

ANALYSE ITRAMEUR BONHEUR



Année universitaire : 2023-2024

I. Introduction

Dans l'univers infini qui constitue le langage, le mot "bonheur" émerge comme une entité complexe, englobant une multitude d'émotions et de perceptions. Grâce au logiciel iTrameur (<https://itrameur.cilliac-arp.univ-paris-diderot.fr/>), nous avons l'opportunité d'explorer ce mot dans la langue française. Pour cela, nous allons tenter de comprendre les différentes nuances de ce mot, mais aussi comment il se tisse dans notre quotidien.

En effet, nous avons décidé de travailler sur des forums en ligne afin d'avoir un corpus qui est enrichi par la vision des gens. Les forums en ligne sont un endroit idéal pour que les gens s'expriment librement et comment ils échangent avec les autres utilisateurs.

En nous plongeant dans les différentes discussions, nous allons pouvoir explorer les diverses perspectives qui convergent au sein de notre société : la recherche du bonheur.

Ainsi, nous allons tenter de comprendre comment le “bonheur” se manifeste dans le langage.

II. Analyse iTrameur

A) Vérification de notre corpus

Avant de commencer notre analyse, nous devons voir si notre script a bien fonctionné, c'est-à-dire mettre les balises correspondantes avec le délimiteur de section. Notre fichier contexte et dumps-text doit avoir ce format :

```
<lang="fr">
<page="fr_1">
<text>Ici, le contenu du fichier fr_1,</text>
</page> §
<page="fr_2">
<text>Là, le contenu du fichier fr_2</text>
</page> §
</lang>
```

Lorsqu'on charge notre corpus, nous pouvons voir que le logiciel iTrameur a bien segmenté notre fichier et il a bien délimité tous nos fichiers. Nous pouvons voir ces carrés qui apparaissent pour voir les différentes pages de notre corpus.

Segmentation terminée (dumps-text_fr.txt :139676 occurrences / 14194 formes)
Carte des sections avec délimiteur de sections : §

Clic sur section : affichage contexte | Clic-droit sur section : sélection section

[illegible]

Ensuite, nous pouvons voir si chaque carré a bien du texte à l'intérieur. Pour cela, nous devons vérifier si dans chaque carré nous avons notre texte/body de chaque lien.

SECTION N°34

Doctissimo : Santé et bien être avec Doctissimo Doctissimo Forums _____ Go Forums Grippe * Jeune intermittent * Diabète type 2 * IMC * Toux grasse * Kamasutra Santé * Coronavirus Coronavirus Actualités coronavirus en direct + Tout savoir sur le coronavirus + Covid 19 : gardez la forme + Prenez soin de vous + En famille + Vaccin anti-Covid 19 * Santé Santé Forums Santé Santé * Médicaments Santé Forums Médicaments Médicaments * Médecines douces Santé Forums Médecines douces Médecines douces Famille * Grossesse Famille Forums Grossesse Grossesse * Bébé Famille Forums Bébé Bébé * Famille Famille Forums Famille Famille * Maison de retraite * Animaux Famille Forums Animaux Animaux Bien-Être * Beauté Bien-Être Forums Beauté Beauté * Forme Bien-Être Forums Forme Forme * Nutrition Bien-Être Forums Nutrition Nutrition * Psychologie Bien-Être Forums Psychologie Psychologie * Sexualité Bien-Être Forums Sexualité Sexualité Autres * Actualités Actualités Actualités + Santé + Médicaments + Grossesse bébé + Nutrition + Beauté + Psycho + Sexualité + Forme + Famille * Vidéos * Podcasts * Tests * Informations Juridiques Espace membre * Se connecter * Club Doctissimo * Le Lab SantéGrossesse BébéBeautéNutritionPsychologieSexualitéMédicamentsForme SportVie pratiqueAnimauxFamille 1. psychologie 2. / 3. Bonheur et joie de vivre le bonheur existe t'il? 17 réponses / Dernier post : 05/02/2019 à 22:17 P Profil supprimé 24/01/2019 à 22:42 (BUTTON) le bonheur existe t'il???...alors si je lance se sujet vous vous doutez bien que c'est par expérience que je pose la question. (ou par inexpérience du bonheur !) donc la question est de savoir si un vrai bonheur sur la durée comme le voudrait sa définition est il réalisable? dans la philosophie certain comme schopenhauer ou pensent qu'il n'existe pas (oui il s'agit là d'un pessimiste) car la vie étant une succession de souffrance, il faudrait soit avoir un désintérêt total pour tout, soit réussir à outrepasser cette pendule qui s'agit "entre l'ennui et la souffrance" (youpi!!) dans le bouddhisme on trouve l'idée que le bonheur s'attend avec l'oubli de son "moi" de sa conscience pour pouvoir traverser l'existence sans trop de peine, en s'enlevant les souffrances inévitables... puis dans les doctrines spiritistes, ésotéristes ou bien religieuses on a des personnes pour nous dire que l'on aura une mission sur terre et que notre bonheur dépendrait de sa réalisation (oui alors là...) bien sûr dans la recherche du bonheur on peut simplement penser à tout les plaisirs, corporels , matériels que nous offrent nos sociétés consuméristes...(on parle là plutôt de plaisir car ces attractions tel que l'alcool, le sexe , les stupéfiants, la consommation de bien matériels sont là des "drogues" qui nous épuisent et nous donnent des émotions fugaces mais très agréables qui ne seraient pas compatible avec le bonheur) et vous, quel est votre avis? avez vous trouvé le bonheur ? si je pose cette question c'est bien sûr car je me pose souvent la question de façon désespéré. mais c'est aussi pour la vertu philosophique, scientifique etc... (BUTTON) 0 (BUTTON) Répondre (BUTTON) Voir la dernière réponse Your browser cannot play this video. P Profil supprimé 25/01/2019 à 07:10 (BUTTON) On érige aujourd'hui le bonheur en maître étalon d'une existence! Le bonheur est un sentiment fugace qui n'est pas sur le long terme... Aujourd'hui on se doit d'être heureux. lire le livre

