

UNIVERSITE TOULOUSE III – PAUL SABATIER
FACULTE DE CHIRURGIE DENTAIRE

ANNEE 2017

2017 TOU3 3011

THESE

POUR LE DIPLOME D'ETAT DE DOCTEUR EN CHIRURGIE DENTAIRE

Présentée et soutenue publiquement

par

Laure MIQUEL

Le 07 février 2017

**INTERETS DU « BIG DATA » POUR L'ESTIMATION DE LA
PREVALENCE CARIEUSE MONDIALE : APPORT DE GOOGLE®
TRENDS**

Directeur de thèse : Dr Jean-Noel VERGNES

JURY

Président :	Pr Cathy NABET
1er assesseur :	Dr Pierre-Pascal POULET
2ème assesseur :	Dr Jean-Noel VERGNES
3ème assesseur :	Dr Paul MONSARRAT



Faculté de Chirurgie Dentaire



➔ **DIRECTION**

DOYEN

Mr Philippe POMAR

ASSESEUR DU DOYEN

Mme Sabine JONJOT

CHARGÉS DE MISSION

Mr Karim NASR

Mme Emmanuelle NOIRRIT-ESCLASSAN

PRÉSIDENTE DU COMITÉ SCIENTIFIQUE

Mme Anne-Marie GRIMOUD

RESPONSABLE ADMINISTRATIF

Mme Marie-Christine MORICE

➔ **HONORARIAT**

DOYENS HONORAIRES

Mr Jean LAGARRIGUE +

Mr Jean-Philippe LODTER

Mr Gérard PALOUDIER

Mr Michel SIXOU

Mr Henri SOULET

➔ **ÉMÉRITAT**

Mr Damien DURAN

Mme Geneviève GRÉGOIRE

Mr Gérard PALOUDIER

➔ **PERSONNEL ENSEIGNANT**

56.01 PÉDODONTIE

Chef de la sous-section :

Mme BAILLEUL-FORESTIER

Professeurs d'Université :

Mme BAILLEUL-FORESTIER, Mr VAYSSE

Maîtres de Conférences :

Mme NOIRRIT-ESCLASSAN, Mme VALERA

Assistants :

Mme DARIES, Mr MARTY,

Adjoint d'Enseignement :

Mme BROUTIN

56.02 ORTHOPÉDIE DENTO-FACIALE

Chef de la sous-section :

Mr BARON

Maîtres de Conférences :

Mr BARON, Mme LODTER, Mme MARCHAL-SIXOU, Mr RÖTENBERG,

Assistants :

Mme GABAY-FARUCH, Mme YAN-VERGNES

Assistant Associé :

Mr BOYADZHIEV

56.03 PRÉVENTION, ÉPIDÉMIOLOGIE, ÉCONOMIE DE LA SANTÉ, ODONTOLOGIE LÉGALE

Chef de la sous-section :

Mr HAMEL

Professeurs d'Université :

Mme NABET, Mr SIXOU

Maîtres de Conférences :

Mr HAMEL, Mr VERGNES

Assistant :

Mlle BARON,

Adjoints d'Enseignement :

Mr. DURAND, Mr PARAYRE, Mr ROSENSWEIG

57.01 PARODONTOLOGIE***Chef de la sous-section :*** ***Mr BARTHET***

Maîtres de Conférences : Mr BARTHET, Mme DALICIEUX-LAURENCIN

Assistants : Mr. RIMBERT, Mr. ANDUZE-ACHER

Adjoints d'Enseignement : Mr. CALVO, Mr. LAFFORGUE, Mr SANCIER, Mr BARRE, Mme KADDECH

57.02 CHIRURGIE BUCCALE, PATHOLOGIE ET THÉRAPEUTIQUE, ANESTHÉSIOLOGIE ET RÉANIMATION***Chef de la sous-section :*** ***Mr COURTOIS***

Maîtres de Conférences : Mr CAMPAN, Mr COURTOIS, Mme COUSTY

Assistants : Mme CROS, Mme COSTA-MENDES

Adjoints d'Enseignement : Mr FAUXPOINT, Mr L'HOMME, Mme LABADIE

57.03 SCIENCES BIOLOGIQUES (BIOCHIMIE, IMMUNOLOGIE, HISTOLOGIE, EMBRYOLOGIE, GÉNÉTIQUE, ANATOMIE PATHOLOGIQUE, BACTÉRIOLOGIE, PHARMACOLOGIE)***Chef de la sous-section :*** ***Mr POULET***

Professeur d'Université : Mr KEMOUN

Maîtres de Conférences : Mme GRIMOUD, Mr POULET, Mr BLASCO-BAQUE

Assistants : Mr BARRAGUÉ, Mme DUBOSC, Mr LEMAITRE,

Assistant Associé : Mme FURIGA-CHUSSEAU

Adjoints d'Enseignement : Mr SIGNAT, Mr PUISSOCHET, Mr FRANC

58.01 ODONTOLOGIE CONSERVATRICE, ENDODONTIE***Chef de la sous-section :*** ***Mr DIEMER***

Professeur d'Université : Mr DIEMER

Maîtres de Conférences : Mr GUIGNES, Mme GURGEL-GEORGELIN, Mme MARET-COMTESSE

Assistants : Mr BONIN, Mr BUORO, Mme DUEYMES, Mme. RAPP, Mr. MOURLAN, Mme PECQUEUR

Adjoints d'Enseignement : Mr BALGUERIE, Mr ELBEZE, Mr MALLET

58.02 PROTHÈSES (PROTHÈSE CONJOINTE, PROTHÈSE ADJOINTE PARTIELLE, PROTHÈSE COMPLÈTE, PROTHÈSE MAXILLO-FACIALE)***Chef de la sous-section :*** ***Mr CHAMPION***

Professeurs d'Université : Mr ARMAND, Mr POMAR

Maîtres de Conférences : Mr BLANDIN, Mr CHAMPION, Mr ESCLASSAN, Mme VIGARIOS, Mr. DESTRUHAUT

Assistants : Mr. CHABRERON, Mr. EMONET-DENAND, Mr. KNAFO, Mme. SELVA, Mme. ROSCA

Adjoints d'Enseignement : Mr. BOGHANIM, Mr. FLORENTIN, Mr. FOLCH, Mr. GHRENASSIA,

Mme. LACOSTE-FERRE, Mr. POGEANT, Mr. RAYNALDY, Mr. GINESTE, Mr. CHAMPION,

Mr. LE GAC, Mr. GAYRARD

58.03 SCIENCES ANATOMIQUES ET PHYSIOLOGIQUES, OCCLUSODONTIQUES, BIOMATÉRIAUX, BIOPHYSIQUE, RADIOLOGIE***Chef de la sous-section :*** ***Mme JONJOT***

Maîtres de Conférences : Mme JONJOT, Mr NASR

Assistants : Mr. CANCEILL, Mme. GARNIER, Mr. OSTROWSKI

Adjoints d'Enseignement : Mr AHMED, Mme BAYLE-DELANNÉE, Mme MAGNE, Mr VERGÉ, Mme BOUSQUET, Mr MONSARRAT

À Maurice et Patrick, dans mon cœur tous les jours.

Remerciements

À **mes parents**, Marie-Line et Jean-Luc. Vous êtes mon socle, mon noyau dur, le phare dans la tempête. Je ne compte pas les fois où j'ai pu bénéficier de votre soutien sans faille. À toi **Maman**, qui a fait que je suis ici aujourd'hui, grâce à ta détermination et à ta pugnacité à tirer de moi le meilleur. Ton exigence est le plus beau des cadeaux. Je te suis éternellement reconnaissante. À toi **Papa**, ta joie de vivre et ton humour sont un moteur exceptionnel. Ton pragmatisme m'apaise et tu m'aides à garder les pieds sur terre. J'admire chaque jour un peu plus l'artiste que tu es à travers tes meubles magnifiques. Je vous aime éperdument.

À **Lionel**, mon frère. Tu supportes mes blagues lourdes et mon humour douteux dès que tu nous sommes ensemble chez Papa et Maman et je t'en suis reconnaissante. Tu es un jeune homme curieux et passionné. Je suis assurée de ta réussite future et tu pourras toujours trouver en moi une oreille attentive. Je t'aime.

À **tous les membres de ma famille** que je ne prends pas forcément le temps de voir aussi souvent que je le voudrais mais qui ont une place énorme dans mon cœur. Chacun à votre façon vous m'avez appris et m'apprendrez encore beaucoup. **Mamie Arlette, Papi Lucien, Mamie Josette, Papi Maurice, Bernard et Christine, Virginie, Ophélie, Séverine, Bernadette, Claudine et Patrick, Julien et Emilie, Martine, Ghislaine et Jean-Luc, Marie-José et Frédéric, Francis, Julien et Kévin, Martine et Michel, Franck et Karine, Anthony et Anaïs**, je vous estime et vous aime.

À **Camille**, la femme de ma vie, maintenant je peux le dire. Nous avons été (et sommes encore) un binôme de choc, le meilleur de toute évidence. Pour tous les souvenirs mémorables en clinique et ailleurs (le sol du Divino s'en souvient aussi), pour les joies et les peines que nous avons traversées ensemble, merci. Tu es le Mr Hyde de mon Dr Jekyll. Pour les fous rires, les appels quasi quotidiens et le sentiment de confiance absolue que j'aie en toi, merci. Tu es la meilleure amie que l'on puisse souhaiter.

À **Sally**, Yazbeckita pour les intimes (que pour moi en fait), Miss Lebanon pour les autres. Tu es un rayon de soleil. Ta douceur et tes conseils sont précieux pour moi. Merci de m'avoir fait découvrir la culture libanaise et de m'avoir permis de côtoyer les personnes formidables qui sont ta famille. Je ne veux faire ma visite guidée du Liban qu'avec toi. Tu as une place toute spéciale dans mon cœur.

À **Oihana et Lucile**, amies de la première heure qui ont été là dans les bons et les mauvais moments. Vous avez été géniales, je ne suis pas toujours facile à supporter et vous avez réussi. Oiha, encore merci pour ce squat de canapé pendant 1 mois. Lulu, encore merci ces weekends au ski. Vous êtes des filles exceptionnelles, douées de qualités humaines rares, je vous surkiffe !!!

Avec vous je n'oublie pas **Claire, Mathilde, Frédérique, Julien et Manu**. Je ne vous voie plus assez souvent à mon goût mais je ne doute pas qu'on s'organise des retrouvailles digne de ce nom régulièrement. Vous êtes des personnes formidables.

À **Kelly et Lousiane**, les belles gosses des Chalets. Que vous dire sinon que vous êtes fabuleuses, toujours disponibles et à l'écoute (d'ailleurs ces remerciements sont tapés depuis ton canapé Kéké). Vous êtes des nanas superbes, intelligentes et indépendantes. Je vous admire et vous adore, vous êtes des amies en or et vous pouvez compter sur moi.

À **Lucie**, tu es mon amie depuis la P1, que de souvenirs ensemble. Ton courage et ta ténacité forgent mon admiration. Je ne doute pas de ta réussite, tu le mérites.

Au **Docteur Chapart**, toujours disponible et à l'écoute même avec un patient sur le fauteuil. Je ne sais pas comment vous remercier de la patience dont vous avez fait preuve à mon égard. J'ai énormément appris à vos côtés, tant sur le plan technique que sur la manière d'aborder le patient. Travailler dans votre cabinet a été une expérience formidable qui laissera sa trace dans ma manière d'aborder ce métier. L'écoute dont vous faites preuve est une qualité rare. Vous avez toute mon estime.

À **Sylvie**, qui a brillamment su endosser les casquettes d'assistante dentaire/psy durant mon séjour au cabinet. Ton oreille bienveillante et patiente a été d'une grande aide. Ta capacité à tout gérer de front me fascine, tu es une femme admirable et je peux t'assurer de ma reconnaissance (et moi j'adore ta voix !).

Au **Docteur Costes**, qui a elle aussi fait preuve d'une patience à toute épreuve et qui a toujours su me conseiller sans perdre de vue l'intérêt des patients. Votre sérénité est un modèle pour moi. J'ai grandement apprécié de travailler à vos côtés et vous assure de mon plus profond respect.

À **Marie** et **Vanessa**, deux assistantes de choc.

Aux **Docteurs Lassara, Benezet** et à **Céline. Damien**, merci d'avoir eu la confiance de me laisser ton magnifique cabinet pour mon premier remplacement, j'ai commencé dans les meilleures conditions possibles. Ton exigence et ta rigueur restent le leitmotiv de ma pratique. **Céline**, tu es une assistante de rêve, merci pour ta gentillesse à mon égard. **Stéphanie**, merci pour ton soutien et ton écoute. Je vous souhaite tout le bonheur possible.

À **Isabelle**, tu as été là durant une période compliquée. Je t'en suis infiniment reconnaissant. Tu es quelqu'un de formidable. Je ne te permets pas d'en douter. Merci d'être toi.

À toutes ces nouvelles personnes formidables rencontrées il y a 2 ans maintenant et que j'apprécie chaque jour un peu plus de côtoyer : **Maylis, Cécile, Dominique, David, Hugues et Olivia, Karim et Mbarka, Nathalie et Olivier**. Je garde de beaux souvenirs de nos soirées.

À tous les copains d'ici et d'ailleurs rencontrés au cours des voyages notamment toi **Diana**, grâce à qui j'ai pu découvrir Florence et à **Javier** et tes soirées pleines de personnes venues des quatre coins du monde qui permettaient de s'enrichir humainement. Aux potes de médecines, vous ferez de supers praticiens je n'en doute pas. À tous ceux que j'oublie de citer ici. Je pense à vous.

