



Regards et approches pluriels sur la construction des chiffres du bonheur, analyse comparée France/Bénin

<https://icimonbonheur.wordpress.com>

Exercice d'écritures transversales

M1 HUMANITE NUMERIQUE

AÏHOU Laurenda, DOSSOU-YOVO Arnaud
MOLINA Marian & RINGWALD Célian



SOMMAIRE

I.	CONTEXTE.....	3
1.1.	Une question transversale aux sciences humaines	3
1.2.	Un indicateur alternatif du bonheur	4
1.3.	Compter le bonheur à partir des traces du Web	5
1.4.	Du contexte numérique du Bénin et de la France	6
1.5.	Notre approche.....	7
II.	MISE EN DONNEES	7
2.1	Jeux de données Open data :	7
2.2	Jeux de données issues du web :.....	8
2.3	Tableau récapitulatif.....	10
2.4	Modèle Logique et modélisation conceptuelle	11
III.	ANALYSES.....	13
3.1	Analyse des données Twitter	14
3.2	Analyse des données Google CSE.....	16
3.3	Analyse des données issues de Google trends	17
3.1	Analyse des données ESS.....	18
3.2	Analyse des données WHR	19
3.3	Analyse croisée des données ESS et WHR	21
IV.	CONCLUSIONS	22
	BIBLIOGRAPHIE	24



I. CONTEXTE

1.1. Une question transversale aux sciences humaines

Etymologiquement, le bonheur c'est la bonne chance¹. La question du bonheur a dans un premier temps été l'adage des philosophes, offrant alors une vaste vision de la problématique. Selon Epicure il pouvait être issu d'un fait négatif atténué : une diminution de la souffrance, ou issu de quelque chose de positif comme le plaisir ou le sentiment de mener une vie vertueuse. Pour Blaise Pascal le bonheur est inatteignable et ne peut se trouver que face à Dieu. Baruch Spinoza aborde cette notion plus largement en y intégrant l'harmonie de l'homme à la nature. Kant le définit comme un état dépendant de la satisfaction de nos besoins, alors que Nietzsche l'associe au sentiment de puissance... Nombre de visages que revêt cette notion, investie par les religions notamment, qui lui attachent une connotation morale.



Depuis le siècle dernier cette question est passée entre les mains des psychologues et des sociologues. Les premiers la saisissent à travers l'étude de la dépression et de sa pathologisation, dont les symptômes peuvent trouver leurs remèdes grâce à un suivi médical et/ou médicamenteux. La psychanalyse se propose de suivre la vie psychologique des patients atteints de ces maladies dont souffre l'humain puisque névrosé par nature. La psychologie positive se veut de son côté ponctuelle et focalisée sur les patients atteints de traumatismes. Sa méthode consiste à faire réaliser au patient une analyse réflexive afin de pouvoir retrouver un équilibre à travers la réminiscence. Les sciences cognitives prétendent être les premières à mesurer le bonheur de manière objective à travers des mesures physiologiques, mais aussi plus largement émotionnelles et psychologiques.

¹ <http://www.cnrtl.fr/etymologie/bonheur>



A travers la sociologie et l'anthropologie le concept de bonheur rencontre le social et le culturel, les dimensions du bonheur et ce qu'il peut représenter se multiplient. Nous citerons le sociologue Ruut Veenhoven, pionnier en matière d'étude du bonheur. Spécialiste des « hapynness studies » il a regroupé une grande part de ces travaux dans une base de données internationale sur le sujet². Dans une perspective économique, politique et prospectiviste les instituts de sondages et les institutions ne cessent depuis les années 80 d'innover dans la construction d'indicateurs. Ils adoptent des points de vue hétérogènes sur la manière de faire et de mesurer l'indicateur par l'agrégations de résultats de questionnaires qualitatifs ou des indicateurs purement économiques, ayant des buts et des usages différents. Afin d'illustrer ce phénomène nous pouvons citer la World Values Survey (1981) de Gallup Word Poll, institut de sondage historique aux Etat Unis, le Latino Barometer (1996), l'Afrobarometer (1999), Better Life Index (2011) de l'OCDE, ou encore l' European Social Survey (2012).



Cette mise en contexte nous permet de nous apercevoir que le concept de bonheur a beaucoup évoluer à travers les époques. La philosophie du bonheur a préparé le terrain à une psychologie rationnelle mais aussi à une mesure comparative du bonheur global ou spécifique d'une ou plusieurs populations données.

1.2. Un indicateur alternatif du bonheur



En 2013 le Boutan³ a fait l'objet de nombreuses controverses à travers l'institutionnalisation d'un indicateur annuel et régional du bonheur issu de grandes enquêtes à l'échelle de son pays. Les valeurs culturelles du pays ont été mises en avant,

² <https://www.worlddatabaseofhappiness.eur.nl/index.html>

³ <http://www.grossnationalhappiness.com/>

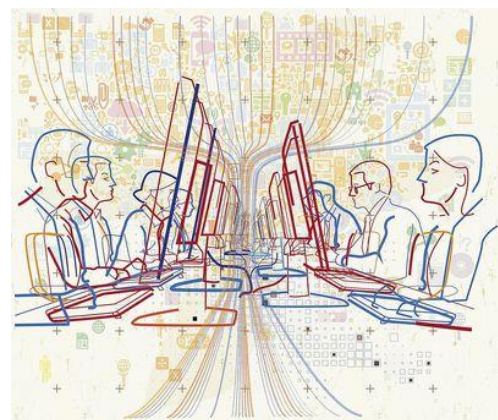


plutôt liées au patrimoine bouddhiste et à contre-pied des grands indicateurs tels que le PIB, ou encore l'IDH produit par les instituts statistiques occidentaux.