En cliquant sur nos carrés, nous pouvons voir qu'il y a bien du contenu textuel à l'intérieur de chaque balise <page></page>. En effet, cette étape est très importante car cela nous permet de faire une analyse concrète et non biaisée avec du contenu qui s'affiche mal.

Ensuite, nous pouvons voir dans l'onglet "Cadre", puis dans l'option "Cadre" l'arborescence de notre fichier.

lang=fr <1:450729>

page=fr_10 <2:11709>

text=text <3:11708>

page=fr_11 <11712:15180>

text=text <11713:15179>

page=fr_12 <15183:17208>

text=text <15184:17207>

page=fr_13 <17211:22613>

text=text <17212:22612>

page=fr_14 <22616:29521>

text=text <22617:29520>

Nous pouvons voir que notre arborescence dans le fichier est correcte et que dans la balise langue nous avons toutes nos pages, et dans chaque page nous avons notre texte.

Lorsque nous avons exécuter notre script pour iTrameur, nous avons rencontré des difficultés au niveau des balises et au niveau du contenu. Par exemple, nous avons remarqué que la première page était bien <page="fr_1">, mais ce n'était pas le contenu de fr_1 mais de fr_10. Nous avons pu modifier cela, en mettant bien le contenant pour chaque lien , mais cela à mis pages dans le désordre. Nous pouvons le voir parce que le premier fichier de page est bien fr_10.

B) Dictionnaire

Après notre vérification de corpus terminée, nous avons décidé de nous diriger dans l'onglet "Trame", puis dans "Dictionnaire".

Le dictionnaire va nous permettre d'afficher toutes les formes de notre mot selon la fréquence : du plus fréquent au moins fréquent.

Avec l'option du dictionnaire, il est possible de projeter chaque item sur :

- la concordance
- la ventilation
- la carte des sections
- la sélection

Dans notre corpus, nous pouvons voir que le mot "bonheur" apparaît de plusieurs manières : au singulier et au pluriel, mais aussi en majuscule et en minuscule.

Dictionnaire

Recherche : bonheur

Item	Fq	Concordance	Ventilation	Carte	Sélection
bonheur	1130				
Bonheur	74				
bonheurs	22				
BONHEUR	5				
bonheur...	2				
>bonheur	1				
'bonheur'	1				
bOnHeUr	1				
bonheur... Anonyme	1				

Nous pouvons justifier ces différentes formes par le fait que ce soit un corpus venant de forums en ligne, donc les utilisateurs écrivent de la manière qu'ils souhaitent. Par exemple, nous pouvons voir la forme "bOnHeUr" ou encore ">bonheur". Ce sont des facteurs qu'il faut prendre en compte parce que gens sont libres d'écrire ce qu'ils veulent et comment ils veulent avec le clavier : avec des fautes de frappe ou d'orthographe.