À **l'Atelier de cartographie de l'Université Toulouse Jean Jaurès** et plus particulièrement à **François Aussaguel** qui permis l'élaboration de la carte Google® Trends. Soyez vivement remercié de votre aide, de votre écoute et de votre disponibilité.

Je clôture ces remerciements sur ces mots :

« Faut savoir s'étendre/ Sans se répandre »
Serge Gainsbourg, Pauvre Lola

Au président du jury,

Madame le professeur Catherine NABET

- Professeur des Universités, Praticien hospitalier d’Odontologie,
- Docteur en Chirurgie Dentaire,
- Diplôme d’Etudes Approfondies de Santé Publique – Epidémiologie,
- Docteur de l’Université Paris XI,
- Habilitation à Diriger des Recherches (HDR),
- Lauréate de la Faculté de Médecine,
- Lauréate de l’Université Paul Sabatier,
- Lauréate de l’Académie Nationale de Chirurgie Dentaire.

Merci de l’honneur que vous nous faites d’accepter la présidence de cette thèse.

Votre enseignement de la santé publique et de l’épidémiologie nous a ouvert à ce domaine passionnant qui viendra enrichir notre pratique.

Nous vous remercions de votre disponibilité tout au long de ce travail.

Veuillez croire en l’expression de notre respect le plus grand.

Aux membres du jury,

Monsieur le docteur Pierre-Pascal POULET

- Maître de Conférences des Universités, Praticien Hospitalier d'Odontologie,
- Responsable de la Sous-Section Sciences Biologiques,
- Docteur en Chirurgie Dentaire,
- Docteur de l'Université Paul Sabatier,
- Lauréat de l'Université Paul Sabatier.

Nous vous remercions vivement d'avoir accepté de siéger à ce jury.

La passion que vous mettez dans vos enseignements, la volonté de faire de nous des praticiens accomplis nous ont permis d'appréhender toute la complexité de notre exercice et de faire preuve de l'exigence nécessaire.

Que soient exprimés ici notre respect et notre reconnaissance.

Monsieur le docteur Paul MONSARRAT

- Adjoint d’Enseignement à la Faculté de Chirurgie Dentaire de Toulouse
- Ex. Assistant hospitalo-universitaire d’Odontologie,
- Docteur en Chirurgie Dentaire.
- Master 1 Recherche : Biosanté
- Master 1 Recherche : Méthodes d’Analyse et de Gestion en Santé Publique,
- Master 2 Recherche : mention : Biologie, santé ; spécialité : Physiopathologie,
- Lauréat de la faculté de Médecine Rangueil et de Chirurgie Dentaire de l’Université Paul Sabatier,
- Diplôme Universitaire d’Imagerie 3D maxillo-faciale.

Nous vous remercions très chaleureusement de siéger à ce jury.

Votre investissement et votre rigueur sont un modèle pour tout futur praticien.

Nous avons apprécié de pouvoir apprendre à vos côtés notre futur métier.

Votre regard et votre analyse sur ce travail sont d’une grande importance pour nous.

Que soient exprimés ici notre reconnaissance et notre gratitude.

À notre directeur de thèse,

Monsieur le docteur Jean-Noël Vergnes

- Maître de Conférences des Universités, Praticien Hospitalier d’Odontologie,
- Docteur en Epidémiologie,
- Docteur en Chirurgie Dentaire,
- Professeur associé, Oral Health and Society Division, Université McGill – Montréal, Québec – Canada,
- Maîtrise de Sciences Biologiques et Médicales,
- Master 2 Recherche – Epidémiologie clinique,
- Diplôme d’Université de Recherche Clinique Odontologique,
- Lauréat de l’Université Paul Sabatier

Nous vous témoignons les plus sincères remerciements d’avoir accepté de diriger cette thèse.

Votre finesse d’analyse et votre acuité nous ont permis de mener à bien ce travail.

Nous avons toujours pu trouver un soutien et une écoute attentive.

Nous avons apprécié votre disponibilité.

Soyez assuré de notre profond respect.

Table des matières

INTRODUCTION	15
1. GENERALITES SUR LA SANTE PUBLIQUE BUCCO-DENTAIRE ET LA MALADIE CARIEUSE	17
1.1. LES POLITIQUES DE SANTE PUBLIQUE POUR LA SANTE BUCCO-DENTAIRE EN FRANCE	17
1.1.1. <i>Le ministère de la Santé et l'action publique</i>	17
1.1.1.1 Le plan de prévention bucco-dentaire.....	17
1.1.1.2 La convention entre l'Assurance maladie et le chirurgien-dentiste	18
1.1.1.3 La Haute Autorité de santé.....	18
1.1.1.4 Les actions locales	18
1.1.2. <i>Un acteur privé : l'Union Française pour la Santé Bucco-dentaire (UFSBD)</i>	19
1.2. LES POLITIQUES DE SANTE PUBLIQUE POUR LA SANTE BUCCO-DENTAIRE AU NIVEAU MONDIAL	20
1.2.1. <i>L'Organisation mondiale de la santé</i>	20
1.2.2. <i>La Fédération Dentaire Internationale</i>	22
1.3. LA CARIE.....	23
1.3.1. <i>Définition</i>	23
1.3.2. <i>Les déterminants de la carie</i>	23
1.3.3. <i>Conséquences</i>	25
1.3.4. <i>Traitement</i>	26
1.3.4.1 Préventif	26
1.3.4.2 Curatif.....	27
1.3.5. <i>Epidémiologie</i>	27
1.3.5.1 En France	27
1.3.5.2 Au niveau mondial.....	28
1.4. SURVEILLANCE ET SUIVI.....	29
1.4.1. <i>En France</i>	29
1.4.2. <i>Au niveau mondial</i>	29
2. GOOGLE[®] ET SON OUTIL GOOGLE[®] TRENDS	33
2.1. PRESENTATION DE GOOGLE [®]	33
2.2. GOOGLE [®] EN QUELQUES CHIFFRES	33
2.3. LES MULTIPLES OUTILS DE GOOGLE [®]	34
2.4. GOOGLE [®] TRENDS.....	35
2.4.1. <i>Origine des données</i>	35
2.4.2. <i>Ajustement des données</i>	36
2.4.3. <i>Export des données</i>	36
2.5. UTILISATION DE GOOGLE [®] TRENDS	36
2.5.1. <i>La page d'accueil</i>	36
2.5.2. <i>La page de résultat</i>	39
2.5.2.1 Evolution de l'intérêt dans le temps.....	39
2.5.2.2 Répartition dans l'espace géographique	40
2.5.2.3 Autres termes associées lors de la recherche.....	42
2.5.3. <i>Astuces de recherche</i>	43
2.5.3.1 Absence de résultats	43
2.5.3.2 Précision de la recherche.....	43

2.5.4. <i>Filtres disponibles</i>	45
2.5.4.1 Filtrage par période et par pays.....	46
2.5.4.2 Filtrage par catégorie.....	46
2.5.4.3 Filtrage par type de recherche	46
2.5.5. <i>Comparaison des termes</i>	47
2.5.5.1 Comparaison des termes	47
2.5.5.2 Comparaison dans l'espace géographique	50
2.5.5.3 Comparaison dans le temps	52
2.5.5.4 Comparaison de termes en langues différentes	54
2.5.5.5 Comparaison sujet / terme.....	55
3. BIG DATA ET SANTE	57
3.1. DEFINITION DU BIG DATA	57
3.2. LE MONDE MEDICAL ET LE BIG DATA	58
3.2.1. <i>Un intérêt accru pour le Big Data</i>	58
3.2.2. <i>Applications en santé publique : des potentialités non négligeables</i>	59
3.2.2.1 Des outils plus performants.....	59
3.2.2.2 Des connaissances accrues sur les comportements des populations.....	59
3.2.2.3 Une aide pour les politiques de santé publique	61
3.2.2.4 Prédire grâce au Big Data	62
3.2.3. <i>Intérêts de Google® Trends en épidémiologie bucco-dentaire</i>	64
3.2.3.1 Utiliser Google® Trends à bon escient : suivre une check-list	64
3.3. LES LIMITES DU BIG DATA	67
3.3.1. <i>Pour qui, pour en faire quoi ?</i>	67
3.3.2. <i>Trop d'informations tuent l'information</i>	67
3.3.2.1 Google® Flu Trends : un échec de l'utilisation du Big Data devant l'abondance d'informations	69
3.3.3. <i>Limites propre à Google® Trends</i>	70
3.4. ETHIQUE ET BIG DATA.....	73
4. ESTIMATION DE LA PREVALENCE CARIEUSE MONDIALE A L'AIDE DE GOOGLE TRENDS®	74
4.1. INTRODUCTION	74
4.2. MATERIEL ET METHODE	75
4.2.1. <i>La carte de l'OMS</i>	75
4.2.1.1 L'indice CAOD	75
4.2.2. <i>Méthode</i>	78
4.2.2.1 Premier préalable : traduction du terme « carie »	78
4.2.2.2 Deuxième préalable : choix des options de recherche : période et catégorie.....	81
4.2.2.3 Troisième préalable : choix de la combinaison des termes et des marques de citations	81
4.2.2.4 Première étape : recherche du terme dans Google® Trends	82
4.2.2.5 Deuxième étape : création d'un« pool de référence »	85
4.2.2.6 Troisième étape : identifier les « niches linguistiques » ou la guerre aux recherches croisées.....	87
4.2.2.7 Quatrième étape : obtention des données permettant de réaliser la carte.....	94
4.2.2.8 La check-list de Nuti et al. : rendre notre méthode transparente et reproductible	98
4.2.2.9 Diagramme de flux de l'étude	102
4.3. RESULTATS.....	103
4.3.1. <i>Méthode de réalisation de la carte</i>	103
4.3.2. <i>Cartes</i>	105
4.3.2.1 Carte Google® Trends.....	105
4.3.2.2 Correspondance entre la carte des requêtes Google Trends® et les cartes de l'OMS	106

4.3.2.3 Comparaison statistique des cartes.....	107
4.3.2.4 Résultats et concordance des cartes	108
4.4. DISCUSSION	109
CONCLUSION.....	111
TABLE DES ILLUSTRATIONS.....	112
TABLE DES TABLEAUX.....	113
BIBLIOGRAPHIE	114
ANNEXE 1 : EXEMPLE DE CATEGORIES ET SOUS-CATEGORIES DISPONIBLES DANS GOOGLE[®] TRENDS (77)	120
ANNEXE 2 : TRADUCTION DU TERME « CARIE » DANS LES LANGUES DISPONIBLES DE GOOGLE[®] TRADUCTION	122
ANNEXE 3 : LISTE FINALE DES TERMES RETENUS POUR LA CONSTITUTION DE L'ECHANTILLON ET RAISONS DE REJETS DES AUTRES TERMES	123
ANNEXE 4 : TABLEAU COMPLET DES PAYS ET DE LEUR POPULARITE EN FONCTION DES REQUETES	124

Introduction

Une des pathologies les plus courantes en santé bucco-dentaire est la maladie carieuse. Malgré un net recul de son incidence et de sa prévalence dans les pays industrialisés, l'Organisation Mondiale de la Santé observe toujours une forte fréquence des complications médicales liées à la carie (1). Des facteurs sociaux, environnementaux et liés au mode de vie vont jouer une grande influence sur la santé bucco-dentaire (2). Les populations en situation de précarité restent largement touchées (2).

Les politiques de santé publique ont un rôle prépondérant pour essayer de faire changer les habitudes et les comportements. En amont de toute action de santé publique, connaître l'épidémiologie des maladies constitue la première étape pour la constitution de programmes de prévention et de traitement efficaces. Ceci est vrai pour la carie dentaire, pour laquelle l'identification de populations à risque constitue un enjeu important (1).

Comme toute profession médicale, le chirurgien-dentiste a depuis de longues années pu jouer un rôle de prévention et de promotion de la santé bucco-dentaire. C'est le cas en France, avec des structures publiques ou associatives telle que l'Union Française pour la Santé Bucco-Dentaire (UFSBD) dès la fin des années 1960(3), via ses interventions dans les écoles ainsi que sa participation à la réalisation d'études épidémiologiques (3). C'est aussi le cas au niveau mondial avec le programme pour la santé bucco-dentaire de l'Organisation Mondiale de la Santé (OMS) dont le plus récent date de 2003 et qui a donné lieu en 2007 à une assemblée afin de mettre en place un plan d'action précis (1).

L'utilisation de plus en plus courante d'internet pour la recherche d'informations médicales (4), notamment via les moteurs de recherche tels que Yahoo® ou Google®, constitue un précieux matériel sur la manière dont les populations s'interrogent sur leur santé et sur les pathologies dont ils pensent souffrir.

Des outils comme Google Trends® permettent d'analyser où, quand et comment les recherches sont réalisées par les internautes. Cette énorme quantité d'informations recueillie s'appelle du Big Data ou données massives.

Cette thèse a pour but d'évaluer comment un moyen moderne de recueil de données, peut apporter sa contribution au domaine de la santé publique en Odontologie. En effet, le

recours à l'analyse des données massives pourraient nous aider à appréhender différemment le problème de santé publique que représente la carie dentaire.

L'hypothèse générale du travail présenté dans cette thèse est que le recours aux données massives pourrait permettre de compléter les connaissances issues des enquêtes de terrain, et d'en optimiser les coûts, ce qui pourrait par exemple permettre de privilégier l'utilisation des ressources disponibles sur des actions de prévention ou des mesures thérapeutiques ciblées. Il est donc important d'évaluer dans quelle mesure le recours aux données massives permet d'estimer les besoins d'une population en matière de santé.

