

Tuto « e-Manulex-Infra »

Ortega, É., & Lété, B. (2010). eManulex: Electronic version of Manulex and Manulex-infra databases. Retrieved from <http://www.manulex.org>.

Comment faire une requête dans cette base de données,
pour trouver des mots en fonction de leur fréquence
et de leur difficulté orthographique

Tuto rédigé par Marie-Line Bosse, le 08/11/2024.

<https://www.manulex.org/fr/infra/request.html>



Accueil Manulex **Manulex-infra** Téléchargements Liens

Recherche

Recherche par niveau(x) Options des recherches pour plusieurs niveaux Affichage des niveaux

CP
 CE1
 CE2 au CM2
 CP au CM2

Afficher les résultats si ...
 les critères dépendants du niveau sont vrais pour **tous** les niveaux sélectionnés
 les critères dépendants du niveau sont vrais pour **au moins un** des niveaux sélectionnés
 les critères dépendant du niveau sont vrais dans **un niveau sélectionné** :
 CP CE1 CE2 au CM2 CP au CM2

Côte-à-côte Interlacés

Tokens/Types
 Calculs basés sur les tokens Calculs basés sur les types

Pour chercher des mots dans la base entière, cochez « CP au CM2 » et décochez « CP »

Recherche

Recherche par niveau(x) Options des recherches pour plusieurs niveaux Affichage des niveaux

CP

CE1

CE2 au CM2

CP au CM2

Afficher les résultats si ...

les critères dépendants du niveau sont vrais pour **tous** les niveaux sélectionnés

les critères dépendants du niveau sont vrais pour **au moins un** des niveaux sélectionnés

les critères dépendant du niveau sont vrais dans **un niveau sélectionné** :

CP CE1 CE2 au CM2 CP au CM2

Côte-à-côte Interlacés

Tokens/Types

Calculs basés sur les tokens Calculs basés sur les types

Le premier critère qu'on vous propose est la catégorie grammaticale; par défaut, elles sont toutes cochées

● Critères


Réinitialiser les critères

Catégories Grammaticales 

Afficher

<input checked="" type="checkbox"/>	GRAM	Catégories Grammaticales	<input checked="" type="checkbox"/>	→	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
					ADJ	ADV	CON	DET	NC	NP	PRE	PRO	VER	

Décochez-les toutes en cliquant ici, puis choisissez uniquement celles que vous voulez; en cochant ADV, PRE et CON, vous obtiendrez uniquement les mots invariables (adverbes, conjonctions et prépositions)









Catégories Grammaticales 

Afficher

<input checked="" type="checkbox"/>	GRAM	Catégories Grammaticales	<input type="checkbox"/>	→	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
					ADJ	ADV	CON	DET	NC	NP	PRE	PRO	VER	

Les critères suivants portent sur la description des mots

Les coches dans les cases de la colonne de gauche, indiquent quels critères vous voulez voir apparaître dans votre liste. Ici, on veut voir uniquement la forme écrite du mot. C'est normalement suffisant.




Descriptions des Mots  			
Afficher			
<input checked="" type="checkbox"/>	ORTHO	Forme Orthographique	<input type="text"/> Wizard 
<input type="checkbox"/>	PHON	Forme Phonologique	<input type="text"/> Wizard 
<input type="checkbox"/>	PSYLL	Segmentation Syllabique	<input type="text"/> Wizard 
<input type="checkbox"/>	GSEG	Segmentation Graphémique	<input type="text"/> Wizard 
<input type="checkbox"/>	PSEG	Segmentation Phonémique	<input type="text"/> Wizard 
<input type="checkbox"/>	GPMATCH	Associations Graphèmes-Phonèmes	<input type="text"/> Wizard 

Si vous voulez juste générer la liste de tous les mots en fonction de la fréquence et de la difficulté orthographique que vous définirez plus loin, rien à faire ici. Si vous voulez chercher uniquement des mots contenant une suite de lettres particulière, par exemple des mots qui contiennent « en », alors vous pouvez l'indiquer comme ceci:

en = « en » n'importe où dans le mot

en* = « en » en début de mot

*en = « en » en fin de mot

Descriptions des Mots  			
Afficher			
<input checked="" type="checkbox"/>	ORTHO	Forme Orthographique	<input type="text" value="*en*"/> Wizard 

Les critères suivants concernent la longueur des mots souhaitée








Je n'ai rien coché car je n'aurai pas besoin d'avoir la longueur du mot dans ma liste finale.