Cet indicateur, le Gross National Happiness Index (GNH), est le résultat de la pondération de près de 33 variables d'évaluation de satisfaction réparties sur 9 thèmes adaptés à la définition locale et culturelle du pays. Institutionnalisé au sein de l'appareil d'état, il permet ainsi de prendre en compte le bien-être de ses citoyens. Cependant sa mise en place a aussi été vue comme une politique de communication faisant de la défense d'un style de vie un argument touristique. Le GNH a ouvert un débat concernant les indicateurs économiques de bien-être défendus par les Etats occidentaux tels que le GHI ou le PIB qui ne prennent pas en compte la réalité sociale des populations. Il n'a pas fallu longtemps pour que l'ONU mette en place une étude annuelle (WHR) sur le bonheur subjectif à l'aide de Gallup, qui est renouvelé tous les ans. Décrite, car classant majoritairement des pays du Nord en haut de son classement, cette étude semble avantager une vision occidentale du bonheur. Il est curieux notamment de remarquer que ces pays bien classés possèdent le plus haut taux de suicide, travail que nos collègues se sont attelés à démontrer dans leur projet.

1.3. Compter le bonheur à partir des traces du Web

L'utilisation de données issues du web n'est pas l'unique affaire des indicateurs de bien-être, mais s'inscrit dans une remise en question plus globale et d'actualité qui touche à de nombreux domaines s'intéressant à la « détection de tendances sociales » : les sciences économiques, sociales, et politiques notamment. Il est ici question de traiter les textualités produites par les utilisateurs d'internet : mots-clefs de recherche, commentaires, discussions, échanges sur les réseaux sociaux afin d'obtenir une fenêtre plus large du social. Cette méthode est moins couteuse que les grandes enquêtes, puisque directement disponible avec une connexion internet.



Guido Rosa/Getty Images



Afin de mener notre projet, nous avons souhaité nous lancer dans la collecte d'information de ce type. Nous sommes partis des travaux réalisés par Yann Algan (notamment cité dans le dernier rapport WHR) et son équipe. Son travail se base sur le service Google Trends fournissant des statistiques de recherche sur le moteur de recherche géolocalisé et temporalisé. A travers l'étude économétrique de près de 845 séries temporelles issues de mots clefs (champs sémantiques liés à près de 11 thèmes), l'équipe du MediaLab a pu créer un modèle prédictif qui permet d'inférer les différentes valeurs enregistrées par Gallup autour de 10 thématiques polarisées : la joie, la peur, etc.

1.4. Du contexte numérique du Bénin et de la France

Afin de remettre en contexte nos résultats il est important de se référer à quelques données statistiques. En France, le taux d'équipement mobile est élevé et entre 86% et 94 % des foyers seraient équipés d'une connexion internet selon un récent rapport du gouvernement⁴. Le Bénin a un taux d'équipement mobile oscillant entre 75 et 86% entre 2014 et 2017. D'après un rapport de l'autorité des régulations des communications électroniques et de la poste de 2017⁵, seul 20% à 25% des utilisateurs mobiles possèdent un abonnement internet. Le Bénin possède une véritable richesse linguistique : près de 26 langues différentes y sont parlées, bien que le français en soit la langue officielle, seul 46% savent le lire et l'écrire en 2011⁶.

⁴ <https://labo.societenumerique.gouv.fr/wp-content/uploads/2018/12/barometredunumerique2018.pdf>

⁵ https://arcep.bj/wp-content/uploads/2018/06/Rapport-Annuel-dActivit%C3%A9s-2017_ARCEP-BENIN.pdf

⁶ https://www.insae-bj.org/images/docs/insae-statistiques/sociales/education-alphabetisation/Taux_alphabetisation_adultes_15_ans_et_plus_2004_2011.pdf



1.5. Notre approche

Nous abordons ce projet de manière pragmatique. A travers l'analyse de 6 jeux de données traitant du bonheur subjectif, nous avons souhaité comparer nos pays respectifs : la France et le Bénin, sous plusieurs angles. L'analyse de discours de contenus produits sur un réseau social, l'offre de contenu fournie en première page sur deux moteurs personnalisés de Google, des données issues des interrogations des utilisateurs du même moteur de recherche via Google Trends. Nous traiterons aussi de données d'enquêtes à travers l'étude de données d'un questionnaire issu d'un institut inter-gouvernemental, la European Social Survey, afin de pouvoir mesurer des particularités de ce mode d'étude à travers le cas de la France. Nous comparerons cela avec les résultats mis en avant dans WHR. De cette manière nous avons pu analyser les facteurs de bonheur mis en avant dans ces différentes constructions du bonheur, et aussi évoluer à notre niveau les limites de ces approches Big Data de plus en plus mises en avant.

II. MISE EN DONNEES



Comme précisé dans la première partie nous avons mis en lien 6 jeux de données : le jeu de données fourni dans le rapport WHR 2018, des données issues de l'enquête ESS sur plusieurs années, et 4 autres jeux de données récupérés via des techniques de *scrapping* ou en utilisant des API. Nous les décrivons ici brièvement et présentons la façon par laquelle ils sont liés.

2.1 Jeux de données Open data :

Jeu de données ESS :

Ce jeu de données est issu des enquêtes de **2012**, **2014** et **2016** réalisée en **France** par la *European Social Society*. Il porte sur les réponses de 5955 individus et vise à évaluer le niveau de bien-être subjectif des résidents en France de plus de 15 ans. Les variables sont toutes qualitatives, certaines ordonnées et d'autres nominales. Dans un souci de cohérence nous avons décidé de considérer les variables relatives au bonheur et à la religiosité comme des



variables qualitatives, le faible nombre de modalités rendant les résultats très peu satisfaisants. Le détail des variables étudiés est consultable ici :

<https://drive.google.com/file/d/1gR-2zTtGCJ2VSzZRSDFPCMj7hgoYZSdc/view?usp=sharing>

Source des données : <https://www.europeansocialsurvey.org>

Jeu de données WHR :

Ce jeu de données est issu du rapport sur l'état du bonheur ou du bien-être subjectif dans le monde entre les années 2006 et 2018, réalisé par le Global Happiness Council. L'étude présente 156 pays, nous nous sommes centrés sur les données de la France et du Bénin. Elles sont regroupées en 16 thèmes qui sont entre autres, le pays, l'année, l'expérience de vie en bonne santé à la naissance, le bonheur, la générosité, la perception de la corruption, l'aide sociale, la liberté de faire des choix de vie, la confiance en son gouvernement, la qualité démocratique. Ces indicateurs sont tous quantitatifs et sont le résultat de l'agrégat de données fournies par Gallup.