1. Généralités sur la santé publique bucco-dentaire et la maladie carieuse

1.1. Les politiques de santé publique pour la santé bucco-dentaire en France

1.1.1. Le ministère de la Santé et l'action publique

Au niveau français, la politique de prévention actuelle pour la santé bucco-dentaire est représentée par plusieurs initiatives:

- un plan de prévention (2),
- un objectif de la loi de Santé publique (2),
- une convention entre l'Assurance maladie et les chirurgiens-dentistes comprenant un dispositif de prévention (2).

On peut aussi observer des actions à l'échelle locale.

1.1.1.1 Le plan de prévention bucco-dentaire

Le dernier plan de prévention a été lancé en novembre 2005 pour la période 2006-2010. Il avait pour volonté principale de réduire de 30% en cinq ans l'indice carieux de la carie chez les enfants de 6 à 12 ans (5) comme énoncé dans l'objectif 91 de la loi de Santé publique de 2004 (2).

Le plan de prévention bucco-dentaire s'organise en 10 mesures avec comme mesure principale un examen de prévention obligatoire pour les enfants de 6 et 12 ans.

Cet examen se fait avec une dispense d'avance de frais et une prise en charge de 100% par l'Assurance maladie de ce dernier et des soins consécutifs (2).

Cet objectif, pour être atteint, profite de la mise en place d'une campagne de communication radiophonique et télévisuelle avec le slogan « M'T Dents », pour promouvoir l'examen de prévention bucco-dentaire dont bénéficient les enfants (5).

Les autres mesures promues dans ce plan concernent une meilleure prise en charge de la santé bucco-dentaire des personnes âgées dépendantes et des personnes handicapées ; ainsi que la prévention accrue des cancers buccaux (5).

D'autres actions sont proposées à titre expérimental avec entre autres des consultations à 4 mois de grossesse et 6 mois après la naissance ou des consultations de prévention à 60 ans. Ces mesures sont testées et évaluées dans quelques départements (2).

1.1.1.2 La convention entre l'Assurance maladie et le chirurgien-dentiste

La convention nationale entre les chirurgiens-dentistes et l'Assurance maladie modifie le dispositif de prévention antérieur en proposant un examen de prévention tous les 3 ans entre 6 et 18 ans, sans avance de frais (2).

1.1.1.3 La Haute Autorité de santé

Cette autorité publique est indépendante et propose une évaluation des politiques de santé publique via un rapport (2). Grâce à ce document, en plus de l'épidémiologie de la maladie carieuse, on a une vue d'ensemble des stratégies de prévention de la carie disponible en 2010 en France ainsi que leur évaluation (2). La Haute Autorité de Santé propose enfin des recommandations pour adapter les politiques de santé publique en fonction des dernières recommandations et études fiables publiées tant au niveau national qu'international (2).

1.1.1.4 Les actions locales

Les collectivités locales peuvent, à leur échelle, promouvoir des actions de prévention. Il en est de même pour les caisses régionales de l'Assurance maladie (2).

Ces actions ont été appuyées par des recommandations en santé publique de la Haute Autorité de Santé sur les stratégies de prévention de la carie dentaire publiées en mars 2010 (2).

1.1.2. Un acteur privé : l'Union Française pour la Santé Bucco-dentaire (UFSBD)



Figure 1 : logo de l'UFSBD (3).

L'Union Française pour la Santé Bucco-Dentaire est une association fondée en 1966. Depuis sa création elle est fortement impliquée dans la politique de prévention de la carie dentaire chez les enfants et propose des ateliers dans les écoles pour enseigner l'hygiène bucco-dentaire (6).

En 1986, elle promeut l'introduction du sel fluoré en France pour la prévention de la carie (6).

Son implication dans la santé publique au niveau national est confirmée par la réalisation de 4 enquêtes épidémiologiques en 1987, 1990, 1993 et 1998 à la demande du ministère de la santé (6).

De même, elle collabore avec l'Organisation Mondiale de la Santé depuis 1998 démontrant ainsi sa reconnaissance au niveau mondial (6).

Enfin, étant le promoteur du programme de prévention « M'T Dents » pour la prévention bucco-dentaire des 6-12 ans depuis 2006 (6) ; et la publication d'une charte de recommandations pour la santé bucco-dentaire des personnes handicapées et des personnes âgées dépendantes en 2011 (6), elle confirme encore son implication dans la politique de prévention menée par le ministère de la santé.

1.2. Les politiques de santé publique pour la santé bucco-dentaire au niveau mondial

1.2.1. L'Organisation mondiale de la santé



Figure 2 : logo de l'Organisation Mondiale de la Santé (7).

L'Organisation Mondiale de la Santé (OMS) est créée en 1948 (7). Elle a pour but de diriger et coordonner la gestion de la santé publique des pays membres des Nations Unies (8). Elle a donc un rôle clé en donnant des normes et des critères sur les questions majeures de santé (9). La santé bucco-dentaire en fait partie.

Depuis 1983, la santé bucco-dentaire est inscrite dans la stratégie de « santé pour tous » de l'OMS (10).

Elle a inscrit en 1989 la promotion de la santé bucco-dentaire comme partie intégrante de son programme de santé pour tous en vue de l'an 2000 (10).

En 1994, elle consacre sa Journée mondiale de la santé à l'information et à la promotion de la santé bucco-dentaire (10).

Elle a édité en 2003 un rapport sur la santé bucco-dentaire avec la volonté de continuer l'amélioration de cette dernière pour le XXI^{ème} siècle (10). Il y est écrit que les maladies bucco-dentaires sont des problèmes majeurs de santé publique par leur forte incidence et prévalence (11).

L'OMS explique que la santé bucco-dentaire ne tient pas seulement au fait d'avoir des dents saines. Cela exclut aussi de souffrir de douleurs oro-faciales, de cancers de la cavité buccale, de lésions des tissus de la cavité buccale, d'anomalies congénitales (fentes labiales, labio-palatines ou palatines) (10).

Les pathologies bucco-dentaires partagent de nombreux facteurs de risques avec les quatre principales maladies chroniques non transmissibles que sont les maladies cardio-vasculaires,

le diabète, le cancer et les pneumopathies obstructives chroniques. Le fait de lutter efficacement contre ces pathologies a donc une incidence directe sur l'état de santé général et la qualité de vie des populations (10).

Il est aussi fréquent que certaines pathologies générales aient une répercussion sur la sphère orale. Un bon suivi bucco-dentaire fait partie de la prise en charge globale du patient. Devant l'intrication des facteurs de risques entre les maladies générales chroniques et les pathologies bucco-dentaires il est important d'agir sur la santé bucco-dentaire afin de traiter le patient dans sa globalité (10).

Les deux principales pathologies bucco-dentaires sont la carie et la maladie parodontale.

Les populations en situation de pauvreté sont les plus touchées. Les traitements disponibles restent très peu accessibles pour ces individus (10).

En 2007, c'est un article reprenant les décisions des instances dirigeantes de l'OMS, l'Assemblée de la Santé et le Conseil exécutif, qui vient confirmer les politiques que l'OMS souhaite mettre en place pour améliorer la santé bucco-dentaire (1).

Il est décidé d'intégrer la santé bucco-dentaire dans les programmes concernant les maladies chroniques non transmissibles (1). Les principales mesures décidées concernent la promotion de la santé bucco-dentaire avec des programmes nationaux, la mise en place de programme de fluoration dans les pays non dotés et la prévention des cancers buccaux (1).

Son message essentiel est que la santé bucco-dentaire ne doit pas être considérée à part mais comme étant un élément déterminant de la santé générale et de la qualité de vie (1).

1.2.2. La Fédération Dentaire Internationale



Figure 3 : logo de la Fédération Dentaire Internationale (11).

La Fédération Dentaire Internationale a fêté ses 100 ans en 2012. Il s'agit d'une fédération dont le but est d'organiser des congrès au sujet de la santé bucco-dentaire, de l'éducation et de l'hygiène bucco-dentaire.

Après son congrès à Mexico en automne 2011, elle a publié un rapport sur sa vision de la santé bucco-dentaire en 2020. Elle affirme qu'un bon état de santé général nécessite une bonne santé bucco-dentaire (11).

Elle note qu'au cours du temps, on a plutôt privilégié le traitement à la prévention des maladies bucco-dentaires, ce qui entraîne des coûts prohibitifs, rendant l'accès aux soins difficiles dans certaines populations défavorisées et dans les pays sous-développés (11).

Ainsi, comme l'OMS, la FDI milite pour une intégration de la santé bucco-dentaire dans les programmes de santé publique des états et non comme un parent pauvre des politiques de prévention. Elle veut passer d'un modèle traditionnel curatif qui tendrait vers un modèle plus « salutogénique », axé sur la prévention et la promotion de la santé bucco-dentaire (11).

1.3. La carie

Au niveau mondial, la carie est la principale pathologie dentaire avec les maladies parodontales (10).

La carie et ses complications peuvent toucher l'ensemble des individus dentés mais avec une fréquence différente selon le type de population et l'âge (2).

1.3.1. Définition

D'après le Dictionnaire médical de l'Académie de Médecine la carie dentaire résulte de la déminéralisation d'une dent (12).

Cette déminéralisation est entraînée par des facteurs endogènes, propre à la physiologie de l'individu que peuvent être la plaque dentaire bactérienne présente en bouche, les maladies gingivales, et des facteurs génétiques qui peuvent atteindre l'émail (hypoplasie ou agénésie) (2).

On a aussi une part de facteurs exogènes dans la maladie carieuse tel que le mode alimentaire plus ou moins riche en sucre et la consommation de tabac ou d'alcool (12).

Ces facteurs exogènes sont fortement reliés à l'environnement social dans lequel va évoluer l'individu.

La carie dentaire procède par des déminéralisations dues à la présence d'acide qui entraînent la perte de substance des différents tissus composant la dent : dentine, émail et pulpe. A terme, cela nécessite des soins plus ou moins importants (2, 12).

1.3.2. Les déterminants de la carie

Les déterminants prédisposant à la carie sont multiples. Ils sont à la fois liés aux habitudes d'hygiène mais sont aussi constitutionnels, sociétaux voire liés à des traitements médicaux ou à un environnement professionnel (2).

Les **déterminants comportementaux** regroupent la qualité de l'hygiène bucco-dentaire dont la personne est capable et ses habitudes alimentaires (2).

Une alimentation très sucrée, couplée à une prise alimentaire répétée (ou grignotage) jouent un rôle déterminant dans l'apparition de la carie. De plus, si le contrôle de plaque n'est pas

assez soutenue, on observe l'installation d'un biofilm (ou plaque dentaire) contenant des bactéries cariogènes (2).

Ils englobent aussi la capacité à recourir à des actes de prévention voire des soins précoces si cela est nécessaire. Il est à noter que plus le délai de recours aux soins sera tardif, plus on diminuera les chances de conserver la dent sur l'arcade (2).

On comprend aisément que ces deux volets, contrôle de plaque et habitudes de vie peuvent largement bénéficier de politique de prévention.

Les **déterminants constitutionnels** sont liés à l'hérédité car ils concernent la résistance de l'émail, la qualité et quantité de salive disponibles ainsi que les caractéristiques de la plaque dentaire. Ils sont propres à chaque personne (2).

Les **déterminants sociétaux** consistent en un modèle holistique présenté par Fisher-Owens et al. permettant de donner trois dimensions à l'individu : l'individu, son environnement familial et sa place dans la société.

Ainsi, la santé orale du sujet va dépendre de tous ces facteurs exogènes et pas seulement ceux propres à l'individu. Ce modèle a été développé pour l'odontologie pédiatrique mais peut aussi s'appliquer aux adultes (13).

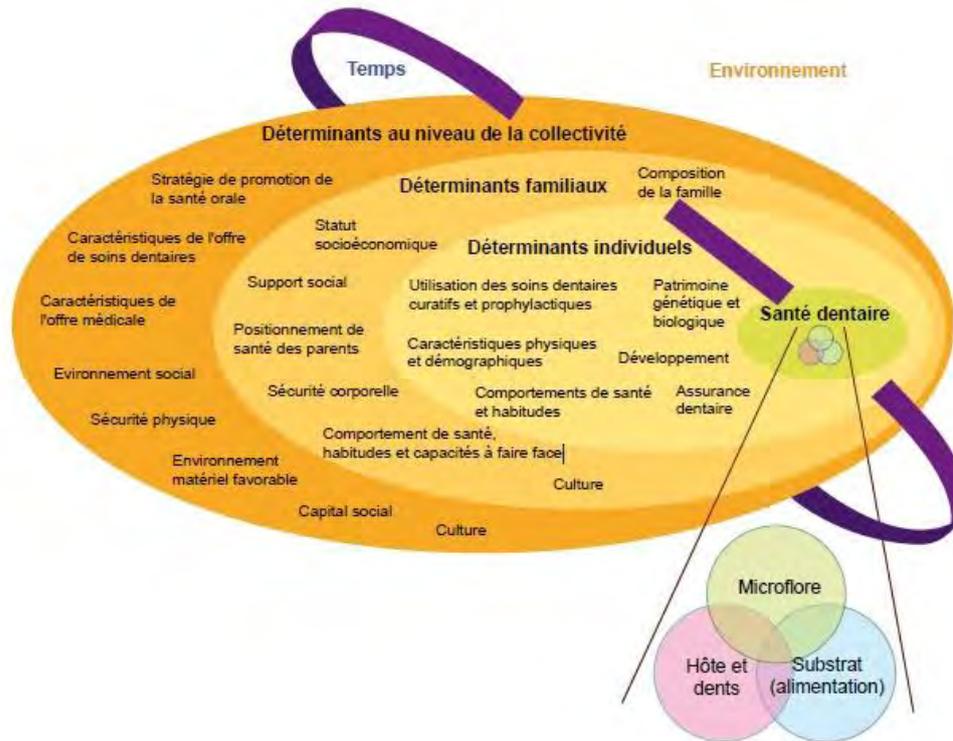


Figure 4 : modèle conceptuel de santé orale des enfants proposé par Fisher–Ovvens et al. (13).