Mais aussi, nous allons nous concentrer sur le mot "bonheur" écrit en minuscule et singulier parce qu'il apparaît 1130 fois. Mais aussi, "Bonheur" apparaît 74 fois mais à l'aide de notre concordancier développé sur notre site, nous avons vu que ce sont des occurrences qui apparaissent au titre de notre forum, tels que : "#La perte du Bonheur" ou encore "#Discussion: Bonheur".

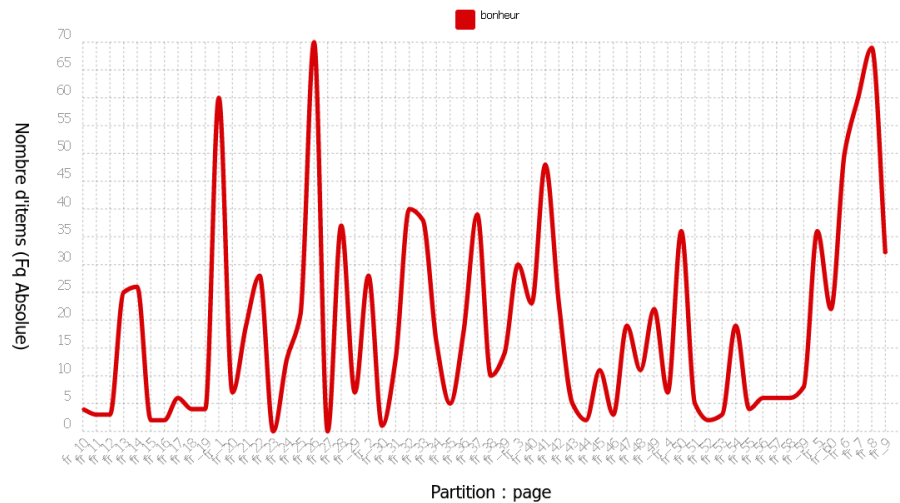
C) Ventilation

Nous allons nous intéresser à l'onglet "Cadre", puis dans l'option "Ventilation". La ventilation nous permet d'avoir trois graphiques. Le premier graphique (en haut) est réalisé en fonction de la fréquence absolue du pôle recherché. Ce graphique sera très intéressant pour notre analyse parce que il est réalisé en fonction du nombre d'occurrences de notre forme du mot dans le corpus. Nous avons décidé de choisir la partition pages dans les paramètres mais il est possible de faire la partition sur d'autres facteurs. La partition selon les pages va nous permettre de d'avoir en vu toutes les pages sur l'axe horizontal. Mais aussi, nous allons voir combien de fois notre page contient le mot "bonheur".

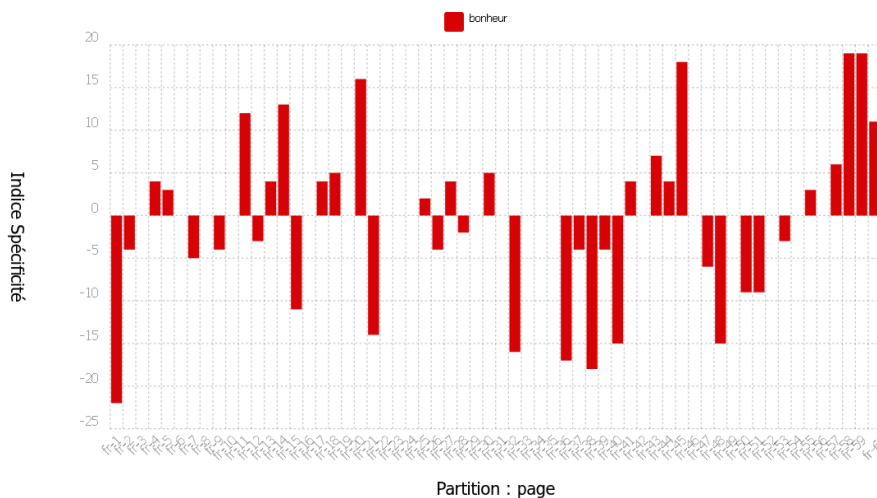
Dans le premier graphique nous pouvons voir qu'il y a des différences au niveau de l'apparition de notre mot.