Source des données : [WHR2018Chapter2OnlineData.xls](#)

2.2 Jeux de données issues du web :

Nous avons choisi de récolter nos données Web à travers le champ lexical français de la joie et de la tristesse, nous nous sommes limités à 12 mots clefs négatifs et 13 mots clefs positifs. Nous avons aussi pris en compte sur Twitter des émoticônes heureux et triste.



Nous calculons à partir des résultats que nous avons obtenu un calcul qui nous permet de créer nous aussi un indicateur alternatif très simpliste :

$$\frac{\sum x_{+}}{\sum x_{+} + \sum x_{-}} * 100$$



Nous comptons alors la part de résultats polarisés positifs sur le total des résultats obtenus pour chacun des jeux de données.

Jeu de données SoWell :

Issu d'un projet du MediaLab Science Po, il est le résultat d'une suite de requêtes liées au bien-être, thématisées sur Google Trends entre 2013 et 2016. Les données sont décrites par régions, et les thèmes adaptés après une étude préalable. Les données sont quantitatives mais reposent sur une méthode d'analyse qualitative.

Google Trends :

Google offre l'opportunité via Google Trends de récupérer le nombre de recherche liées à un mot clef ou à une requête sur une période de temps très fine (au jour) et une échelle géographique que nous avons choisi d'explorer au niveau pays. Nous avons donc questionné l'application à travers nos mots clefs polarisés présentés plus bas sur les périodes 2015, 2016, et 2017 au Bénin et en France. Ce jeu de données nous permet d'en apprendre sur les routines d'utilisations des utilisateurs du moteur de recherche.

Données Google CSE :

Cette application proposée par Google permet de récupérer la liste des résultats fournis par un moteur de recherche personnalisé. Payante à partir de plus de 100 requêtes par jours nous avons donc limité les requêtes de l'API Google CSE aux premières pages retournées pour chacune de nos requêtes concernant nos mots clefs polarisés pour nos deux pays. Nous avons par la suite extrait le contenu des sites web proposés (données uniquement qualitatives : contenus, date de publication, langue).

Données Twitter :

Twitter fournis aussi une API permettant de requêter la base de données du réseau social sur les 30 derniers jours, nous avons créé à cette occasion un script de *scrapping* nous permettant d'interroger l'application très localement en France et au Bénin en fonction des mots clefs polarisés et de récupérer les tweets sur la période 2015-2017 que nous étudions.



2.3 Tableau récapitulatif

DETAILS DES DONNEES

<i>Jeux de données</i>	<i>Périodes</i>	<i>granularité temporelle</i>	<i>granularité géographique</i>	<i>granularité individus</i>	<i>Nombre de variables</i>	<i>Information Langue</i>
ESS	2012, 2014, 2016	Année	Pays	oui	10	X
WHR	2005 à 2018	Année	Pays	non	14	
SoWell	2013 à 2016	Jour	Région	non	8	
Google trends	2015, 2016, 2015	Jour	Région	non	31	X
Google CSE	2015, 2016, 2016	Année	Pays	non	38	X
Twitter	2015, 2016, 2017	Année	Ville	oui	39	X

ANALYSE DES COMMUNS

Jeux de données	Sociabilité	Santé	Famille	aide sociale	esperance de vie	liberté	discrimination	Handicap	securité	Générosité	corruption	Evaluation bonheur	Evaluation malheur	confiance etat	democratie	Indicateurs Economiques	Contenu textuel
ESS	X	X					X	X	X			X					
WHR	X	X		X	X	X				X	X	X	X	X	X	X	
SoWell	X	X	X						X							X	
Google trends												X	X				X
Google CSE												X	X				X
Twitter	X											X	X				X



2.4 Modèle Logique et modélisation conceptuelle

Afin de lier toutes ces informations très hétérogènes nous avons axer notre schéma sur ces axes d'analyses :

- Temporel : avec la table période (regroupement des données au niveau année)
- Géographique : Table Pays et ville
- Linguistique : Table Langue

Ce modèle assez général nous laisse la possibilité future d'ajouter de nouvelles données pour d'autres périodes, d'autres pays...

MODELE LOGIQUE

Cities (city_id, city_name, city_population, city_lat, city_long, country_id#)

Countries (country_id, country_name, population)

ESS_data (ESS_id, genre, annee_naissance, niv_bonheur, sociabilite, sentiment_securite, niveau_sante, handicap, religiosite, discrimination, citoyen, minorite, country_id#, period_id#)

GoogleCSE_data_queries (cse_query_id, nb_total_results, country_id#, keyword_id#, period_id#, language_id#)

GoogleCSE_data_results (cse_results_id, lang_website, index_result, url, title, source, description_google, content, description_meta, country_id#, cse_query_id#)

GoogleTrends_data (GoogleTrends_id, number_of_search, country_id#, period_id#)

Keywords (keyword_id, keyword_label, polarity, origin, lang_id#)

Lang_Country (lang_country_id, official, country_id#, lang_id#)

Lang_ESS (lang_respondant_id, maternal, lang_id#, ess_id#)

Languages (lang_alpha3b, lang_alpha3t, lang_alpha2, name_en, name_fr)

Period (period_id, year)

