groupé 0.08% -  
1 groupée 0.08% -  
1 groupées 0.08% -  
1 regroupée 0.08% -  
0 grouper 0% -  
0 grupa 0% -  
0 groupaient 0% -  
0 groupais 0% -  
0 groupait 0% -  
0 groupant 0% -  
0 groupent 0% -  
0 groupera 0% -  
0 groupements 0% -

108. nation

nation  
nations  
national  
nationale  
nationales  
nationaux  
nationaliste  
nationalistes  
nationalité  
nationalisme

109. grouper

grouper  
groupe  
groupes  
groupement  
regroupant  
regroupe  
regroupement

108. nation

nation  
nations  
z\_national  
z\_nationale  
z\_nationales  
z\_nationalisme  
z\_nationaliste  
z\_nationalistes  
z\_nationalité  
z\_nationaux

109. grouper

grouper  
groupe  
groupes  
z\_groupement  
z\_regroupant  
z\_regroupe  
z\_regroupement

# Évaluation

# Compleat Web VP v.2.6 Million-word VPs

Profile texts with \*9\* list frames in \*2\* langs. and \*2\* grain-sizes at \*1\* interface

Units: Note that BNL, Coca-Core, CEFR, & Classic AWL are not *full* 1000-family lists (see [?] details); & that NGSL, BNC-25k, and French are (F)lemmas not Families

## Related

How to make list framework comparisons? Last Demo [From here](#)

Does lex frequency predict text complexity? Check [these](#)

Check VP-sorted concordances, easiest contexts first [here](#)

**NEW! Morpholex integration @ k-level**



**Input mode A** Type or paste text up to 100,000 words into window (VP WILL LIMIT INPUT TO ~ 15,000 WORDS), choose options, and click **Submit\_window** button

Title: Fr\_camus\_etranger.txt [+](#) Eng+Fr! Cognates  (Lists) [+](#) Sentence Count  [+](#) Bar Chart  [+](#) Count Index  [?](#) [+](#) Basic  [?](#)

Aujourd'hui, ma mère est morte. Ou peut-être hier, je ne sais pas. J'ai reçu un télégramme de l'asile: Mère décédée. Enterrement demain. Sentiments distingués. Cela ne veut rien dire. C'était peut-être hier. L'asile de vieillards est à Marengo, à quatre-vingts kilomètres d'Alger. Je prendrai l'autobus à deux heures et j'arriverai dans l'après-midi. Ainsi, je pourrai veiller et je rentrerai demain soir. J'ai demandé deux jours de congé à mon patron et il ne pouvait pas me les refuser avec une excuse pareille. Mais il n'avait pas l'air content. Je lui ai même dit: Ce n'est pas de ma faute. Il n'a pas répondu. J'ai pensé alors que je n'aurais pas dû lui dire cela. En somme, je n'avais pas à m'excuser. C'était plutôt à lui de me présenter ses condoléances. Mais il le fera sans doute après-demain, quand il me verra en deuil. Pour le moment, c'est un peu comme si maman n'était pas morte. Après l'enterrement, au contraire, ce sera une affaire classée et tout aura revêtu une allure plus officielle. J'ai pris l'autobus à deux heures. Il faisait très chaud. J'ai mangé au restaurant, chez Céleste, comme d'habitude. Ils avaient tous beaucoup de peine pour moi et Céleste m'a dit: On n'a qu'une mère. Quand je suis parti, ils m'ont accompagné à la porte. J'étais un peu étourdi parce qu'il a fallu que je monte chez Emmanuel pour lui emprunter une cravate noire et un brassard. Il a perdu son oncle, il y a quelques mois. J'ai couru pour ne pas manquer le départ. Cette hâte, cette course, c'est à cause de tout cela sans doute, ajouté aux cahots, à l'odeur d'essence, à la réverbération de la route et du ciel, que je me suis assoupi. J'ai dormi pendant presque tout le trajet. Et quand je me suis réveillé, j'étais tassé contre un militaire qui m'a souri et qui m'a demandé si je venais de loin. J'ai dit oui pour n'avoir plus à parler. L'asile est à deux kilomètres du village. J'ai fait le chemin à pied. J'ai voulu voir maman tout de suite. Mais le concierge m'a dit qu'il fallait que je rencontre le directeur. Comme il était occupé, j'ai attendu un peu. Pendant tout ce temps, le concierge a parlé et ensuite, j'ai vu le directeur: il m'a reçu dans son bureau. C'était un petit vieux, avec la Légion d'honneur. Il m'a regardé de ses yeux clairs. Puis il m'a serré la main qu'il a gardée si longtemps que je ne savais trop comment la retirer. Il a consulté un dossier et m'a dit: Mme Meursault est entrée ici il y a trois ans. Vous étiez son seul soutien. J'ai cru qu'il me reprochait quelque chose et j'ai commencé à lui expliquer. Mais il m'a interrompu: Vous n'avez pas à vous justifier, mon cher enfant. J'ai lu le dossier de votre mère. Vous ne pouviez subvenir à ses besoins. Il lui fallait une garde. Vos salaires sont modestes. Et tout compte fait, elle était plus heureuse ici. J'ai dit: Oui, monsieur le Directeur. Il a ajouté: Vous savez, elle avait des amis, des gens de son âge. Elle pouvait partager avec eux des intérêts qui sont d'un autre temps. Vous êtes jeune et elle devait s'ennuyer avec vous. C'était vrai. Quand elle était à la maison, maman passait son temps à me suivre des yeux en silence. Dans les premiers jours où elle était à l'asile, elle pleurait souvent. Mais c'était à cause de l'habitude. Au bout de quelques mois, elle aurait pleuré si on l'avait retirée de l'asile. Toujours à cause de l'habitude. C'est un peu pour cela que