Par exemple, dans la page fr_8, nous avons 74 occurrences, dans fr_26 nous avons 71 occurrences, et dans fr_1 on en a 60.

Nous pouvons donc voir que dans notre corpus, nous avons des pages qui ont généré beaucoup de discussion entre les utilisateurs car le nombre d'occurrences apparaît beaucoup de fois.



Au milieu, nous avons la fréquence relative, mais notre graphique ressemblait à celui de l'absolue donc dans notre analyse cela ne semblait pas intéressant car nous voulions expliquer de nouvelle chose et non nous répéter.



Pour finir, nous avons le graphique de spécificité qui nous permet d'analyser la spécificité d'apparition d'un événement textuel dans une partie d'un corpus plutôt qu'une autre. Nous pouvons voir que les pages que nous citons auparavant sont incluses dans les indices positifs et non dans le négatif.

Nous pouvons observer que les pics les plus hauts deviennent positifs et les pages qui ont moins d'occurrences sont négatives.

D) Cooccurents

Dans cette partie nous allons dans l'onglet "Coocs", puis dans "Cooccurents"

L'onglet "Coocs" va nous permettre de rechercher les cooccurents de notre mot, entré dans les paramètres en cliquant sur le bouton "cooccurents*". Le calcul de cooccurrence est réalisé en prenant en compte les cooccurents dont l'indice de spécificité est supérieur à la valeur, c'est-à-dire dans nos graphiques.

En traitement automatique du langage, l'analyse de la cooccurrence peut être utilisée pour extraire des relations sémantiques entre des mots et en examinant les patterns dans lesquels ils se trouvent. Ainsi, cela peut nous permettre de comprendre les associations et les liens entre le mot choisi.

Tout d'abord, nous allons voir les mots qui apparaissent le plus souvent près du mot "bonheur". Pour cela, nous avons pris des captures de la table en nous concentrant sur les différentes fréquences mais aussi sur l'indice de spécificité qui est supérieur.

(cliquez sur mot : contextes)

Cooc	FqCooc	CoFreq	IndSP
le	2552	561	**
te	470	184	**
du	1101	267	**
est	2289	370	**
quoi	132	54	32
mesure	41	31	31
Définition	22	22	28
c	877	134	25
Dents	32	22	21
Re	123	39	19
parles	30	20	18
Qu	43	21	15
définition	28	17	15
dents	219	45	14
i	154	35	13

absolu	16	12	13
atteindre	34	16	12
argent	51	19	12
recherche	44	16	10
Pour	125	27	10
existe	46	16	10
notion	15	10	10
moi	327	48	9
accessible	15	9	9
pour	1062	105	8
Message	92	20	8
imposer	9	7	8
que	1904	163	7
peut	289	38	7
on	841	83	7
c	279	37	7
par	578	64	7
au	409	49	7

ultime	4	4	6
conception	6	5	6
construit	6	5	6
n	662	64	5
pas	1481	123	5
quête	20	7	5
son	200	26	5
traque	3	3	5
mesurer	17	6	5
définit	5	4	5
simple	37	9	5
compatible	3	3	5
égoïste	5	4	5
Notre	3	3	5
tremplin	3	3	5
tabou	6	4	5
perception	6	4	5
journaliers	3	3	5

Tout d'abord, le premier mot qui a retenu notre attention est "Dents". Au début, nous nous sommes demandé pourquoi le mot "dent" avait un indice de spécificité important pour être placé dans les mots les plus importants. Nous avons donc cliqué sur ce mot pour avoir le contenu des phrases. Lorsqu'on clique dent :

Re : Dents du bonheur tu parles! coucou

Re : Dents du bonheur tu parles! Hello

Re : Dents du bonheur tu parles! si

Re : Dents du bonheur tu parles!

Re : Dents du bonheur tu parles!