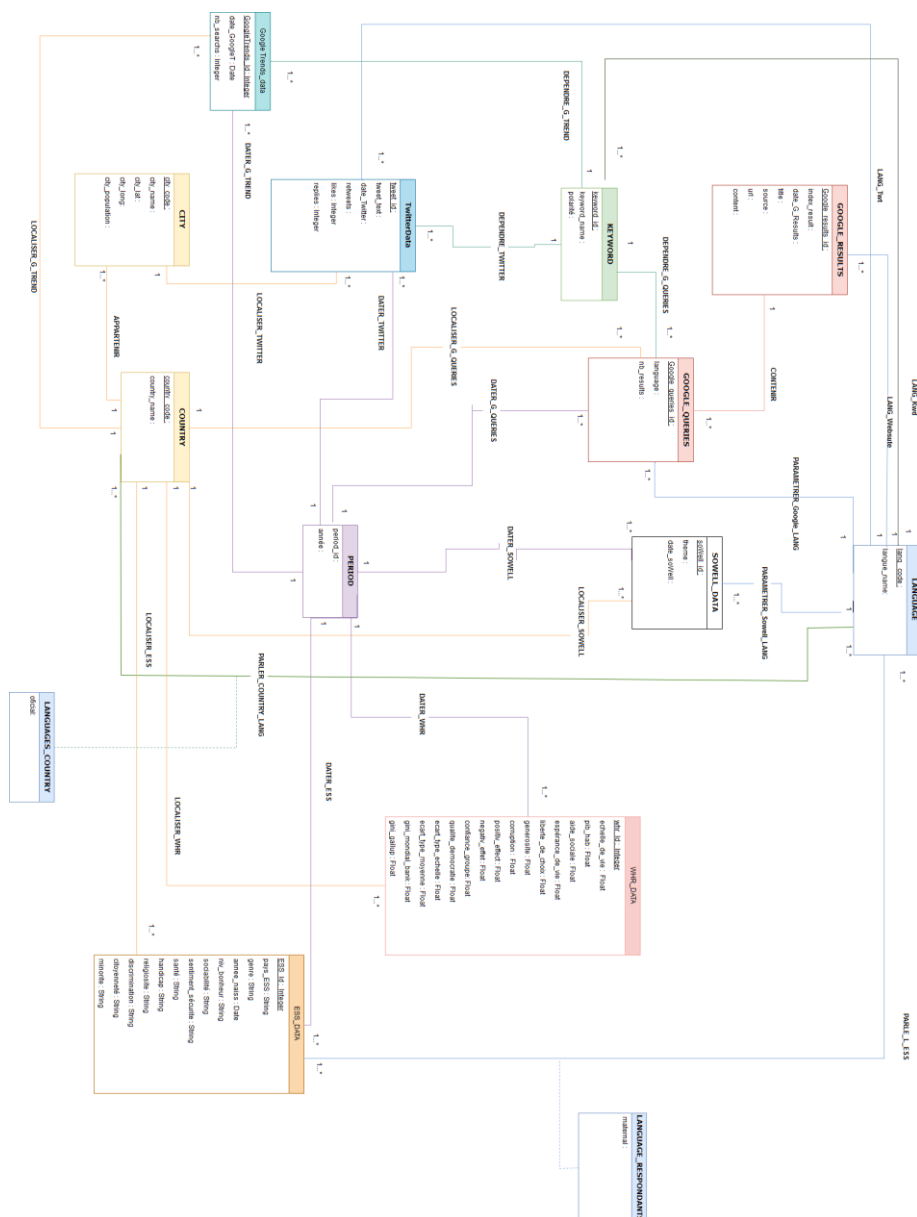
Sowell_data (sowell_id, topic, period_content, number_of_search, country_id#, lang_id#, period_id#)



Twitter_data (tweet_id, date_tweet, tweet_orig_content, retweets, replies, hashtag_contened, content_url, tweet_content_clean, period_id#, city_id#, lang_id#, keyword_id#)

WHR_data (id_whr, echelle_de_vie, pib_hab, aide_sociale, esperance_de_vie_bonne_sante, liberte_choix_de_vie, generosite, corruption, effet_positif, effet_negatif, conf_gouv, qualite_democratie, niveau_confiance, ecart_type_echelle, ecart_type_moyenne, gini_mondial_bank, gini_mondial_bank_2015, gini_gallup, period_id#, country_id#)

MODELE CONCEPTUEL





III. ANALYSES

De nombreux possibles d'analyses se sont présentés à nous durant ce projet, et nous aurions pu porter notre regard sur une infinité de réflexions. Nous nous sommes chacun consacrés à l'étude d'un jeu de données ce qui a finalement donné naissance à la rédaction d'analyses diverses. Un vrai travail d'écriture que ce projet transversal. Les supports dossier et site web n'offrant pas les mêmes libertés nous avons dû revoir la forme et le contenu de nos analyses. Nous les présentons ici de manière succincte mais contextualisée tout en sachant qu'une autre lecture plus complète et illustrée et possible en ligne⁷.

Le premier type d'analyse réalisé concerne deux jeux de données issus de réponses de questionnaires. Celui téléchargé dans la base de données de la *European Social Survey* a pour particularité de regrouper des données relatives directement à des individus humains, un individu statistique correspondant à un résident français. La granularité qui en découle permet d'évaluer le comportement de chaque variable à une échelle humaine. Le jeu de données du WHR quant à lui a pour individus statistiques des pays, les valeurs relatives à chacun étant issu d'une moyenne des résultats obtenus pour un échantillon d'individu de chaque pays. Comme cela a déjà été évoqué précédemment, les deux jeux de données sont issus de questionnaires constitués de questions fermées. Les réponses se font principalement par un positionnement sur une échelle de valeur ou un positionnement binaire, l'enquête est très dirigé donc, ce qui a des implications non négligeables sur l'interprétation des résultats et leur véritable représentabilité de la réalité. La manipulation de ces données a pour but principal de déterminer quelles variables sont les plus à mêmes de rendre compte du bonheur.

En ce qui concerne l'analyse de nos données web, les données de GoogleCSE et Twitter nous ont permises d'entrevoir une fenêtre d'informations textuelles très bien contextualisées mais à la fois hétérogènes et complexes (pages web, conversations). La clef d'entrée de ces corpus est une requête constituée de mots clefs et même si GoogleTrends ne fournit qu'une série temporelle, la stratégie de requêtage que nous avons mise en place nous a permis d'offrir à celui-ci une dimension sémantique, et de posséder un moyen de comparaison avec nos deux autres jeux de données. Chacune de ces sources fournissent des données hétérogènes par leurs structures, leurs contenus et leurs contextes de création. Un aller-retour entre la macroanalyse et la micro-analyse était donc crucial dans notre démarche.