Demos: Lit (1) (2) | Graded (1) (2) | Novel Science (1) (2) | News NYT WP Rex Speech Adults Kids | Legal | GSL+AWL 1k 2k AWL | FRENCH News Novel Comment [Select](#)

→ WORDS TO HOLD @ 1k (Type/Dbf-click)  
E.g., missed by algorithms →

PROPER NOUN HANDLING ?  
Off-list mid-sentence caps...

COMPOUNDS HANDLING ?  
When Off word is 2 wds...

Class as offlist   
Class as 1k/1c

Class as offlist   
Break apart

**SUBMIT\_window**

<input type="radio"/> BNC-COCA 1-25k Paul Nation	[?]	Lists
> <input type="radio"/> BNC-COCA NFL7 (v.2) Cobb & Laufer	[?]	Lists
> <input type="radio"/> BNC-COCA 1-25 C ("c-series," 100 heads)	[?]	Lists
<input type="radio"/> CLASSIC (GSL/AWL) Nation & Averil Coxhead	[?]	Lists
BNC <input type="radio"/> Fams 20k <input type="radio"/> (f)Lems 25k <input type="radio"/> (f)Lems 50c	[?]	K C
<input type="radio"/> BNC-COCA Core-4 (Dee Gardner)	[?]	Lists
NGSL + <input type="radio"/> NAWL OR + <input type="radio"/> TOEIC OR + <input type="radio"/> BIZ C. Browne	[?]	Lists
<input type="radio"/> IZ - English (New curriculum 2022)		Lists
<input type="radio"/> BNL (Steve Neufeld)	[?]	Lists
<input checked="" type="radio"/> FRENCH <input type="radio"/> 1-25k f(LEMS) <input checked="" type="radio"/> 1-3k FAMS	[?]	Lists

→ FRAMEWORKS →

→ RECAT

↓ COLOUR TEXT ↓ K-LISTS TYPES + extract + mo

**e-Cats + Propers (Off-List Mid-Sentence Capped Words) => 1k:** (28 types / 238 tokens): Albert\_1 II\_2 III\_1 Lyon\_1 MM\_2 Marengo\_5 Masson\_32 Mauresque\_2 Mauresques\_1 Meursault\_9 Péré\_3 Voil\_8