Mesures de Longueur						Min	Q1	Q2	Q3	Max
Afficher						Statistiques du CP au CM2				
<input type="checkbox"/>	NBLET	Nb de Lettres	> 3	ET	< 9	1	6	8	9	20
<input type="checkbox"/>	NBSYLL	Nb de Syllabes	>	ET	<	1	2	2	3	7
<input type="checkbox"/>	NBPHON	Nb de Phonèmes	>	ET	<	1	5	6	7	16
<input type="checkbox"/>	NBGRAPH	Nb de Graphèmes	>	ET	<	1	5	7	8	17

J'ai indiqué >3 et <9 pour n'avoir que des mots de 4 à 8 lettres.

Ici, à droite, sont indiqués, pour la plupart des critères, les caractéristiques de l'ensemble des mots. Min = valeur minimum (ici, 1 lettre); Max = valeur maximum (le plus grand mot de la base Manulex a 20 lettres);
Q1 = 1^{er} quartile: il y a 25% des mots qui ont une valeur inférieure à celle-là;
Q2 = 2^{ème} quartile: il y a 50% des mots qui ont une valeur inférieure à celle-là;
Q3 = 3^{ème} quartile: il y a 75% des mots qui ont une valeur inférieure à celle-là (et donc 25% qui ont une valeur supérieure).

Les critères suivants concernent la discriminabilité


Mesures de Discriminabilité 							Min	Q1	Q2	Q3	Max		
<i>Afficher</i>							<i>Statistiques du CP au CM2</i>						
<input type="checkbox"/>	NBON	Nb de Voisins Orthographiques	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	4	88786
<input type="checkbox"/>	NBPGN	Nb de Voisins Phonographiques	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	1	84166
<input type="checkbox"/>	NBHP	Nb d'Homophones	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	6	27586
<input type="checkbox"/>	NBHG	Nb d'Homographes	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	0	27586
<input type="checkbox"/>	NBHPNG	Nb d'Homophones non Homographes	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	5	26806
<input type="checkbox"/>	NBHGNP	Nb d'Homographes non Homophones	> v	<input type="text"/>	ET	< v	<input type="text"/>		0	0	0	0	11978
<input type="checkbox"/>													


On ne va pas s'en soucier ici, rien à faire pour une recherche simple.

(pour info: les voisins orthographiques sont tous les mots de même longueur et qui ne diffèrent que d'une lettre avec le mot; les voisins phonologiques, c'est pareil mais que d'un phonème)

Le critère suivant concerne la fréquence des mots, critère fondamental !




Je coche car je veux avoir l'indicateur de la fréquence dans mon tableau final, pour pouvoir classer les mots obtenus par ordre de fréquence.

Fréquence Lexicale (par million) 

		Min	Q1	Q2	Q3	Max
<i>Statistiques du CP au CM2</i>						
Afficher	<input checked="" type="checkbox"/>	0.00	0.02	0.34	2.29	31333.79
U	Fréquence de la Forme Orthographique	> v	2,29	ET	< v	

Je veux des mots très fréquents; j'ai indiqué >2,29, car les données de droite m'indiquent que c'est la valeur Q3: je limite donc ma recherche au quart des mots les plus fréquents.

Les critères suivants concernent la fréquence de parties de mots (syllabes, bigrammes...)