⁷ <https://laurendaaihou.wordpress.com/articles/>



3.1 Analyse des données Twitter



Nous avons donc collecté de manière automatique **15 916** tweets entre 2015 et 2018 en ciblant les 10 plus grandes villes de nos deux pays de référence à travers 14 mots clefs « négatifs » et 15 mots clefs « positifs ». Nous avons pour cela ciblé les productions réalisées dans 10 grandes villes de chacun de nos pays : Paris, Marseille, Lyon, Toulouse, Nantes, Bordeaux, Metz et Lille pour la France et Cotonou, Porto Novo, Abomey-Calavi, Parakou, Djougou, Bohicon. Seulement nous n'avons récupéré que très peu de résultats pour le Bénin (164) contre plus de 14 000 tweets en France. Seulement trois villes nous ont ramenées des résultats : Cotonou, Porto Novo et Abomay-Calavi, dont la majorité des tweets proviennent. Plusieurs hypothèses peuvent expliquer ce résultat : le système de géolocalisation peut ne pas être effectif dans toutes les villes de ce pays via le moteur de recherche Tweeter, les mots clefs choisis peuvent ne pas correspondre aux habitudes langagières des Béninois. Mais nous pensons que cela est sûrement dû au fait que les utilisateurs du réseau sont moins nombreux dans ce pays, du fait de son contexte numérique.

La publication de tweets liés à nos mots-clefs est globalement en baisse depuis 2015, notons que cette année est réellement marquée par les attentats qui ont secoués la France, et cela est aussi observable sur les productions béninoises. Temporellement, le nombre de tweets Béninois produits en un jour est dépendant de la semaine, alors qu'en France cela est plus lié aux événements ayant eu lieu durant le mois courant. Nous avons relevé généralement une hausse de la production entre septembre et décembre et une baisse de celle-ci durant les vacances scolaires, ce qui nous laisse deviner le profil des utilisateurs ayant produits ces contenus : ce sont en général des jeunes. Sans surprise nous avons pu remarquer que les métriques d'engagement du réseau social sont interdépendantes : plus un tweet est *liké* plus il est *retweeté*, et plus les utilisateurs se répondent entre eux plus il y a de tweets. Cependant la polarité (positive ou négative) de ceux-ci a une forte influence sur les réflexes de partage ou de réponse des utilisateurs.

Au Bénin, la plupart des tweets provenait donc d'Abomay, ville Béninoise assez négative sur nos trois périodes : près de 25% du contenu est étiqueté comme négatif. Une partie des messages que nous avons récoltés étant des tweets non francophones que nous avons capté en majeure partie à travers le smiley souriant ou le smiley qui boude, cette partie du corpus est une population qui voyage : allemand, espagnol, anglophone. Une autre partie de ceux-ci était centré sur le « Je », introduit lors de conversation ou encore lors du partage d'un moment de vie : retrouvaille avec des amis, la famille.



Enfin une dernière partie des messages était liée au partage d'idées, de réactions face à l'actualité, se voulant plus globale, peut être aussi plus impersonnelle, liée notamment aux formalités évènementielles : la nouvelle année, Noël... Une AFC réalisée avec Iramuteq nous a permis de nous rendre compte que les tweets négatifs étaient plutôt associés à l'année 2015, l'année 2016 plutôt liée à des démonstrations de joie : le rire, le bonheur partagé, enfin 2017 semble être l'année d'un bonheur calme : plus largement marquée par la prospérité, le fait d'être satisfait, heureux. Finalement nous pensions que le réseau était principalement utilisé au Bénin pour des besoins de communications officielles, cependant seul 15% de nos données peuvent être classées comme tel.

En France, la polarité des tweets est beaucoup plus négative qu'au Bénin, près de 55% de la production de 2015, ce qui va en s'amenuisant années après années. Comme exprimé précédemment c'est un stigmate des attentats de 2015, confirmé par l'analyse lexico-métrique. Il faudra attendre août 2015 avant que la part de tweets positifs dépasse celle négative. Le tableau de bord que nous avons construit à cet usage nous a permis de visualiser la distribution de notre corpus sur une carte. Nous avons remarqué que ce flux négatif provenait majoritairement de Paris. Une analyse plus approfondie du corpus nous aura permis de repérer que ces tweets portaient sur un vocabulaire très lié à la mort, la perte, la famille avec une forte dimension émotionnelle. Nous avons également pu remarquer que Toulouse, Lille et Bordeaux étaient les moins touchés par ce phénomène de négativisme. Finalement, les tweets positifs pouvaient avoir tendance à être plus commerciaux que les autres (occurrences de noms d'entreprises) et une autre partie de ces tweets est plus liée à des conversations entre amis, liées à des moments de vie.

Les smileys sont les mots clefs ramenant le plus de résultats, en France le bonhomme boudeur a tendance à surpasser le smiley souriant, alors que c'est l'inverse au Bénin. La plupart des résultats béninois dans ce sens proviennent de mots positifs, alors qu'en France la tendance est plutôt à l'inverse.