**Processing Notes:** In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class nu

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
<b>K-1</b>	793 (53.8)	2269 (57.23)	29527 ( <u>90.1</u> )	90.1
<b>K-2</b>	413 (28.0)	795 (20.05)	1806 ( <u>5.5</u> )	95.6
<b>Coverage 95</b> 				
<b>K-3</b>	269 (18.2)	413 (10.42)	806 ( <u>2.5</u> )	<b>98.1</b>
<b>Coverage 98</b>				
<b>Off-List:</b>	??	507 (12.79)	619 (1.89)	99.99
<b>Total (unrounded)</b>	1475+?	3965 (100)	32758 (100)	≈100.00

# 8 000 lemmes ou 3 000 familles de mots?

Vocabulaire	Nombre de mots-occurrence	Nombre de lemmes	Nombre de lemmes (nombre cumulé)	Couverture
1 000	29 776	711	711	86,0%
2 000	1 874	449	1160	91,5%
3 000	993	331	1491	94,3%
4 000	518	229	1720	95,8%
5 000	288	163	1883	96,7%
6 000	155	85	2058	97,1%
7 000	172	94	2062	97,6%
8 000	152	82	2144	98,0%
9 000	91	61	2205	98,2%
10 000	70	54	2259	98,5%
11 000	55	41	2460	98,7%
12 000	67	48	2553	98,9%
13 000	43	36	2589	99,1%
14 000	42	38	2643	99,2%
15 000	48	28	2460	99,3%
20 000	98	93	2553	99,7%
25 000	44	36	2589	99,8%
Off list	69	54	2643	100

Fr\_camus\_etrange.txt

↓ COLOUR TEXT ↓ K-LISTS TYPES + extract + mo

-Cats + Propers (Off-List Mid-Sentence Capped Words) => 1k: (28 types / 238 tokens): Albert\_1 II\_2 III\_1 Lyon\_1 MM\_2 Marengo\_5 Masson\_32 Mauresque\_2 Mauresques\_1 Meursault\_9 Père\_3 Voil\_8

Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class nu

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1	793 (53.8)	2269 (57.23)	29527 (90.1)	90.1
K-2	413 (28.0)	795 (20.05)	1806 (5.5)	95.6
Coverage 95				
K-3	269 (18.2)	413 (10.42)	806 (2.5)	98.1
Coverage 98				
Off-List:	??	507 (12.79)	619 (1.89)	99.99
Total (unrounded)	1475+?	3965 (100)	32758 (100)	≈100.00

# Pistes à suivre

- Évaluation empirique auprès d'enseignants et d'apprenants
- La connaissance de ces mots permet-elle de lire et comprendre des textes?
- Connaissance à quelle profondeur?

# Références et lecture recommandée

Bauer, L. & P. Nation. 1993 « Word families ». *International Journal of Lexicography*, 6 (4), p. 253-279.

Cobb, T. & Laufer, B. 2021. The nuclear word family list: a list of the most frequent family members, including base and affixed words. *Language Learning* 71(3), p. 834-871.

Lonsdale, D. & Y. Le Bras. 2009. *A Frequency Dictionary of French*. New York : Routledge.

Nation, I. S. Paul, 2006: How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* no 1. 59–82.

Ramnäs, M. 2019. « Étendue du vocabulaire et compréhension écrite – le français à l'université en Suède ». *Bergen Language and Linguistic Studies*, vol. 10.

Ramnäs, M. & Lindqvist, C. 2021. Le concept de famille de mot appliqué au français : réflexions et simulations. *Les langues modernes* 3, p. 45-53.



[cobb.tom.3@gmail.com](mailto:cobb.tom.3@gmail.com)  
[christina.lindqvist@sprak.gu.se](mailto:christina.lindqvist@sprak.gu.se)  
[marten.ramnas@sprak.gu.se](mailto:marten.ramnas@sprak.gu.se)

**UQÀM**

Faculté des sciences de l'éducation  
**Département de didactique des langues**