261 Dents du bonheur = dents écartées.

Les "Dents du bonheur" Problèmes dentaires

J'avais les dents du bonheur et d'après ce cher

mes dents du bonheur. Et je reste assurée

quoi des "dents du bonheur"??? [

'ai les dents "du bonheur" comme beaucoup de

qui avait les dents du bonheur et très jeune

J'avais les dents du bonheur quand j'étais petite et

mes dents du bonheur, je trouve ça sympa,

dis souvent que dents du bonheur ou grains de

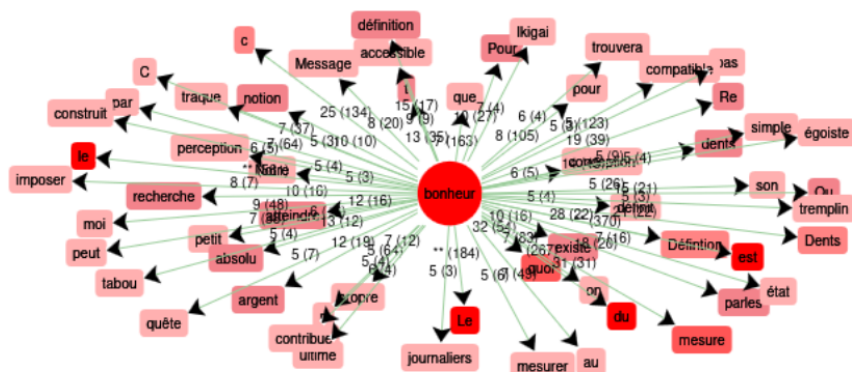
Lorsque l'on pense au mot "bonheur" on peut ne pas penser au mot "dent" mais en remarquant les phrases on comprend mieux pourquoi "dents" est placé dans ce tableau. Nous pouvons voir que "dents" fait référence au dent de bonheur, et les utilisateurs demandent des renseignements ou expriment leur point de vue/ expérience sur les dents du

Ensuite, nous pouvons voir que les premier résultats sont des déterminants ou bien des prépositions, mais cela est tout à fait normal dans la langue française. Ensuite, nous pouvons voir “quoi” et “atteindre”. Nous pouvons voir que ces deux apparaissent le plus souvent parce que lorsqu’on regarde leur contexte d’apparition leurs résultats sont intéressants.

on peut pas atteindre le bonheur = idéal mais on peut

quoi mesure-t-on le bonheur ? Salut,

Par ailleurs, le verbe être “est” apparaît avant car les gens communiquent en définissant pour eux ce que c’est le bonheur. Ce schéma ressemble à un cycle parce que les gens se demandent ce qu’est le bonheur, comment l’atteindre ou encore comment peut-on le mesurer. Mais aussi, après avoir écrit les questions sur les forums, les autres peuvent définir ce qu’est pour eux le bonheur.



Ensuite, nous pouvons trouver une liste de mots qui correspond à un champ lexical de l'exploration et de la quête, tels que : “recherche”, “conception”, “ultime”, “notion”, “quête”, “perception”, “propre” etc...

Mais aussi, un champ lexical plutôt négatif avec “argent”, “égoïste” ou encore “tabou”.

Ainsi, nous pouvons voir que le bonheur est un terme riche sémantiquement car il englobe une gamme complexe de mots. Pour certain elle peut représenter une moment de satisfaction mais pour d'autre c'est bien plus. Ce mot incarne une quête constante et prend en compte plusieurs facteurs dans la vie humaine. Cela passe par un sentiment individuel de bien être, à la recherche d'accomplissement. Sa polysémie se manifeste à travers la diversité des expériences humaines qui contribuent à la recherche de cet état.

Ainsi, explorer ces nuances nous ont permis de montrer les multiples facettes de ce mot et de voir qu'il contient une richesse lexicale et une complexité à cause de l'expérience des individus.

E) Spécifs-totales* et Spécifs-partie*

Pour finir, cette analyse nous avons voulu voir l'onglet “Cadre” et les options “Spécifs-totales*” et “Spécifs-partie*”.