ANALYSE COMPLETE :

https://icimonbonheur.wordpress.com/2019/03/27/analyse_1/



3.2 Analyse des données Google CSE



Nous avons utilisé l'api Google CSE afin de simuler 2 moteurs de recherche, l'un au Bénin et l'autre en France, sur nos périodes temporelles d'analyse 2015, 2016 et 2017 autour de nos mots clefs polarisés. Nous avons ainsi réalisé 167 requêtes auprès de Google, nous avons pour chaque requête, enregistré les 10 premiers résultats renvoyés par le moteur de recherche, ainsi

que l'index de chacun et enregistré le nombre de requêtes estimé par Google même si nous savons qu'il est faux. Mais celui-ci nous permet malgré tout d'évaluer le volume estimé. Par cette analyse nous nous sommes rendu compte de l'offre disponible de contenu liés aux notions du bonheur sur le web. Nous avons pu nous apercevoir que le web francophone indexé par Google était quasiment semblable en France et au Bénin. Pour la plupart ces résultats de première page renvoyaient vers des dictionnaires détaillant les notions requêtées. Nous nous sommes penchés sur les sites « uniquement référencés » pour chacun de nos pays. Au Bénin, Google a tendance à valoriser les contenus commerciaux et liés à la presse, alors qu'en France les solutions de coaching de personnels sont plus présentes, ainsi que le contenu culturel (pas du tout proposé sur le Google Bénin).

Cette analyse nous aura aussi permis de souligner l'ambiguïté de certains termes utilisés : fier est aussi un verbe, la rage peut faire référence à la maladie (faux positif plutôt Béninois), ou faire référence à la rage de dents (faux positif plutôt Français). Enfin les résultats liés au terme « content » n'ont ici aucun sens puisque quasiment lié à son utilisation anglophone.

Malgré la ressemblance des résultats en terme géographique, nous avons pu souligner que tout les ans le moteur de recherche réorganisait ces résultats, ainsi aucune des pages d'une année n'est revenu celle d'après.

L'analyse du contenu des pages fournies nous aura permis d'aller plus loin sur le contexte d'utilisation de nos mots clefs, ainsi le mot « morose » semble être plus associé au social dans son utilisation journalistique plutôt critique dans sa démarche. L'utilisation de la tristesse est dans notre corpus associée à une idée de nostalgie. Enfin nous avons pu découvrir un véritable noyau sémantique articulé autour de la notion de renouveau, de prise en charge psychologique, de prise en main, d'action. Une action pouvant se manifester sur internet par un acte d'achat « satisfait » ou bien « remboursé ».



Nous calculons dans le même sens qu'avec Twitter un indicateur alternatif du bonheur web sur la base du **nombre de résultats estimé par google par polarité pour la France et le Bénin** sur nos trois périodes et remarquons une légère tendance à la baisse pour la France au fil des années, alors que la part de contenu « positif » augmente au Bénin.

ANALYSE COMPLETE :

<https://icimonbonheur.wordpress.com/2019/04/15/analyse-des-donnees-google-cse/>

3.3 Analyse des données issues de Google trends



Introduction

Pour pouvoir mesurer le bonheur au Bénin et en France, nous avons utilisé l'outil de recherche Google Trends pour collecter des données via des mots clés préalablement définis comme des indices de bonheur.

Les cibles de notre étude sont des citoyens du Bénin et de la France qui ont effectués des recherches avec ces mots clés entre 2015 et 2017. Avant d'analyser ce jeu de données, nous avons d'abord présenté les variables. Ensuite nous avons utilisé la méthode d'Analyse de Correspondances Multiples (ACM). Il a été question pour nous d'analyser les variables.

Ces variables sont-elles toutes indépendantes ? Quelles sont les variables qui expliquent le mieux le bonheur ? Il y a-t-il une corrélation entre les variables ?

Résultats

On retient de cette analyse que :

L'étude des coordonnées, de la contribution, des COS^2 des valeurs propres de nos modalités montre que le côté positif de l'AXE 2 est déterminé par les modalités BJ.

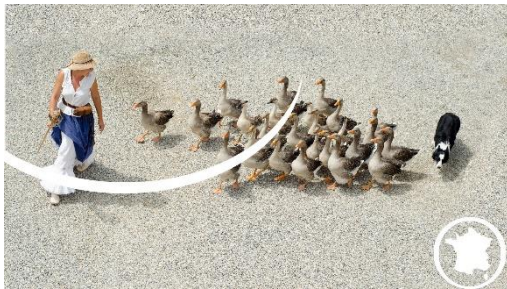
Sur l'Axe 1, les modalités comme positif et négatif arrivent en tête et influencent l'axe. La variable FR influence l'axe 2.

ANALYSE COMPLETE :

<https://icimonbonheur.wordpress.com/2019/04/26/analyse-des-donnees-google-trends/>



3.1 Analyse des données ESS



Dans la logique de notre problématique première nous voulions mettre en avant les déterminants influençant le bien-être, en comparant une variable directement relative à un niveau ressenti de bonheur à d'autres variables annexes tel que le niveau de santé, le handicap ou le fait d'avoir la citoyenneté française ou non. C'est ce

que nous avons fait pour la France avec l'analyse de ce jeu de données. Nous avons étudié la corrélation entre les variables et réalisé une analyse factorielle des données.

Nous avons constaté que le comportement de 7 des 10 variables étaient très lié avec celui de la variable du niveau de bonheur : l'âge, la sociabilité, le sentiment de sécurité, la santé, le handicap, la religiosité et la discrimination. Les variables relatives à la sociabilité, à la santé et au handicap en tête.

Le fait d'être entouré et d'être en bonne santé seraient donc des facteurs déterminants du niveau de bonheur.

Nous avons ensuite souhaité visualiser l'évolution de ces résultats dans le temps. Peut-on constater des tendances selon la période ? Cette partie de l'analyse n'a été réalisée qu'avec trois variables sélectionnées parmi celles ayant un fort taux de corrélation avec le niveau de bonheur : la sociabilité, le sentiment de sécurité et le niveau de santé.

Nous avons également créé un individu pour chaque année (2012, 2014 et 2016) en leur assignant pour valeur de chaque variable la moyenne des valeurs de tous les individus sur une année donnée.

Une croissance conjointe du niveau de bonheur et de sociabilité sur les 4 ans est observée. C'est la variable la plus concluante sur l'intégralité de la période, les variables relatives à la sécurité et au niveau de santé ayant un comportement différent de celui du niveau de bonheur soit sur l'intervalle 2012-2014 pour le premier et 2014-2016 pour le second.