Les **déterminants liés à des traitements médicaux** tels que ceux soignant des pathologies à l'aide de médicaments avec du saccharose, qui fait partie des sucres fermentescibles, entraînent divers problèmes bucco-dentaires; d'autres favorisent la xérostomie, comme par exemple la radiothérapie cervico-faciale pour le traitement du cancer (2).

Les **déterminants liés à certains environnements professionnels** se réfèrent à tous les métiers où le travailleur est exposé aux sucres. Cela a pour effet d'entraîner des caries multiples sur les faces vestibulaires des incisives et des canines alors que cela est plus rare en population générale (2).

1.3.3. Conséquences

Une fois que la carie a commencé la destruction du tissu dentaire de manière irréversible, selon son stade, on va assister à des complications. La plus courante est la douleur mais on peut assister à la perte de la dent si la lésion est trop importante voire à des problèmes infectieux locaux ou locorégionaux qui peuvent eux-mêmes se compliquer en infections à distance (2).

La perte de la dent peut générer des troubles fonctionnels en gênant la nutrition et la phonation ainsi que des problèmes esthétiques. À terme, cela peut entraîner des conséquences sur le plan social (2).

Il est fort probable que les requêtes des internautes au sujet de la carie dentaire proviennent essentiellement de la douleur dentaire et des barrières ou conséquences sociales du traitement (ou du non traitement).

1.3.4. Traitement

1.3.4.1 Préventif

Tout un volet de moyens de prévention existe avant que la carie atteigne de manière irréversible la dentine. La prévention va agir en faveur du processus de reminéralisation de l'émail.

Ces moyens sont :

- Le fluor via une action topique du fluor (gel, vernis fluoré, dentifrice) plus qu'une action au niveau systémique (2).
- Le brossage des dents et autres mesures d'hygiène bucco-dentaire dans le but d'éliminer la plaque dentaire (2).
- Le scellement des sillons dentaires à l'aide d'un matériau adhésif fluide (2).
- Les conseils hygiéno-diététiques en insistant ce sur ce qu'on appelle le « grignotage », qui est à éviter au maximum (2).
- Les consultations régulières chez le chirurgien-dentiste qui vont permettre de dépister les potentielles caries et de mettre en place les actions de prophylaxie décrites ci-dessus(2).

Dans tous les cas, les moyens de prévention sont à la fois d'ordre individuel tel que les habitudes alimentaires, l'hygiène bucco-dentaire ou le recours aux soins mais aussi d'ordre collectif via la supplémentation en fluor via le sel de table (2).

Devant la diversité des déterminants, on imagine aisément que les stratégies de prévention vont passer par de multiples canaux et des acteurs variés. Cette pluralité exige de coordonner les actions au sein d'une politique de santé publique afin que la confusion ne s'installe pas(2).

1.3.4.2 Curatif

Selon la perte de substance due à la lésion carieuse le traitement de la carie varie.

Si la dent reste vivante, on peut réaliser une obturation étanche, de manière directe (au fauteuil) à l'aide de matériaux comme des résines appelées « composites » ou indirecte (avec une étape de laboratoire) grâce aux inlays-onlays (2).

Si le traitement des racines doit être effectué, il est fréquent que la dent nécessite une prothèse dentaire (2).

Enfin, si la dent doit être avulsée, des thérapeutiques prothétiques doivent être mise en place pour compenser l'édentement (2).

Ainsi, dans tous les cas les soins conservateurs et prothétiques nécessaires entraînent un coût pour le patient et pour la collectivité.

1.3.5. Epidémiologie

1.3.5.1 En France

Pour sa dernière politique d'envergure pour la santé bucco-dentaire via un plan national de prévention, le ministère de la Santé donne les chiffres suivants pour motiver son engagement : plus d'un tiers des enfants de 6 ans sont atteints par la carie dentaire, 45 % à 12 ans et plus de 75% des adultes (14).

On note la très forte prévalence de la carie dans la population française, ce qui en fait un problème majeur de santé publique.

Il est cependant difficile d'obtenir des chiffres précis selon les diverses tranches d'âge. Les dernières données fiables publiées en France sont celles tirées des recommandations de la Haute Autorité de santé concernant la lutte contre la carie dentaire et concerne la période entre les années 2000 et 2009 (2).

Chez les enfants de moins de 6 ans, il existe peu d'études mais on estime à 20 à 30% des enfants de 4 à 5 ans atteints d'au moins 1 carie ainsi qu'une forte corrélation entre le lieu de vie ou le niveau socio-économique. En effet, les enfants en ZEP ou en milieu précaire semblent largement plus touchés (2).

Chez les enfants de 6 à 12 ans, de nombreuses études sont disponibles. A l'âge de 6 ans, les enfants exempts de carie sont de 50 à 70% selon les études et pour les enfants de 12 ans, le pourcentage est de 45-55% selon les études. La disparité entre les populations à faible niveau socio-économique ou vivant en ZEP et la population d'un niveau socio-économique plus élevé persiste avec une majorité des soins concentrée sur une faible partie de la population (2).

Pour la population adolescente, les études sont peu nombreuses et concernent surtout les moins de 15 ans. Une grande disparité dans les chiffres existe car on oscille entre 20 et 50% des adolescents indemnes de caries. On retrouve toujours le lien entre précarité et une plus grande prévalence de la carie. Les études concernées ont aussi démontrées que les adolescents pratiquant un suivi régulier avaient moins de caries à soigner (2).

Chez l'adulte comme chez l'enfant, la difficulté de créer un échantillon représentatif limite la possibilité de réaliser des études mais on estime chez l'adulte qu'entre un tiers et la moitié de la population présente au moins une lésion carieuse non traitée. La catégorie socioprofessionnelle ou la situation de précarité restent liées à une plus grande atteinte par la carie (1).

Les personnes âgées, d'après les études, ont peu recours aux soins. La MSA en a réalisé une où il apparaît que 37% des séniors avaient une dent cariée. Dans les établissements de type EPHAD c'est 30 à 40% des résidents qui nécessitent des soins conservateurs (2).

A la vue de ces chiffres, la carie apparaît comme une pathologie touchant toutes les tranches d'âge, avec une fréquence accrue dans l'enfance due à la minéralisation incomplète des dents définitives post-éruptives (2). Cependant, au cours de la vie, la carie continue à toucher les populations notamment les plus précaires.

1.3.5.2 Au niveau mondial

Le rapport sur la santé bucco-dentaire de l'OMS publié en 2003 parle de 60 à 90% des élèves touchés par la carie et la grande majorité des adultes (10).

La prévalence reste plus élevée dans les groupes de population défavorisés des pays développés ou en développement (10).

On remarque une plus grande prévalence de la maladie carieuse dans les pays industrialisés type Europe et Amérique du nord. Elle l'est aussi dans plusieurs pays d'Asie (10).

Cependant, on l'observe dans une moindre mesure en Afrique. Le rapport évoque l'hypothèse qu'avec l'évolution des conditions de vie impliquant une plus grande consommation de sucres sans politique de fluoration dans les pays en voie de développement on devrait voir augmenter la prévalence de la maladie carieuse en Afrique et se rapprocher de celle des pays industrialisés (10).

Tout comme en France, on retrouve dans les études réalisées, le rôle important des facteurs sociaux, environnementaux et comportementaux dans le développement des maladies dentaires.

Il apparaît que l'avulsion ou l'absence de soins est fréquente dans les pays en développement où l'offre de soins reste encore limitée (10).

1.4. Surveillance et suivi

1.4.1. En France

De manière régulière, des plans de prévention de la santé bucco-dentaire sont proposés avec des objectifs chiffrés.

Les divers acteurs de santé tel que les caisses de santé régionales, l'UFSBD coordonnent des études afin d'avoir une idée de l'indice CAOD de l'échantillon choisi (3).

Ces données sont synthétisées sous formes de recommandations telles que les recommandations de bonne pratique publiées par la HAS (2) qui permettent, grâce à la synthèse des chiffres fournis par des études fiables et évalués, de donner une analyse transversale de la situation bucco-dentaire en France, de proposer des moyens de prévention novateurs ou au moins, de réévaluer ceux en place.

1.4.2. Au niveau mondial

En 1979, l'OMS s'est fixé comme objectif d'avoir un indice CAOD moyen ne dépassant pas 3 pour les enfants de 12 ans en 2000 (1).

Pour surveiller l'évolution des maladies bucco-dentaires, notamment la carie chez l'enfant en priorité, mais aussi chez l'adulte, l'OMS compile le maximum d'études épidémiologiques effectuées selon les critères et la méthodologie souhaitée. Elle obtient donc depuis 1969, une base

de données de l'indice CAOD moyen en fonction des pays (10). Elle peut produire des cartes du monde où cet indice CAOD moyen est représenté, à la fois pour les enfants de 12 ans et les adultes de la tranche 35-44 ans. Les dernières cartes publiées datent de 2003 (10).

Nous n'avons malheureusement aucun accès à la méthodologie employée par l'OMS pour analyser en détail la méthode de compilation de telles données.

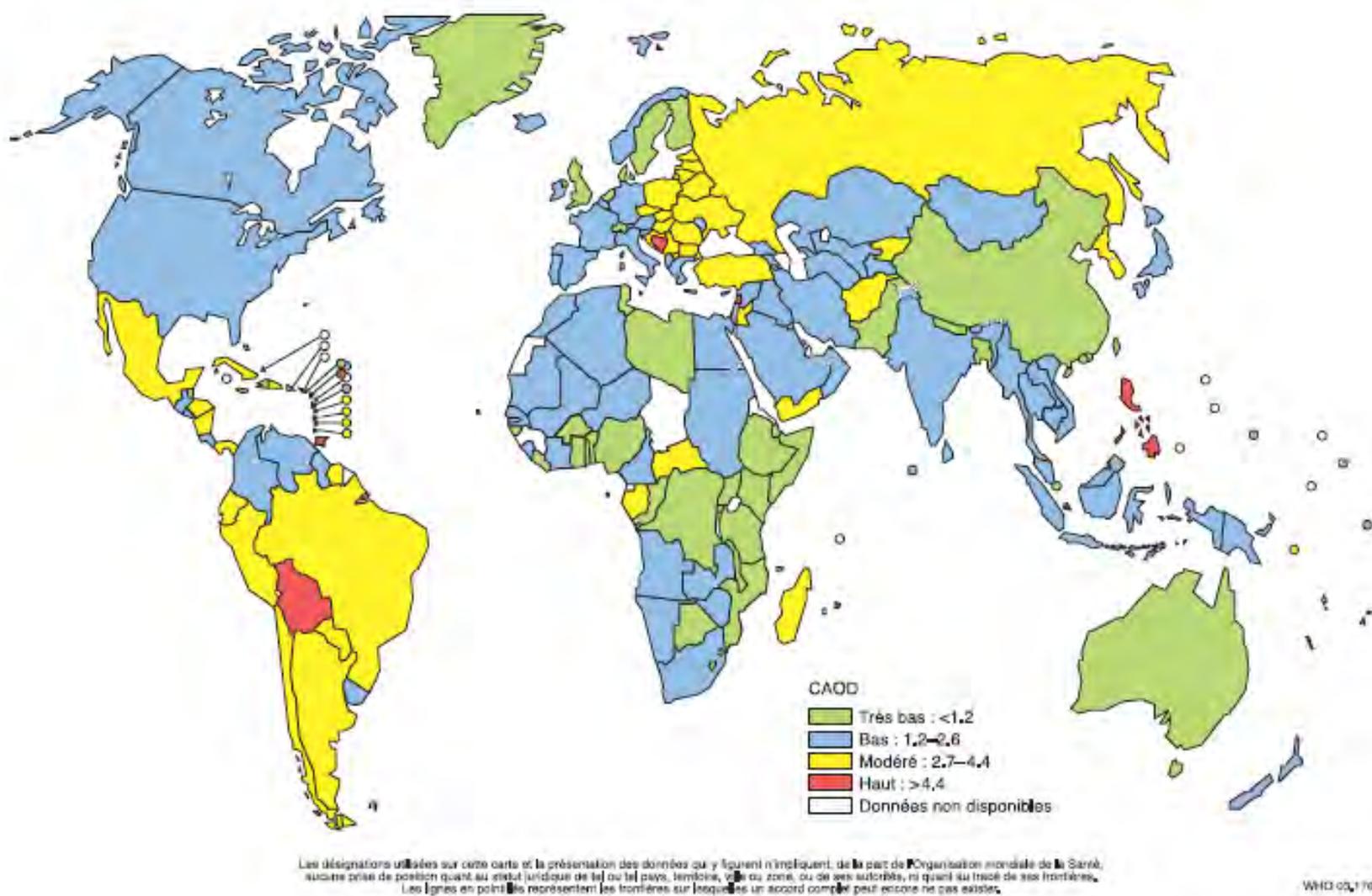
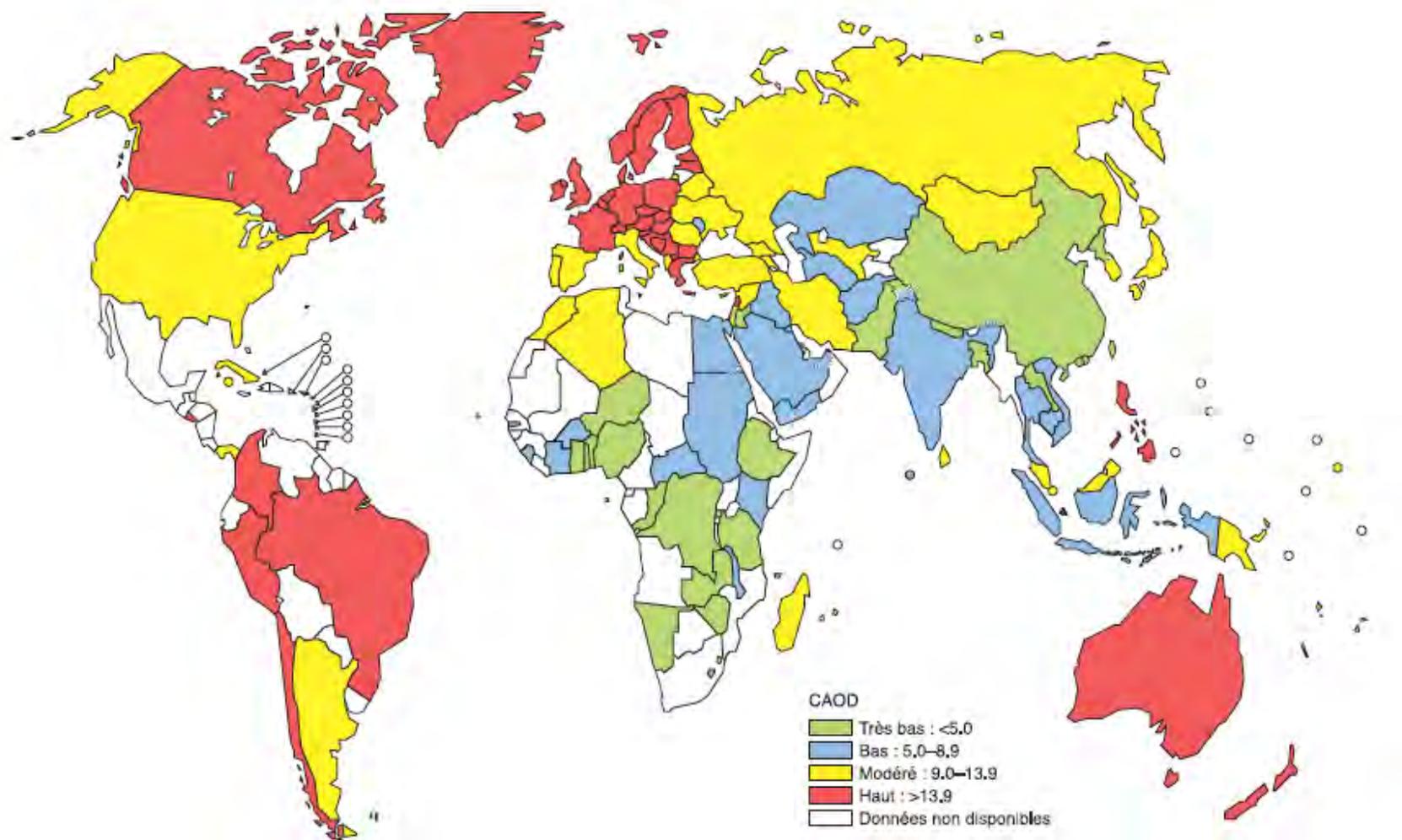