Fréquence des Bigrammes 						Min	Q1	Q2	Q3	Max			
<i>Afficher</i>						<i>Statistiques du CP au CM2</i>							
<input type="checkbox"/>	FRBIGT	MOYENNE (totalité du mot)	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.20	6826.35	10389.60	13742.80	55602.20
<input type="checkbox"/>	FRBIGI	Bigramme du DÉBUT	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.01	1850.93	3752.41	8769.99	52714.40
<input type="checkbox"/>	FRBIGM	Bigramme(s) du MILIEU (moyenne)	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.30	6104.90	9415.95	13102.23	46535.80
<input type="checkbox"/>	FRBIGF	Bigramme de la FIN	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.00	3161.52	17247.53	27945.07	58489.99
<input type="checkbox"/>													
Fréquence des Syllabes 						Min	Q1	Q2	Q3	Max			
<i>Afficher</i>						<i>Statistiques du CP au CM2</i>							
<input type="checkbox"/>	FRSYLI	Syllabe du DÉBUT	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.00	286.03	1446.61	4668.90	47703.87
<input type="checkbox"/>	FRSYLM	Syllabe(s) du MILIEU (moyenne)	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.00	66.90	275.50	734.10	2040.10
<input type="checkbox"/>	FRSYLF	Syllabe de FIN	> v	<input type="text"/>	ET	< v	<input type="text"/>		0.00	93.35	476.18	1628.52	8179.12
<input type="checkbox"/>													

On ne va pas s'en soucier ici, rien à faire pour une recherche simple.

Les critères suivants concernent la consistance

La consistance OP concerne la complexité au niveau de la lecture.

La consistance PO concerne la complexité au niveau de l'écriture, donc de l'orthographe.

Je coche car je veux avoir l'indicateur de la consistance PO dans mon tableau final, pour pouvoir classer les mots obtenus par ordre de difficulté orthographique.

Consistance Orthographe-vers-Phonologie (OP)							Min	Q1	Q2	Q3	Max	
Afficher							Statistiques du CP au CM2					
<input type="checkbox"/>	COGPT	MOYENNE OP (totalité du mot)	> v		ET	< v		3.23	75.11	81.68	87.97	100.00
<input type="checkbox"/>	COGPI	Consistance OP du DÉBUT	> v		ET	< v		0.00	99.31	100.00	100.00	100.00
<input type="checkbox"/>	COGPM	Consistance OP du MILIEU (moyenne)	> v		ET	< v		0.00	69.27	78.75	88.35	100.00
<input type="checkbox"/>	COGPF	Consistance OP de la FIN	> v		ET	< v		0.00	51.35	96.31	99.13	100.00
<input type="checkbox"/>												
Consistance Phonologie-vers-Orthographe (PO)							Min	Q1	Q2	Q3	Max	
Afficher							Statistiques du CP au CM2					
<input checked="" type="checkbox"/>	COPGT	MOYENNE PO (totalité du mot)	> v		ET	< v	62,84	0.07	62.84	71.78	79.81	255.15
<input type="checkbox"/>	COPGI	Consistance PO du DÉBUT	> v		ET	< v		0.00	58.26	98.81	100.00	100.00
<input type="checkbox"/>	COPGM	Consistance PO du MILIEU (moyenne)	> v		ET	< v		0.00	64.37	75.78	86.33	100.00
<input type="checkbox"/>	COPGF	Consistance PO de la FIN	> v		ET	< v		0.00	29.66	37.75	37.75	100.00
<input type="checkbox"/>												

Je veux des mots orthographiquement difficiles, donc « inconsistants »; j'ai indiqué <62,84, car les données de droite m'indiquent que c'est la valeur Q1: je limite ainsi ma recherche au quart des mots les moins consistants dans le sens Phonologie-vers-orthographe.