ANALYSES COMPLETES :

<https://icimonbonheur.wordpress.com/2019/04/14/analyses-de-donnees-ess/>



<https://icimonbonheur.wordpress.com/2019/04/27/ess-evolution-du-bien-etre-des-francais-de-2012-a-2014/>

3.2 Analyse des données WHR

a. Introduction

Chaque année, le Global Happiness Council sort une enquête sur l'état du bonheur ou du bien-être subjectif regroupant tous les pays du monde. Le dernier en date est celui du 20 mars 2019. Ce projet, ayant vu le jour à partir de Février 2019 donc avant la nouvelle enquête de WHR, nous avons privilégié



le rapport des années 2006 à 2018. L'étude présente 156 pays selon des indicateurs de bien-être, nous avons conservé les données concernant la France et le Bénin.

Dans la partie de notre analyse, nous avons essayé de mettre en lien les 16 variables que contient ce rapport avec d'autres indicateurs issus Web (Twitter, Google Trends). Et nous avons focalisé notre attention sur trois années clefs : 2015 – 2016 – 2017.

Voir le fichier Excel WebIndicators_WHR2 dans le dossier :

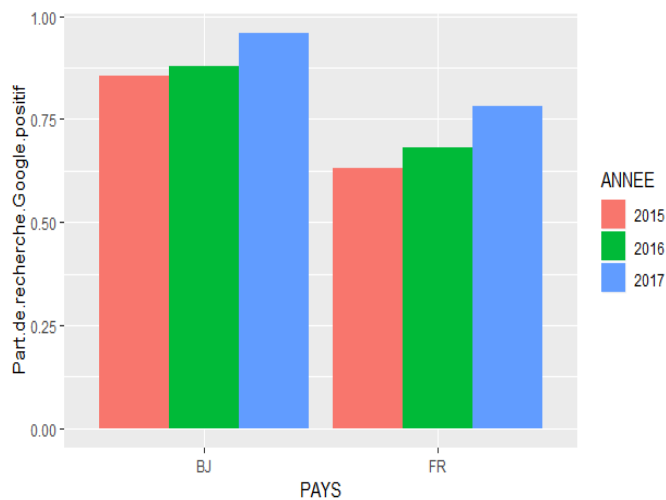
https://drive.google.com/drive/folders/1XDxs30kjT2DudzA8m0BPAQhm1V0IRR_N

b. Résultats

Avant de mettre en relation ces variables et pouvoir les comparer, il convient de se poser quelques questions : peut-on déduire, en analysant ces différentes variables que ces deux pays francophones se rejoignent sur le thème du bonheur ? Il y a-t-il une corrélation entre ces différentes variables ?



Après ces analyses, nous sommes parvenus aux résultats que les variables peuvent influencer l'un sur l'autre suivant les années. Et que le thème du bonheur varie en fonction du moment et de l'année des



enquêtes. Chaque moment avec son aperçu de la vie et du thème qui se présente à soi à ce moment précis. Dans ces résultats, il apparaît également que les indicateurs Web et les données du WHR peuvent aller dans le sens contraire. Quand on aperçoit une évolution de la positivité des thèmes sur les recherches Google, les enquêtes WHR, eux montrent le contraire donc un effet négatif.

c. Conclusion

On peut en déduire que ces deux pays, bien qu'étant francophones et quelques similitudes ne se rejoignent pas sur tous les thèmes de bonheur. Ils évoluent suivant chaque année et le bien-être que les individus ressentent à ce temps T.

ANALYSE COMPLETE :

<https://icimonbonheur.wordpress.com/2019/04/28/analyse-des-donnees-issues-de-whr/>



3.3 Analyse croisée des données ESS et WHR



A quel point est-ce que les modalités de l'enquête influencent les résultats obtenus ? Que se passe-t-il quand des questions différentes sont posées pour essayer de mettre un nombre sur une idée ? Pour répondre à ces questions nous avons comparé deux enquêtes menées en France (celles de ESS et WHR

présentées dans les deux parties précédentes) sur des variables relatives au bonheur et à la sociabilité, deux notions qui d'après nos premiers résultats semblent très liées.

Les résultats obtenus dans cette analyse croisée dénotent un comportement assez différent d'une variable à l'autre, trois variables sont relatives au bonheur mais chacune l'aborde d'un point de vue différent. Il est étonnant de constater qu'une évolution positive du niveau exprimé de bonheur s'accompagne d'une baisse de la perception de la qualité de sa vie.

ANALYSE COMPLETE :

<https://icimonbonheur.wordpress.com/2019/04/28/ess-et-whr-analyse-croisee-de-deux-enquetes-sur-le-bonheur-en-france/>



IV. CONCLUSIONS

Nous avons pu à travers ce projet prendre conscience des limites et avantages de chacune des



méthodes évoquées. Le défi était ambitieux et le travail que nous avons réalisé a sûrement dépassé l'attente originelle liée à ce projet. Cependant il semblait essentiel de pouvoir creuser les aspects qualitatifs de la question afin de pouvoir sous-peser les textualités offertes par le web avant de pouvoir s'imaginer construire une sorte d'indicateur totalement déconnecté de celles-ci.

Le web a l'avantage de fournir une description fine : géolocalisé, à une échelle temporelle journalière voir horaire. Directement disponible, cette matière est peu coûteuse à extraire, mais demande une vraie connaissance des techniques de traitement du langage, de la linguistique : les textualités peuvent être ambiguës, nous détachons parfois leurs contenus de leur contexte de création (ex : échanges de tweets entre amis) et les résultats fournis par Twitter sont très éparses (abréviations, fautes d'orthographe etc.).

Nous avons pu remarquer que l'indexation en première page de Google variait peu sur des requêtes francophones concernant nos deux pays. Celle-ci variant entre ressources lexicales (dictionnaire, définition, citations des mots recherchés), sites commerciaux (industrie, offre culturelle), et une offre grandissante de site de coaching personnel. Bien que nous ayons pu remarquer une corrélation les indices de bonheur de WHR, les indices maisons de bonheur GoogleTrends et Twitter, il faut aussi prendre conscience que les utilisateurs ne sont pas dupes, et se rendent compte que leurs requêtes ne trouvent pas directement leurs réponses à travers leurs moteurs de recherche favoris. Ceci peut avoir plusieurs effets notamment la simplification de la demande par les utilisateurs qui va ainsi faire gonfler artificiellement le nombre de recherches liées à un terme, mais aussi sur la spécification qui peut au contraire faire totalement disparaître certaines demandes.