Figure 5 : indice CAOD moyen dans le monde chez les enfants de 12 ans (10).



Les désignations utilisées sur cette carte et la présentation des données qui y figurent n'impliquent, de la part de l'Organisation mondiale de la Santé, aucune prise de position quant au statut juridique de tel ou tel pays, territoire, ville ou zone, ou de ses autorités, ni quant au tracé de ses frontières. Les lignes en pointillés représentant les frontières sur lesquelles un accord complet peut encore ne pas exister.

WHO 03,177

Figure 6 : indice CAOD moyen dans le monde chez les adultes de 35- 44 ans (10).

2. Google® et son outil Google® Trends

2.1. Présentation de Google®

Google® est une entreprise fondée par deux étudiants américains alors âgés d'une vingtaine d'années, Larry Page et Sergey Brin, qui se sont rencontrés sur le campus de l'université de Stanford (Californie) en 1995 (15).

Le nom Google® dérive du terme "googol" (ou "gogol" en français), qui désigne le chiffre 10^{100} , soit 1 suivi de 100 zéros (15).

L'entreprise est officiellement créée en septembre 1998. Auparavant, Brin et Page travaillaient depuis 1996 sur un moteur de recherche, d'abord baptisé BackRub (littéralement "massage du dos") permettant de trouver rapidement les pages web les plus pertinentes à partir des termes recherchées (15).

L'efficacité de Google® repose sur un algorithme dont un des composants, le Page Rank, a été inventé par Larry Page. Le Page Rank a pour but de mesurer la popularité d'une page web, avec un score de 0 à 10, via notamment le nombre de liens hypertextes qui renvoient vers elle. Plus le score de popularité d'une page sera élevée, plus elle sera dans les premiers choix disponibles proposés par Google® (15). Ainsi, si la page est populaire sur un sujet, on peut dire par simplification qu'elle semble pertinente. Cependant, le Page Rank n'est qu'un composant de l'algorithme total qui a permis à Google® de devenir le leader des moteurs de recherche.

2.2. Google® en quelques chiffres

Les chiffres concernant l'entreprise sont édifiants. Ils démontrent la large suprématie de l'entreprise comme moteur de recherche (16). Ils prouvent aussi qu'il s'agit d'un acteur de premier plan dans le développement des technologies et des services liés au Web.

Ainsi en 2015, on comptait 88,66% des recherches sur Internet faites via Google® (17).

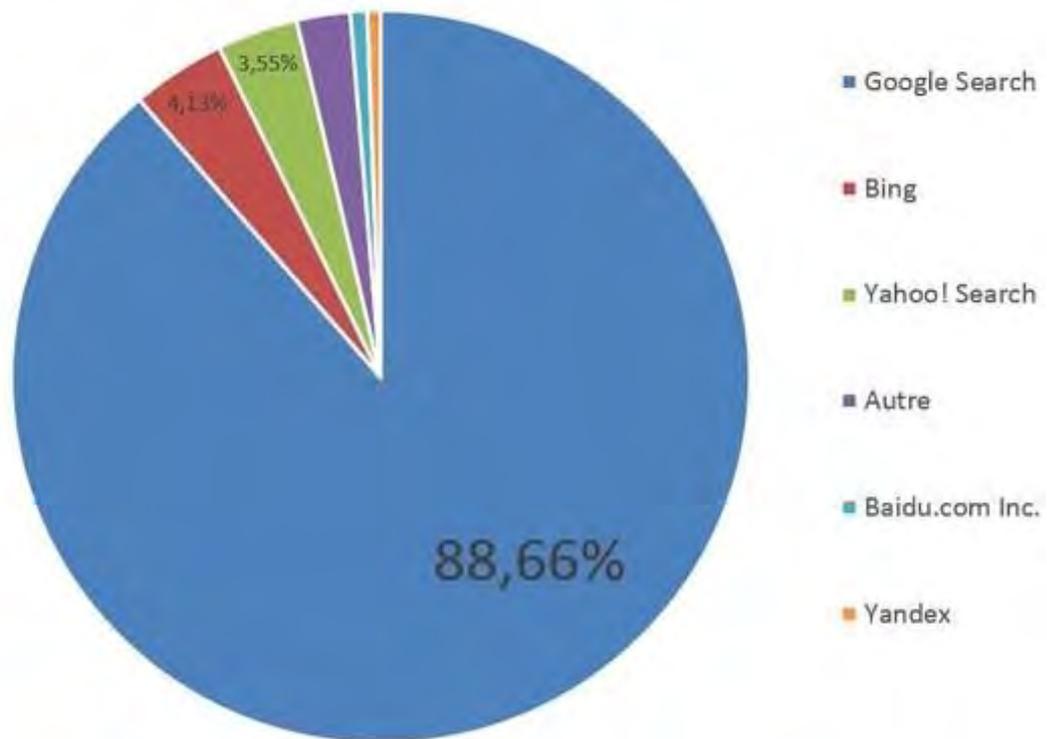


Figure 7 : parts de marché des moteurs de recherche en 2015 (17).

2.3. Les multiples outils de Google®

Cette thèse s'intéresse à Google® Trends, mais il s'agit là d'un des nombreux outils que propose Google®.

De manière non exhaustive, cette liste cite des services proposés par Google® (18):

- Google® : un moteur de recherche
- Gmail : une boîte mail pour recevoir et envoyer des courriels
- Google® Agenda : un agenda en ligne
- Google® Drive : un cloud de stockage
- Google® + : un réseau social
- Google® Sheets, Docs, Slides : des équivalents à un tableur, un traitement de texte et à un logiciel de présentation telle une suite bureautique
- Google® Chrome : un navigateur internet
- YouTube® : une plateforme de partage de vidéo
- Google® Scholar : une base de données de publications académiques.

- Et de nombreux autres services tels que de la cartographie, le référencement des pages web ou un logiciel en ligne de photo.

On compte ainsi plus de 220 outils proposés (19).

On le voit, l'entreprise propose une multitude d'outils touchant à des nombreux domaines du quotidien. Ainsi, la somme d'informations obtenue par Google® dans des domaines très variés, est gigantesque.

2.4. Google® Trends

Google® Trends ou Google® Tendances de recherche en français est lancé en 2006. Il est traduit en français en 2009 (20).

C'est un outil qui permet de connaître la fréquence et la répartition géographique avec laquelle un terme a été recherché dans Google®.

Pour le terme recherché, on obtient une courbe qui indique le nombre de recherches en fonction du temps et une carte montrant les régions du monde où ce dernier est recherché et dans quelle proportion.

On peut comparer l'évolution d'un à cinq termes différents.

Pour chaque terme, on va avoir un indice de popularité de 1 à 100, 1 étant un terme très peu recherché et 100 étant un terme très recherché.

2.4.1. *Origine des données*

Pour proposer ces courbes et cartes, Google® Trends va analyser un pourcentage des recherches effectuées sur Google® pour ce terme et va ainsi déterminer le nombre de recherche sur la période demandée (21).

Certaines données ne seront pas prises en compte :

- Terme de recherche faible : si le volume de recherche est trop faible pour le terme, il est considéré comme nul.
- Terme recherché de manière répétée : si un terme est recherché de manière répétée sur une courte période, il n'est pas inclus.
- Caractères spéciaux : l'apostrophe et les caractères spéciaux sont exclus.

2.4.2. *Ajustement des données*

Pour faciliter la comparaison des données, Google® Trends ajuste ces dernières.

Cet ajustement est nécessaire sinon les endroits du monde où le volume de recherche global est élevé, là où la majorité de la population a accès à Internet, arriveraient toujours en tête du classement même si le terme est moins recherché que dans un pays où il y a moins de volume de recherche global car la facilité d'accès à internet est moindre (22).

Ainsi, Google® Trends divise chaque point de données par le nombre total de recherches pour la période et la zone donnée. La popularité relative obtenue peut ainsi être comparée. Les nombres issus du calcul sont mis à l'échelle sur une plage de 1 à 100 (22).

2.4.3. *Export des données*

Google® Trends permet d'exporter les résultats obtenus lors de la recherche d'un terme sous la forme d'un fichier au format CSV (abréviation de Comma-Separate Value). Il s'agit de données tabulaires séparées par des virgules pouvant être traités par des tableurs.

Ce format de fichier est ouvert et peut donc être traité sans restriction par tout type de tableur acceptant ce format (Excel de Microsoft Office ou Maths d'Open Office par exemple).

2.5. Utilisation de Google® Trends

Pour utiliser Google® Trends, il faut se rendre à l'adresse suivante : www.google.fr/trends/ (23)

2.5.1. *La page d'accueil*

Sur la page d'accueil, en haut, on trouve la barre de recherche permettant de taper le ou les termes dont on veut connaître la popularité. La barre de recherche peut contenir jusqu'à 25 termes à la fois (23).

Un volet « Sélection des tendances » est consacrée aux sujets les plus recherchés du moment en France (dépend des paramètres fixés par l'utilisateur) ou en lien avec l'actualité internationale via des articles montrant des courbes ou des cartes pouvant être obtenus avec ces termes "en une". Le volet en dessous « Actualités du moment » présente un classement des

termes les plus recherchés en France (dépend des paramètres fixés par l'utilisateur) avec les tendances de recherche sur les dernières 24 heures avec une mise à jour toutes les heures.

En dessous de la barre de recherche, on trouve la date du jour, et à droite, deux filtres de recherche possible :

- Selon la catégorie : on peut choisir de voir les termes les plus recherchés du moment, toute catégorie confondue, ou seulement dans les domaines suivants : à la Une, divertissement, entreprise, santé, sciences/technologies, sports.
- Selon le pays : on peut voir les tendances de recherche en France ou dans d'autres pays : Allemagne, Argentine, Australie, Autriche, Belgique, Brésil, Canada, Chili, Colombie, Inde, Irlande, Italie, Japon, Malaisie, Mexique, Norvège, Nouvelle-Zélande, Pays-Bas, Pérou, Philippines, Pologne, Portugal, Royaume-Uni, Russie, Suède, Suisse, Turquie, Vietnam.