Je n'ai plus qu'à appuyer sur n'importe quel bouton « envoyer » pour obtenir la liste des mots qui correspondent à tous les critères

Consistance Orthographe-vers-Phonologie (OP)							Min	Q1	Q2	Q3	Max	
Afficher							Statistiques du CP au CM2					
<input type="checkbox"/>	COGPT	MOYENNE OP (totalité du mot)	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	3.23	75.11	81.68	87.97	100.00
<input type="checkbox"/>	COGPI	Consistance OP du DÉBUT	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	99.31	100.00	100.00	100.00
<input type="checkbox"/>	COGPM	Consistance OP du MILIEU (moyenne)	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	79.27	78.75	88.35	100.00
<input type="checkbox"/>	COGPF	Consistance OP de la FIN	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	51.35	51.31	99.13	100.00
<input type="checkbox"/>												
<input type="checkbox"/>												
Consistance Phonologie-vers-Orthographe (PO)							Min	Q1	Q2	Q3	Max	
Afficher							Statistiques du CP au CM2					
<input checked="" type="checkbox"/>	COPGT	MOYENNE PO (totalité du mot)	> ▾	<input type="text"/>	ET	< ▾	62,84	0.07	62.84	71.78	79.81	255.15
<input type="checkbox"/>	COPGI	Consistance PO du DÉBUT	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	68.26	98.81	100.00	100.00
<input type="checkbox"/>	COPGM	Consistance PO du MILIEU (moyenne)	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	64.37	75.78	86.33	100.00
<input type="checkbox"/>	COPGF	Consistance PO de la FIN	> ▾	<input type="text"/>	ET	< ▾	<input type="text"/>	0.00	29.66	37.75	37.75	100.00
<input type="checkbox"/>												
<input type="checkbox"/>												

Envoyer

Envoyer

Tous mes critères sont récapitulés ici.

Voici les 46 mots, avec toutes les informations que j'avais cochées. J'ai un bouton pour exporter la liste si je veux

Je sais que j'ai obtenu 46 mots.

Retour | Ecran OFF | Réinitialiser le Tri | Export fichier (Inclure l'en-tête) | Code ASCII

CP au CM2:
SYNTH IS IN (ADV, CON, PRE)
3 < NBLET < 9
AND U > 2.29
AND COPGT < 62.84

Résultats 1 – 46 sur 46
Nombre de Résultats | Changer

Précédente | 1 | Suivante

GRAM	ORTHO	U	COPGT
ADV	ailleurs	91.32	61.08
ADV	ainsi	348.32	47.61
ADV	aisément	4.39	54.24
PRE	après	683.99	59.44
ADV	après	126.31	59.44
ADV	assez	233.70	35.86
ADV	aussitôt	170.87	57.62
ADV	beaucoup	574.70	54.07
ADV	bientôt	243.83	57.59
ADV	certes	11.93	58.23
ADV	chaud	85.12	45.57
PRE	chez	638.46	55.23
ADV	combien	255.68	61.62
CON	comme	1921.43	50.59
ADV	comme	383.19	50.59
ADV	comment	537.43	46.85
CON	comment	274.14	46.85
ADV	déjà	351.69	51.88
PRE	derrière	209.56	55.55
ADV	derrière	96.95	55.55
ADV	enfin	323.36	62.00
ADV	épais	3.73	45.60

Je peux trier les mots ici : en cliquant 1 fois sur U (la fréquence), par exemple, je les classe en fonction de ce critère, du moins fréquent au plus fréquent (de la valeur la plus petite à la plus grande).
 En cliquant une 2^{ème} fois sur U, j'inverse le classement, cette fois le plus fréquent apparait en haut de la liste.

CII

Précédente | 1 | Suivante

GRAM	ORTHO	U	COPGT
ADV	ailleurs	91.32	61.08
ADV	ainsi	348.32	47.61
ADV	aisément	4.39	54.24
PRE	après	683.99	59.44
ADV	après	126.31	59.44
ADV	assez	233.70	35.86
ADV	aussitôt	170.87	57.62
ADV	beaucoup	574.70	54.07
ADV	bientôt	243.83	57.59
ADV	certes	11.93	58.23
ADV	chaud	85.12	45.57
PRE	chez	638.46	55.23
ADV	combien	255.68	61.62
CON	comme	1921.43	50.59
ADV	comme	383.19	50.59
ADV	comment	537.43	46.85
CON	comment	274.14	46.85
ADV	déjà	351.69	51.88
PRE	derrière	209.56	55.55
ADV	derrière	96.95	55.55
ADV	enfin	323.36	62.00
ADV	épais	3.73	45.69