Il serait pour le moment impensable de réaliser une enquête internationale du type de WHR via internet, notamment à cause de la situation numérique de chacun, des pluralités linguistiques de chaque pays, de la diversité de discours offerte par la toile, l'ambiguïté de ces discours. Surinterpréter ces fragments à travers une procédure de traitement automatique pourrait bien entendu mener à des dérives ce qui est déjà le cas, notamment en économie où ce genre de mesure de *mood* peuvent influencer des actes d'achat ou de vente. L'ouvrage de Yann Algan, préfacé par notre actuel président est en ce sens assez inquiétant, puisque celui-ci y décrit un Etat s'adaptant à des tendances d'opinions réalisés à travers ce type de démarche.

La mise en parallèle des deux jeux de données de ESS et WHR nous permet de faire au moins un constat, dans WHR comme dans ESS, la sociabilité sous les deux formes qu'on a abordée semble très lié au niveau de bonheur exprimé (*echelle_de_vie* pour WHR). A se demander si les enquêtés, au moment où on leur demande de situer leur niveau de bonheur sur une échelle de 0 à 10 ne représentent pas l'état de leur vie sociale pour répondre à cette question.

Cette mise en parallèle nous apporte aussi plusieurs pistes de réflexion : combien de questions doivent-être posées sur le bonheur pour pouvoir se rapprocher d'une appréciation « réelle » du niveau de bonheur des gens ? Ici, on observe clairement que deux angles d'approches peuvent donner des résultats assez différents. Demander aux individus d'évaluer leur niveau de bonheur n'est pas quelque chose d'anodin et les termes employés pour formuler la question ne peut qu'orienter la réponse. Quelles questions finalement permettraient de cerner ce qui constitue le bonheur d'un individu, et est-ce que ça ne serait pas différent pour chacun selon ses représentations propres ? Auquel cas est-ce pertinent de poser les mêmes questions à 66 millions d'habitants ?

Nous avons par ce travail exploré deux méthodes actuellement utilisées pour mesurer le bonheur. Nous avons mis en évidence les particularités de ces méthodes et les conséquences que ça implique sur les résultats obtenus. C'est un travail et une réflexion à poursuivre si l'on veut espérer s'approcher d'une mesure effective du bonheur.



BIBLIOGRAPHIE

- ❖ Alain (1925). Propos sur le bonheur
- ❖ Karma Ura Sabina Alkire Tshoki Zangmo Karma Wangdi (2012). A Short Guide to Gross National Happiness Index - The Centre for Bhutan Studies
- ❖ Christian Baudelot, Michel Gollacet (2001). Bonheur, travail et sociologie
- ❖ Jean-Pierre Algoud (2004). La société face à la demande de bonheur - Connexions N°81
- ❖ John F. Helliwell, Richard Layard and Jeffrey D. Sachs – WHR report 2018
- ❖ John F. Helliwell, Richard Layard and Jeffrey D. Sachs – WHR report 2019
- ❖ Alain Desrosières (2008). Gouverner par les nombres, L'argument statistique II
- ❖ Emmanuel Didier, Isabelle Bruno, Julien Prévieux (2014)- Statactivisme: Comment lutter avec des nombres
- ❖ Emmanuel Didier, Isabelle Bruno (2013). Benchmarking: l'État sous pression statistique
- ❖ Ruut VEENHOVEN (1997). PROGRÈS DANS LA COMPRÉHENSION DU BONHEUR
- ❖ Ruut Veenhoven (2004) . Subjective Measures of Well-being
- ❖ Benjamin Radcliff (2013). The Political Economy of Human Happiness How Voters' Choices Determine the Quality of Life
- ❖ Schwartz, H. A., J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, R. E. Lucas, M. Agrawal, G. J. Park, S. K. Lakshmikanth, S. Jha, M. E. Seligman, et al. (2013). Characterizing geographic variation in wellbeing using tweets. In ICWSM, pp. 583–591.
- ❖ Algan, Y., F. Murtin, E. Beasley, K. Higa, and C. Senik (2019). Wellbeing through the lens of the internet. PloS one 14 (1), e0209562.
- ❖ Bollen, J., H. Mao, and X. Zeng (2011). Twitter mood predicts the stock market. Journal of computational science 2 (1), 1–8.
- ❖ Algan, Y., T. Mayer, and M. Thoenig (2013). The economic incentives of cultural transmission: Spatial evidence from naming patterns across France.
- ❖ Algan, Y., F. Murtin, E. Beasley, K. Higa, and C. Senik (2019). Wellbeing through the lens of the internet. PloS one 14 (1), e0209562.
- ❖ Hills, T., E. Proto, and D. Sgroi (2017). Historical analysis of national subjective wellbeing using millions of digitized books. Working Paper



- ❖ Robert Kozinets (2009). *Netnography: Doing Ethnographic Research Online*, SAGE Publications Ltd,
- ❖ Agata Jackiewicz (2017). Outils notionnels pour l'analyse des controverses – Question de Communication N°31
- ❖ **Julien Boyadjian et Julien Velcin** (2017). L'analyse quantitative des médias sociaux, une alternative aux enquêtes déclaratives? La mesure de la popularité des personnalités politiques sur Twitter – Question de Communication N°31
- ❖ Jacobo Levy Abitbol, Márton Karsai, Jean-Philippe Magué, Jean-Pierre Chevrot, Éric Fleury
- ❖ (2018). Socioeconomic dependencies of linguistic patterns in Twitter: a multivariate analysis
- ❖ WWW '18 – World Wide Web Conference, Apr 2018, Lyon, France.