Item	FQ	fr / fq	fr / sp
bonheur	1130	3390	0
Bonheur	74	222	0
bonheurs	22	66	0

Item	FQ	fq	Sp
bonheur	1131	1131	0
Bonheur	72	72	0
bonheurs	22	22	0

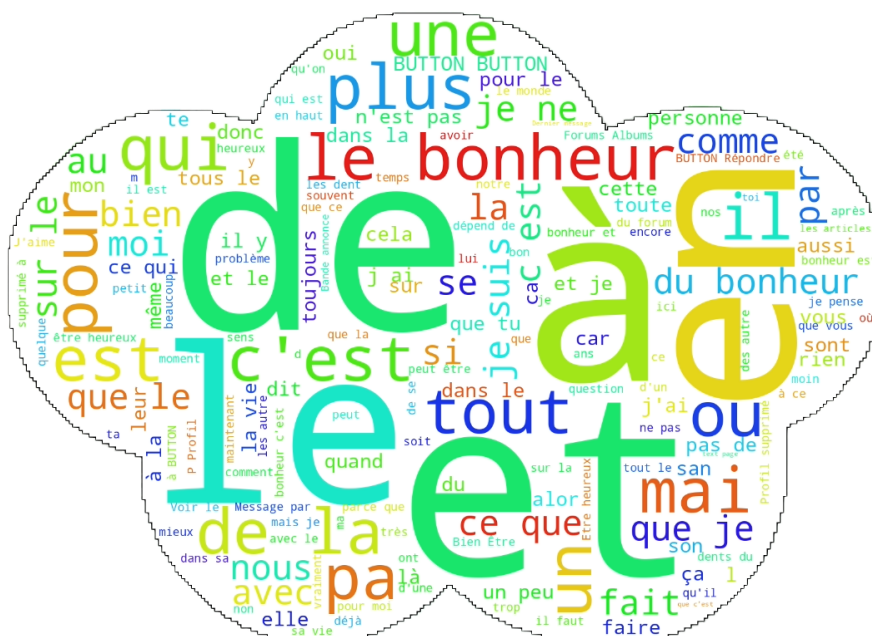
Ces deux options nous permettent de confirmer nos hypothèses du début de notre analyse. Le mot “bonheur” est utilisé : “bonheur”, “Bonheur” ou bien “bonheurs”. Les autres configurations que nous avons vu correspondent surement au faute de frappe ou bien au choix de l'utilisateur.

III. Nuage de mot

Dans cette partie , nous allons observer le nuage de mots que nous avons fait avec le logiciel wordcloud_cli. Voici les commandes que nous avons fait après avoir créé un environnement virtuel et installé la librairie :

```
wordcloud_cli --text itrameur/contextes_fr.txt --image contextes_fr.png --scale 3 --mask
cloud.png --background white --contour_width 1 --contour
    color black
```

```
wordcloud_cli --text itrameur/dumps-text_fr.txt --image dumps-text_fr.png --scale 3 --mask
cloud.png --background white --contour_width 1 --contour_color black
```



Voici le résultat du nuage de mot qu'on a généré à partir de nos dumps-text qui regroupent tous notre contenu HTML de chaque page. Nous pouvons voir que ce résultat affiche plus les stopwords du français, mais aussi les différentes terminaisons du mot "bonheur". Nous pouvons voir que dans les forums le mot "bonheur" était central et que ce mot a généré beaucoup de discussion autour de cette symbolique. Nous pouvons voir des mots tels que "vie", "heureux" ou "personne". Nous pouvons voir que les utilisateurs ont utilisé ont expliqué leur point de vue car on voit l'utilisation des pronoms associé à des verbes ou non : "je suis", "c'est", "c est", "elle", "j'ai", "sont", "vous", "moi" etc...

Ensuite, nous nous sommes plus concentrés sur le contexte parce que il était généré pour avoir le contexte gauche et droit de notre mot.



Même si, nous retrouvons les stopwords du français, nous avons aussi des résultats qui confirment de nouveau notre analyse. Tout d’abord, on retrouve bien les cooccurents qu’on a vu précédemment “dents”, “heureux” , “un état”, “mesure”, “simple”, “joie” etc... Si on ne compte pas le mot “dents”, on voit que ce mot est entouré de mot considéré comme positif. Par ailleurs, il peut générer des questionnements, car nous retrouvons “définition du”, “question”, “c’est quoi “, “je trouve” etc... Les personnes se demandent comment on pourrait atteindre le bonheur mais surtout se demander qu’est-ce que c’est.

De plus, nous pouvons voir que “Re” apparait. Ce token correspond aux discussions qui ont été générées par les utilisateurs, donc les utilisateurs ont répondu au questionnaire des personnes ou bien des discussions en rapport avec le bonheur.

Attention, il ne faut pas oublier que ce mot à aussi son antonyme qui est “malheur” et où il peut aussi apparaître et représenter aussi un champ lexical négatif.