CII

Après 1 clic

Précédente | 1 | Suivante

GRAM	ORTHO	U ▲1	COPGT
ADV	ouais	2.47	52.41
ADV	épais	3.73	45.69
ADV	aisément	4.39	54.24
ADV	frais	4.86	62.27
ADV	plein	6.15	60.32
ADV	sitôt	7.40	61.54
ADV	gaiement	9.04	48.00
ADV	long	9.31	60.30
ADV	certes	11.93	58.23
ADV	tantôt	20.93	56.64
ADV	grand	37.06	62.06
ADV	faux	44.70	43.36
PRE	plein	45.58	60.32
ADV	guère	46.09	35.21
ADV	sûrement	64.84	52.96
ADV	chaud	85.12	45.57
ADV	ailleurs	91.32	61.08
PRE	grâce	92.90	50.87
ADV	derrière	96.95	55.55
ADV	près	97.19	60.78
ADV	haut	102.97	53.54
ADV	après	126.31	59.44

CII

Après 2 clics

Précédente | 1 | Suivante

GRAM	ORTHO	U ▼1	COPGT
CON	mais	3205.27	53.85
CON	comme	1921.43	50.59
ADV	très	1709.90	60.59
CON	quand	1464.14	34.62
PRE	sans	1343.85	58.65
ADV	même	733.86	53.43
PRE	après	683.99	59.44
PRE	chez	638.46	55.23
ADV	beaucoup	574.70	54.07
ADV	trop	558.86	59.70
ADV	comment	537.43	46.85
ADV	comme	383.19	50.59
ADV	déjà	351.69	51.88
ADV	ainsi	348.32	47.61
ADV	enfin	323.36	62.00
PRE	près	303.15	60.78
CON	comment	274.14	46.85
ADV	combien	255.68	61.62
ADV	bientôt	243.83	57.59
ADV	vraiment	240.33	60.16
ADV	assez	233.70	35.86
ADV	tant	220.45	60.83

Si la liste ne me convient pas, j'utilise le bouton « retour » pour revenir en arrière et modifier mes critères

Retour Plein-écran OFF Réinitialiser le Tri Export fichier Inclure l'en-tête Code ASCII

CP au CM2:
SYNTH IS IN (ADV, CON, PRE)
3 < NBLET < 9
AND U > 2.29
AND COPGT < 62.84

Résultats 1 – 46 sur 46

Nombre de Résultats

Précédente | 1 | Suivante

GRAM	ORTHO	U	COPGT
ADV	ailleurs	91.32	61.08
ADV	ainsi	348.32	47.61
ADV	aisément	4.39	54.24
PRE	après	683.99	59.44
ADV	après	126.31	59.44
ADV	assez	233.70	35.86
ADV	aussitôt	170.87	57.62
ADV	beaucoup	574.70	54.07
ADV	bientôt	243.83	57.59
ADV	certes	11.93	58.23
ADV	chaud	85.12	45.57
PRE	chez	638.46	55.23
ADV	combien	255.68	61.62
CON	comme	1921.43	50.59
ADV	comme	383.19	50.59
ADV	comment	537.43	46.85
CON	comment	274.14	46.85
ADV	déjà	351.69	51.88
PRE	derrière	209.56	55.55
ADV	derrière	96.95	55.55
ADV	enfin	323.36	62.00
ADV	épais	3.73	45.60

Bonne utilisation de « Manulex-Infra »

Les données des bases Manulex et Manulex-infra (ce qui inclut les résultats des requêtes sur e-manulex) sont mises à disposition selon les termes de la [licence Creative Commons Paternité - Pas d'Utilisation Commerciale – Partage des Conditions Initiales à l'Identique 3.0 Unported](#).

Toute utilisation en dehors du champ de cette licence doit faire l'objet d'une autorisation préalable.

Ecrire à bernard.lete@manulex.org