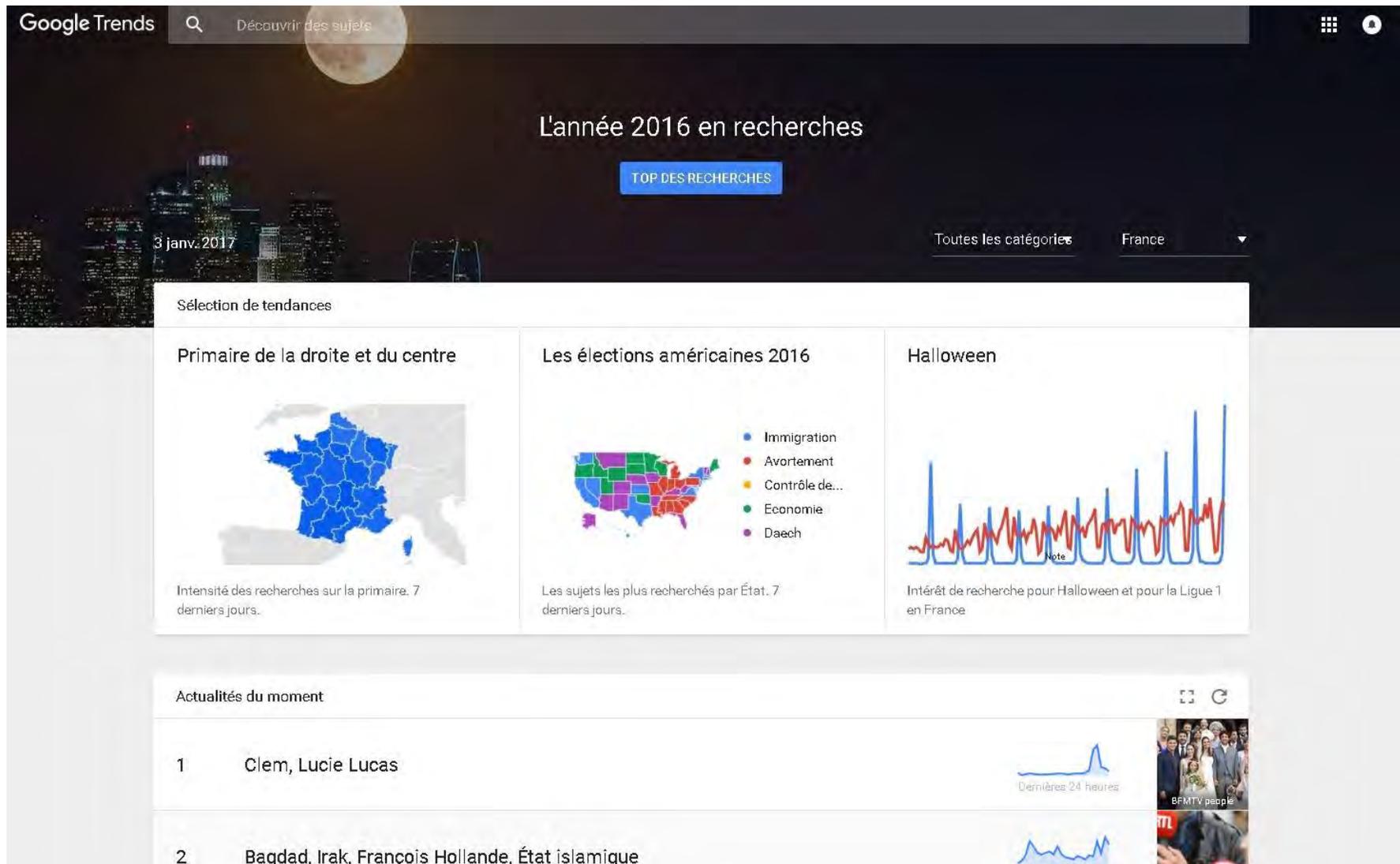


Figure 8 : page d'accueil de Google® Trends (23).

2.5.2. La page de résultat

Une fois la recherche effectuée, on obtient différents encarts indiquant comment le terme est recherché dans le temps et dans l'espace géographique. Des filtres divers peuvent être appliqués pour préciser la période, le lieu ou la catégorie de la recherche.

2.5.2.1 Evolution de l'intérêt dans le temps

Cette évolution est matérialisée par une courbe représentant en abscisse la période choisie et en ordonné l'indice de popularité de recherche (de 0 à 100). En passant la souris sur la courbe, on lit l'indice de popularité de chaque mois.

Les nombres qui s'affichent indiquent la popularité de recherche du terme sur Google® au fil du temps. Par rapport au mois où la valeur est la plus élevée, les valeurs pour les autres mois sont donc exprimées en proportions relatives.

Sans réglage préalable, la période est celle des cinq dernières années jusqu'au mois précédant le jour de la recherche.

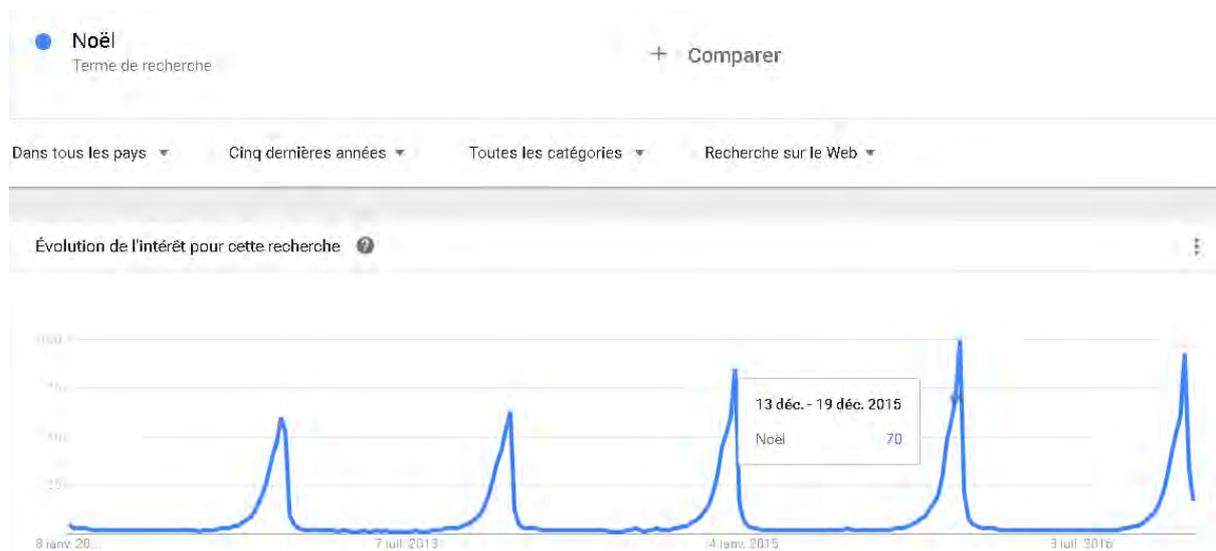


Figure 9 : évolution dans le temps de la popularité de recherche du terme « Noël » (24).

Les options « Titres des actualités » et « Prévisions » ont disparu de la version de Google Trends® proposé depuis l'été 2016.

Durant notre étude, selon le terme, on pouvait avoir des lettres qui apparaissaient à certains endroits de la courbe. Elles renvoyaient vers des articles de presse ; ceci dans le but d'essayer d'expliquer la popularité soudaine d'un sujet. Il suffisait de cocher la case « Titre des actualités ».

Il existait une case « Prévisions », non disponible à chaque recherche. Il fallait que la période concernée soit assez longue pour que la prédiction ait une forte chance de se réaliser. Il s'agissait d'une estimation mathématique de la popularité du terme recherché dans l'avenir immédiat.

2.5.2.2 Répartition dans l'espace géographique

Une carte du monde modélise la répartition géographique de la popularité de recherche du terme avec une plage de couleur allant du bleu clair au bleu soutenu selon que le terme est plus ou moins recherché dans le pays. En passant la souris sur chaque pays, on obtient l'indice de popularité (de 0 à 100). Les pays de très petite superficie (type Monaco, Luxembourg) sont exclus.

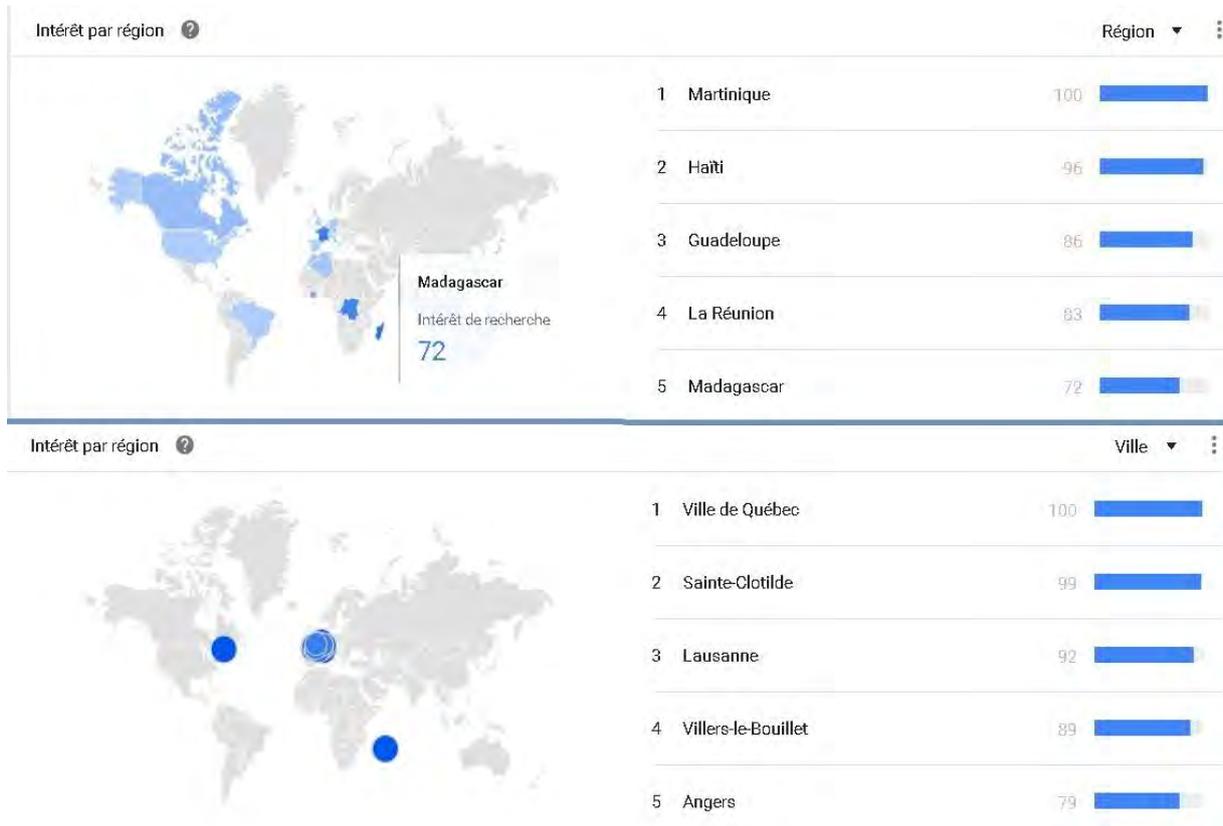


Figure 10 : évolution dans l'espace géographique de la popularité de recherche du terme « Noël » au cours des cinq dernières années (24).

A droite de la carte du monde, de manière non exhaustive, une liste de pays avec l'indice de popularité du terme recherché, accompagné d'une réglette plus ou moins coloré, selon que l'indice est plus ou moins important.

On peut aussi choisir d'afficher les régions ou les villes du monde où le terme est le plus recherché. La liste des villes fonctionne de la même manière que la liste des pays, avec le même système de réglette et de codes couleur. La liste des villes est elle aussi non exhaustive.

Pour avoir la totalité des résultats par ville et par pays, il faut exporter les résultats sous forme de fichier .csv pour pouvoir consulter la liste complète des villes et des pays concernés dans un tableur en cliquant sur les trois points alignés en haut à droite.

Il est parfois possible de cliquer sur un pays pour voir de manière plus précise la répartition de la recherche dans les subdivisions administratives, voire les villes.