IV. Script Python : PALS

A
titre
personnel
parce
que
je
suis
plus
au
faits
pour
parler
de
ce
que
je
sais
me
concernant
le
bonheur
tel
que
je
le
conçoit
aujourd'hui

Pour finir notre analyse, nous avons décidé de créer un script PALS pour compléter nos différentes analyses sur les autres logiciels : iTrameur et le Wordcloud.

Tout d'abord, nous avons élaboré un script qui permet de reprendre notre second fichier contextes_fr.txt et où une seule ligne correspond à un mot. Une des difficultés était de nous adapter aux caractéristiques de chaque langue, dont les accents, la ponctuation et les caractères diacrités.

Par exemple, en français nous devons prendre en compte tous les accents aigus, graves et circonflexes des voyelles, la cédille de la consonne "c". Ainsi, dans notre rgex nous avons voulu être les plus précises car sinon lorsqu'on lançait les script cela pouvait poser des problèmes à certains endroit.

De plus, nous avons voulu diviser nos phrases afin qu'on puisse mieux se retrouver dans le document. Ainsi, dans notre script on explique que lorsqu'il trouve la ponctuation du point alors il saute une ligne.

Lorsqu'on a obtenu notre document qui nous semblait correct, nous l'avons lancé avec le script Python que les professeurs nous ont fourni et comparer les résultats.

Les résultats que nous avons obtenus pour le français ne nous ont pas beaucoup étonné parce qu'on retrouve pratiquement les mêmes résultats.

```
target» frequency
regex=[Bb][Oo][Nn][Hh][Ee][Uu][RrSs]?» 1212
bonheur»1134
Bonheur»72
BONHEUR»5
bOnHeUR»1

token» corpus size»all contexts size» frequency» co-frequency» specificity
le» 24995» 15566» 895»719»32.18
du» 24995» 15566» 368»323»27.65
Le» 24995» 15566» 211»198»26.13
Re» 24995» 15566» 83» 83» 17.11
BUTTON» 24995» 15566» 41» 40» 7.05
est»24995» 15566» 264»200»6.02
dents» 24995» 15566» 46» 43» 5.94
À» 24995» 15566» 28» 28» 5.76
quoi» 24995» 15566» 63» 55» 5.12
mesure» 24995» 15566» 34» 32» 4.68
t» 24995» 15566» 38» 35» 4.55
Dents» 24995» 15566» 22» 22» 4.53
Définition» 24995» 15566» 22» 22» 4.53
pour» 24995» 15566» 250»184»4.39
définition» 24995» 15566» 21» 21» 4.32
parles» 24995» 15566» 20» 20» 4.12
pas»24995» 15566» 345»242»3.39
```

Tout d'abord, en prenant compte toutes les possibilités que le mot "bonheur" apparaît dans notre corpus, on peut voir que sa fréquence totale est 1212. Mais aussi, que "bonheur" a une fréquence beaucoup plus élevée car il apparaît 1134 fois.

De plus, on peut voir que les tokens sont classés par spécificité, c'est-à-dire avec quel mot "bonheur" est le plus associé. En français, les premiers résultats sont toujours les stopwords. Par exemple, le token "le" a la plus grande fréquence d'apparition avec le mot cible, suivi de "du" etc...

Ensuite, on retrouve de nouveau "Re" qui nous permet d'affirmer que les forums reprenant le thème du "bonheur" a suscité beaucoup d'intérêt chez les utilisateurs car cela a diffusé un échange. Lorsqu'un utilisateur répond à un message, au début du titre il y aura "Re", ce qui signifie "réponse".

Mais aussi, on retrouve de nouveau le mot "dents" qui a une spécificité élevée, 7.05. Par ailleurs, nous pouvons confirmer notre hypothèse car on voit que la fréquence et la co-fréquence sont liées. Le token "dents" apparaît 46 fois, mais ce qui est intéressant c'est que 43 fois le token apparaît en même temps que le mot cible.

Pour finir, on retrouve encore nos différents champs lexicaux avec les tokens comme "mesure", "quoi", "Définition", "est", "état", "vivre", "esprit", "argent", "joie", "absolu" ou encore "malheur". Ainsi, notre mot cible "bonheur" est utilisé fréquemment dans les forums, avec différentes variations, et il est souvent associé à des contextes spécifiques et à différents champs lexicaux.