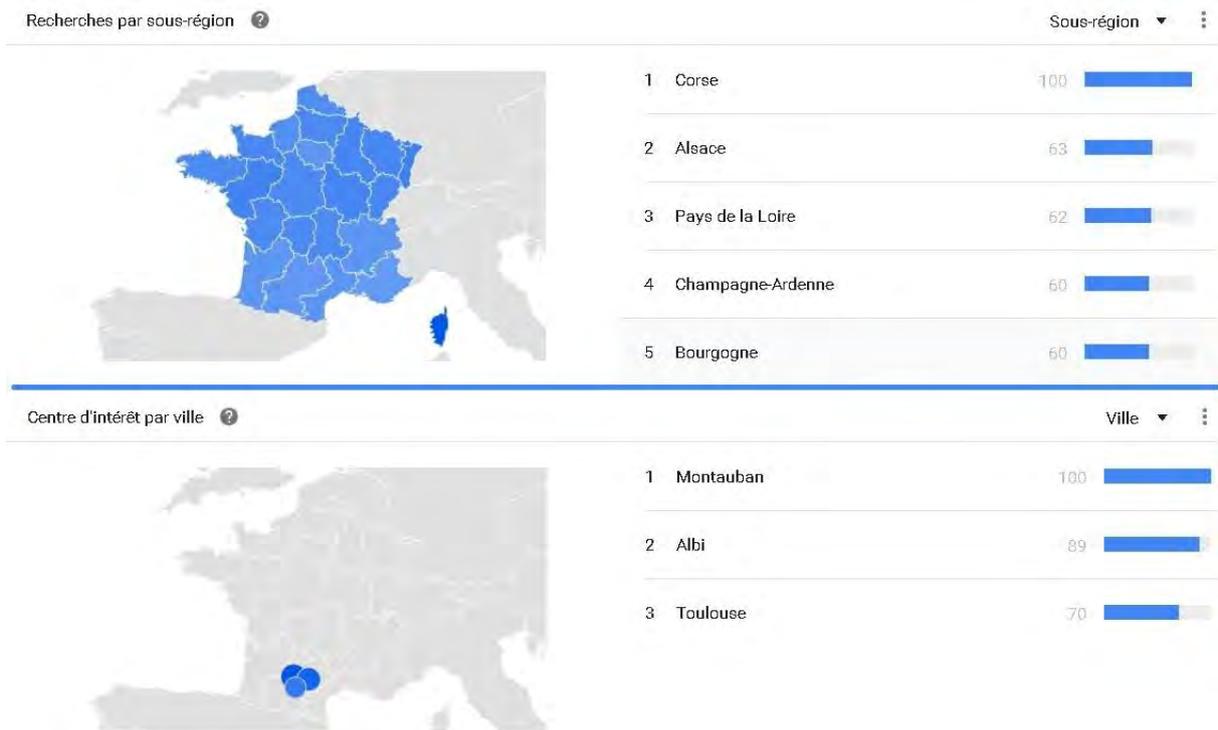


Figure 11 : évolution dans l'espace géographique de la popularité de recherche du terme « Noël » au cours des cinq dernières années au niveau de la France et de la région Occitanie (24).

Avant la nouvelle version de l'été 2016, la possibilité était offerte de visualiser l'évolution de la popularité mois par mois en cliquant sur « Voir l'évolution au fil du temps ».

2.5.2.3 Autres termes associés lors de la recherche

Le dernier encart propose les sujets et les requêtes associées qui font l'objet de recherches étroitement associées avec le terme recherché (24).



Figure 12 : sujets et requêtes associés au terme Noël lors des recherches au cours des 5 dernières années (24).

Le pourcentage représente la croissance en volume sur la période choisie par rapport à la période précédente. S'il est écrit « Record » au lieu d'un pourcentage c'est que la hausse dépasse 5000 % (25).

On peut aussi avoir les sujets et requêtes associés les plus fréquemment en déroulant en cliquant sur la flèche. Ceci va donner une tendance de fond plus générale sur les sujets et requêtes associées au sujet.

2.5.3. Astuces de recherche

2.5.3.1 Absence de résultats

Pour des informations sur les recherches en progression ou qui sont les plus fréquentes, il faut un recul de 7 jours minimum. Les tendances ne pourront pas être données pour une période plus récente (26).

L'absence de graphique lors des résultats est causée par un volume de recherche trop bas. Afin d'obtenir un volume plus important on peut (26):

- Elargir la recherche si elle comporte trop de termes la rendant trop précise
- S'assurer de l'absence de fautes d'orthographe empêchant de reconnaître le terme
- Etendre la période de recherche pour augmenter le volume

2.5.3.2 Précision de la recherche

Lorsque le ou les termes sont tapés sans ponctuation dans la barre de recherche, le résultat portera sur toutes les recherches où chacun des termes a pu apparaître ensemble (27).

Par exemple pour « fleur lotus », Google® Trends comptabilise les recherches pour « fleur de lotus ». D'autres termes non mentionnés peuvent aussi être comptabilisés comme dans « planter une fleur de lotus » ou « fleur de lotus rouge ».

On peut souhaiter rechercher le terme ou le groupe de terme de manière exacte. Pour cela on devra l'encadrer de guillemets dans la barre de recherche (27).

On peut souhaiter rechercher plusieurs termes à la fois mais désirer exclure les données où les termes sont ensemble. Pour cela, on devra écrire « fleur+lotus ». Les résultats porteront sur les recherches de « fleur » ou de « lotus » seuls.

Le signe « - » va permettre d'exclure des résultats toutes les recherches où le terme après le signe moins apparaît (27). Dans notre cas « fleur – lotus » ôterait tous les résultats contenant « lotus ».

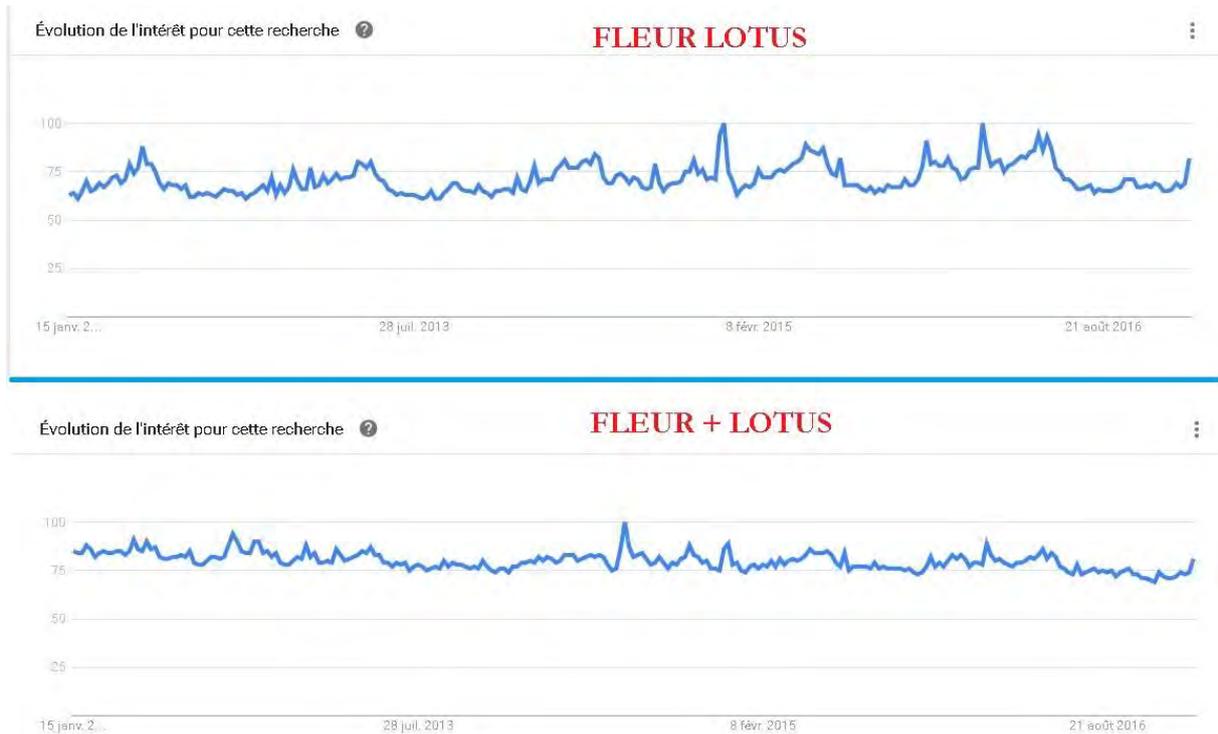


Figure 13 : différences de résultats obtenus selon le niveau de précisions de la recherche (28, 29).

2.5.4. Filtres disponibles

Lorsqu'on lance une recherche pour un ou plusieurs groupes de terme, les filtres disponibles apparaissant sont :

- Un filtre par pays
- Un filtre par période
- Un filtre par catégorie
- Un filtre par type de recherche

Dans ce cas, ils s'appliquent à tous les groupes de termes recherchés, sans pouvoir les comparer les uns aux autres.

Le filtrage global va être intéressant si on compare des groupes de termes non plus identiques mais différents ; ceci dans l'optique de voir s'il existe un lien entre l'évolution de la popularité de ces termes.

Par exemple, chercher « dentifrice » et « brosse à dents » et observer si sur une même période et sur un même endroit s'ils sont recherché de manière similaire où si les personnes ont tendance à plus chercher l'un des termes.

Pour appliquer un filtrage par période ou par lieu à un seul des groupes de termes, il faut cliquer sur les trois petits points alignés verticalement à côté du terme concerné.

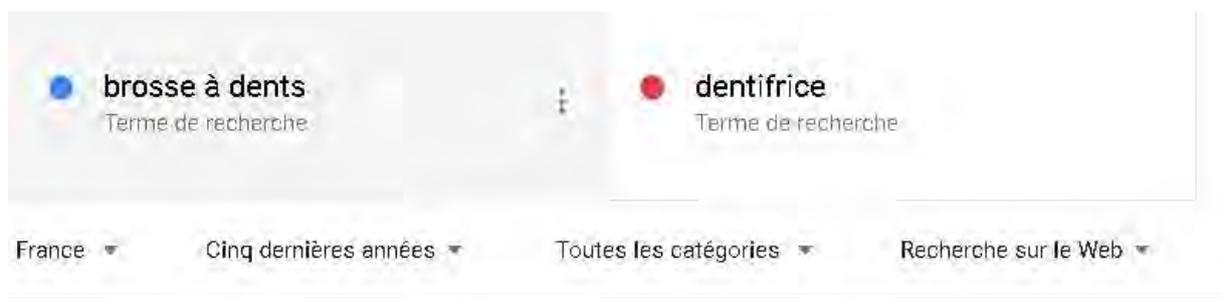


Figure 14 : filtrage global disponible sous la barre de recherche et filtres de période ou de lieu appliqués à un groupe de terme en cliquant sur les trois points à droite de ce dernier (30).

2.5.4.1 Filtrage par période et par pays

On peut observer la manière dont tous les termes sont recherchés dans un pays particulier ou durant une période précise. Ce filtrage peut être appliqué à seulement un des termes via les trois points gris alignés verticalement à droite du terme.

2.5.4.2 Filtrage par catégorie

Le filtre par catégorie propose de centrer la popularité du groupe de termes seulement sur un domaine. Les domaines proposés par Google Trends[®] sont assez divers tels que : actualités, finances, maison et jardinage ou sports (liste non exhaustive). Chaque catégorie ayant elle-même des sous catégories permettant de préciser au maximum à quel « univers » appartient le groupe de terme (31).

Ce filtrage va être utilisé si le mot peut avoir plusieurs sens. Par exemple, le terme « jaguar » est à la fois un animal mais aussi une marque de voiture. Si on veut préciser sa recherche et ne pas voir les résultats mêlant les deux domaines : animal et automobile on aura accès aux catégories « animaux et animaux de compagnie » et « automobiles et véhicules ».

Ce filtrage implique d'avoir un mot avec plusieurs sens et que ce qu'il désigne ne soit pas trop abstrait pour pouvoir s'affilier sans trop de biais dans une des catégories proposées.

La liste des catégories et sous-catégories proposées par Google[®] Trends dans le domaine de la santé est disponible en annexe n°1.

2.5.4.3 Filtrage par type de recherche

Le filtre par type de recherche permet de centrer l'évolution de la popularité sur le moteur de recherche Google[®] en général ou seulement sur un des domaines qu'il propose : recherche d'images, recherche d'actualité, Google[®] Shopping (regroupement des offres commerciales) et recherche YouTube.

Si on ne filtre pas par type, on a donc l'évolution de la popularité dans Google[®] en général mais on ne sait pas si cela a poussé les utilisateurs à chercher plus souvent des images (recherche d'images), des informations dans la presse numérique (recherche d'actualité), à vouloir voir s'il existait une offre commerciale (Google[®] Shopping) ou à regarder des vidéos (recherche YouTube[®]).

2.5.5. Comparaison des termes

2.5.5.1 Comparaison des termes

Google Trends® propose de comparer jusqu'à 5 groupes de termes différents pouvant contenir 25 termes au maximum.

Une fois la comparaison de 2 à 5 termes lancée, on va obtenir les mêmes trois rubriques :

- Evolution de l'intérêt dans le temps
- Répartition dans l'espace géographique
- Termes de recherche associés

La différence notable est que l'encart montrant l'évolution de la popularité dans le temps va pouvoir comporter jusqu'à 5 courbes, soit une courbe par groupe de termes comparé. Chaque groupe de termes va avoir une couleur différente attribué. On va ainsi pouvoir comparer pour chaque groupe de termes l'évolution de la popularité dans le temps les uns par rapport aux autres

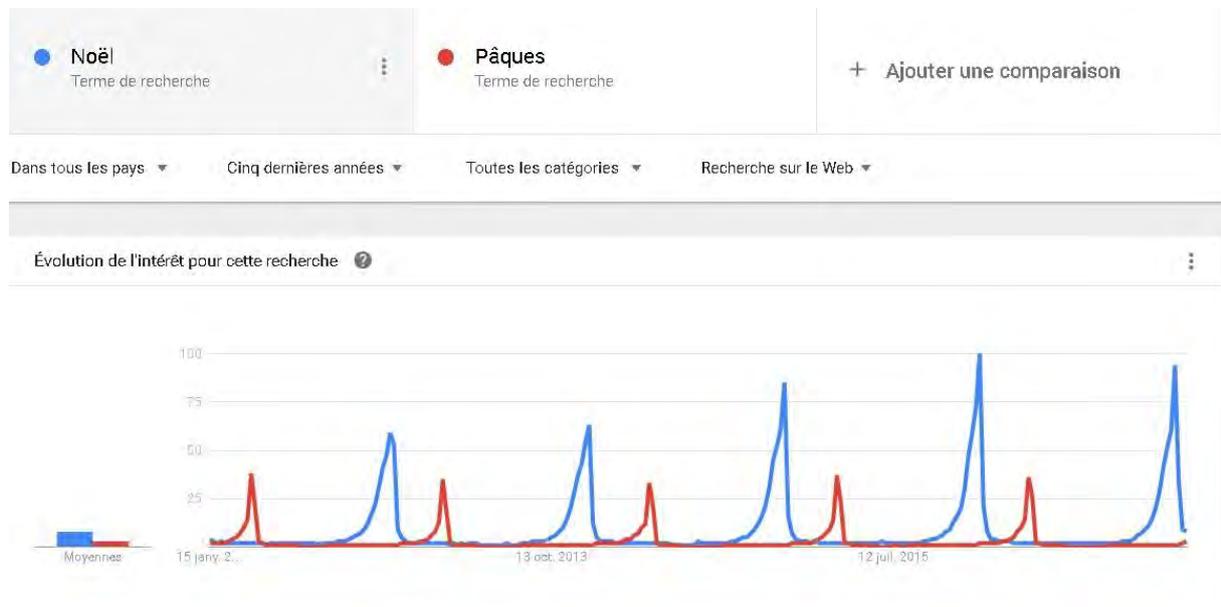


Figure 15 : comparaison de l'évolution de la popularité de recherche entre « Noël » en bleu et « Pâques » en rouge au cours des 5 dernières années (32).

La popularité moyenne de la recherche apparaît à gauche de la courbe. On peut voir deux à cinq histogrammes selon le nombre de groupe de termes comparé et dont la couleur est la même que celle attribué à chaque groupe de terme sur la courbe. Cet histogramme fait une

moyenne de la popularité de la recherche sur le temps donné. On peut comparer la popularité moyenne des groupes de termes les uns par rapport au autres grâce aux histogrammes.

Pour la répartition géographique et les requêtes associées à chaque groupe de termes, la seule différence est qu'on obtient des cartes distinctes pour chacun des termes comparés.

Une grande carte synthétise la popularité de chaque terme pour tous les pays concernés. Si les mêmes zones géographiques sont concernées, les couleurs vont se superposer. La couleur dominante sera celle du lieu où la popularité de recherche aura été la plus forte. Il faudra passer la souris sur chaque pays coloré pour voir l'indice de popularité de chaque pays en fonction du lieu.

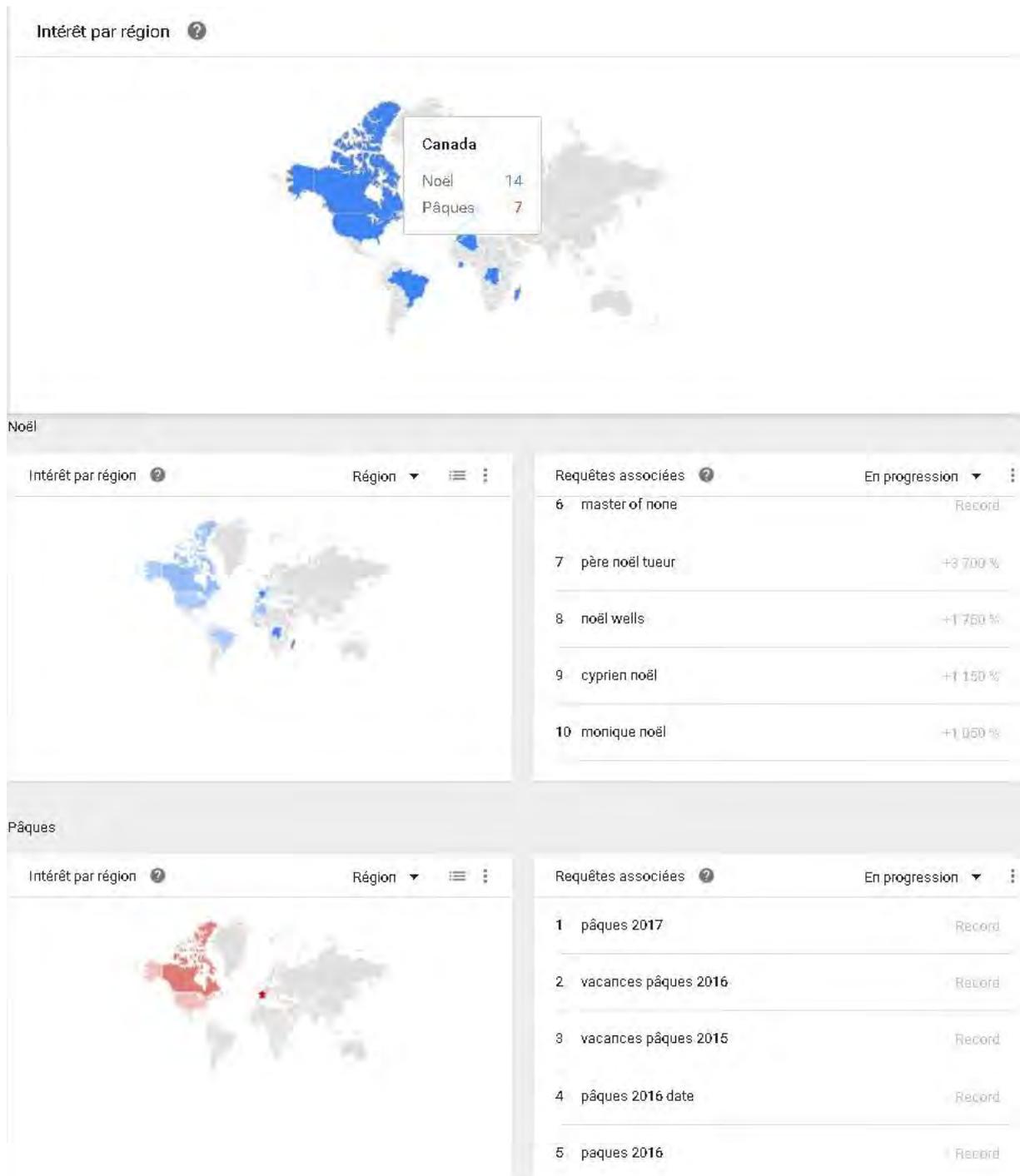


Figure 16 : cartes et requêtes associées obtenues lors de la comparaison des termes « Noël » et « Pâques » (32).

2.5.5.2 Comparaison dans l'espace géographique

On peut aussi choisir d'observer la manière dont est recherché un terme ou un groupe de termes en fonction des lieux. De la même façon que pour la comparaison de groupe de termes différents, on peut entrer jusqu'à cinq lieux différents.

Pour cela, dans la barre de recherche on rentre d'abord le groupe de terme concerné. Ensuite, on le retape de manière identique dans l'encart de recherche supplémentaire proposé, et cela, jusqu'à cinq groupes de terme maximum.

On va ensuite sélectionner au niveau de chaque terme le filtrage désiré, matérialisé par trois petits points à côté du terme.

Cet onglet va permettre d'accéder à l'option suivante : modifier les filtres.

On va pouvoir jouer sur le filtre du lieu en choisissant pour chaque groupe de termes un pays différent parmi tous ceux proposés.

Par défaut, quand on lance une recherche sans utiliser de filtre géographique, on obtient les résultats pour tous les pays et donc une carte mondiale.

Une fois les pays souhaités choisis, on obtient :

- Répartition dans l'espace géographique
- Requêtes associés

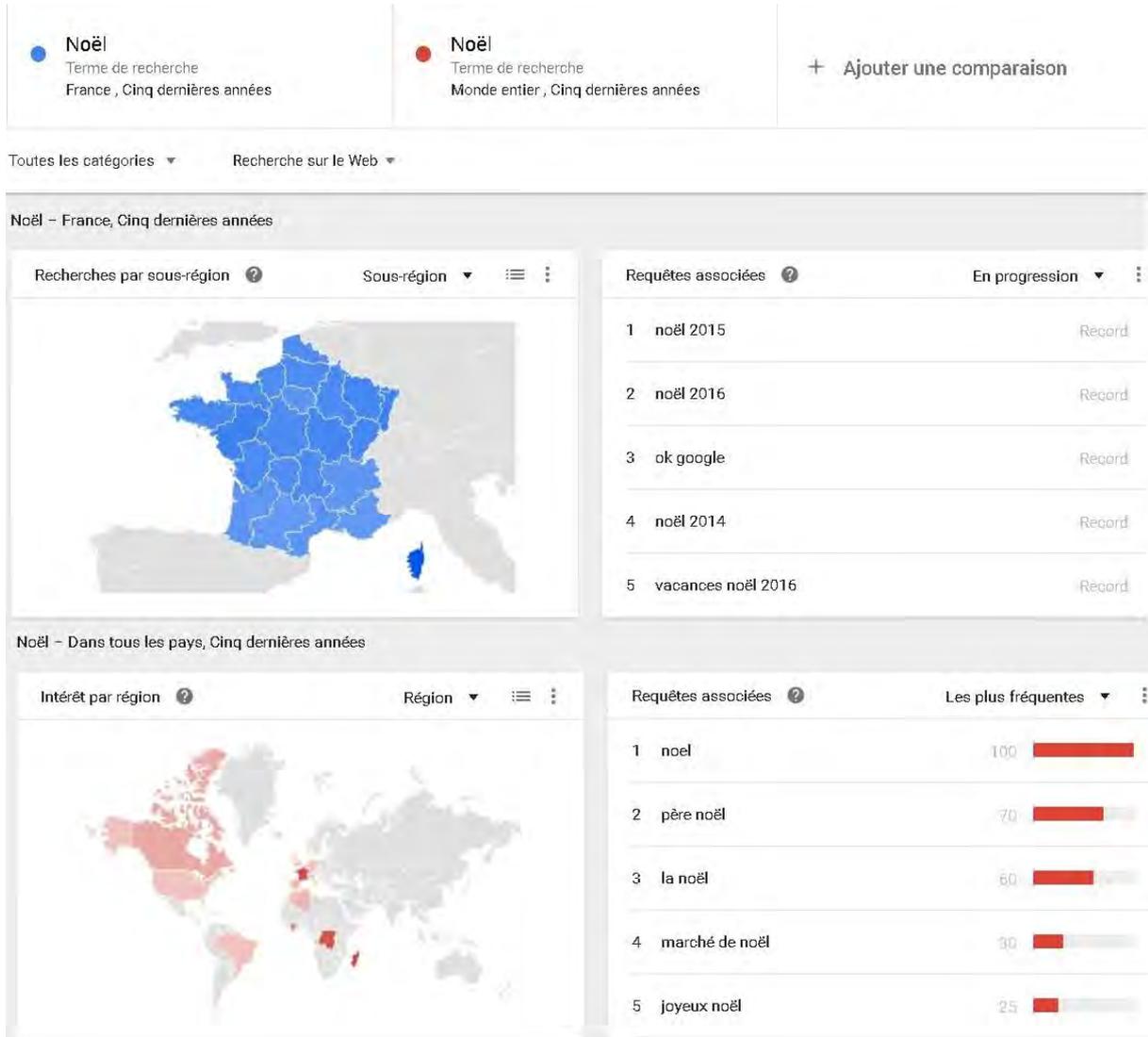


Figure 17 : comparaison de la popularité de recherche du terme « Noël » entre la France et le monde (33).

La répartition de la recherche dans l'espace géographique pour chaque lieu est visible sur une carte si les résultats sont suffisants pour en matérialiser une.

Enfin, pour chaque terme, on obtient aussi les requêtes associées les plus fréquemment à ce groupe de terme mais aussi les requêtes associées en progression, qui peuvent mettre en évidence un épiphénomène ou au contraire une tendance de fond en lien avec le groupe de terme recherché.

2.5.5.3 Comparaison dans le temps

Un autre filtrage que le filtrage géographique est possible, il s'agit du filtrage par période. Google® Trends propose de visualiser comment le groupe de terme a été recherché pendant différentes périodes.

De la même manière que pour toute comparaison de terme ou de lieux, on peut rentrer jusqu'à 5 groupe de termes maximum et donc comparer jusqu'à cinq périodes différentes.

Pour cela, comme pour appliquer un filtre de lieux, on tape le groupe de terme de manière identique de deux à cinq fois dans la barre de recherche. Ensuite, on filtre la période à l'aide de l'onglet de filtrage matérialisé par les trois points.

Une fois « modifier les filtres » sélectionné, on peut choisir la période voulue, soit parmi celle proposée par Google® Trends: de 2004 à ce jour ; cinq dernières années, 12 derniers mois, 90 derniers jours, 30 derniers jours ; 7 derniers jours ; moins d'un jour ; moins de 4 heures ; moins d'une heure. On peut aussi personnaliser la période désirée en tapant précisément les dates ou choisir de voir les résultats pour une année complète.

Par défaut, la période de recherche est fixée aux cinq dernières années.

Il est à noter que Google® ne va pas appliquer le même type de calcul de données selon qu'on recherche sur une période ancienne ou plus récente. Il est indiqué : « Nous calculons les données pour la semaine passée différemment des données plus anciennes. Nous ne pouvons donc pas comparer les deux ensembles de données. »

Les données récentes sont toutes celles remontant à 7 jours inclus avant la date de la recherche. Au-delà, cela est considéré comme des données anciennes. On ne peut donc pas comparer la popularité d'un groupe de terme calculé sur la période englobant la semaine précédente avec la popularité de ce même groupe de terme sur l'année en cours par exemple.

Une fois les périodes à comparer choisies, de la même manière que pour comparer des termes ou des lieux on obtient les rubriques suivantes :

- Evolution de l'intérêt pour cette recherche
- Répartition dans l'espace géographique ou intérêt par région
- Requêtes associés

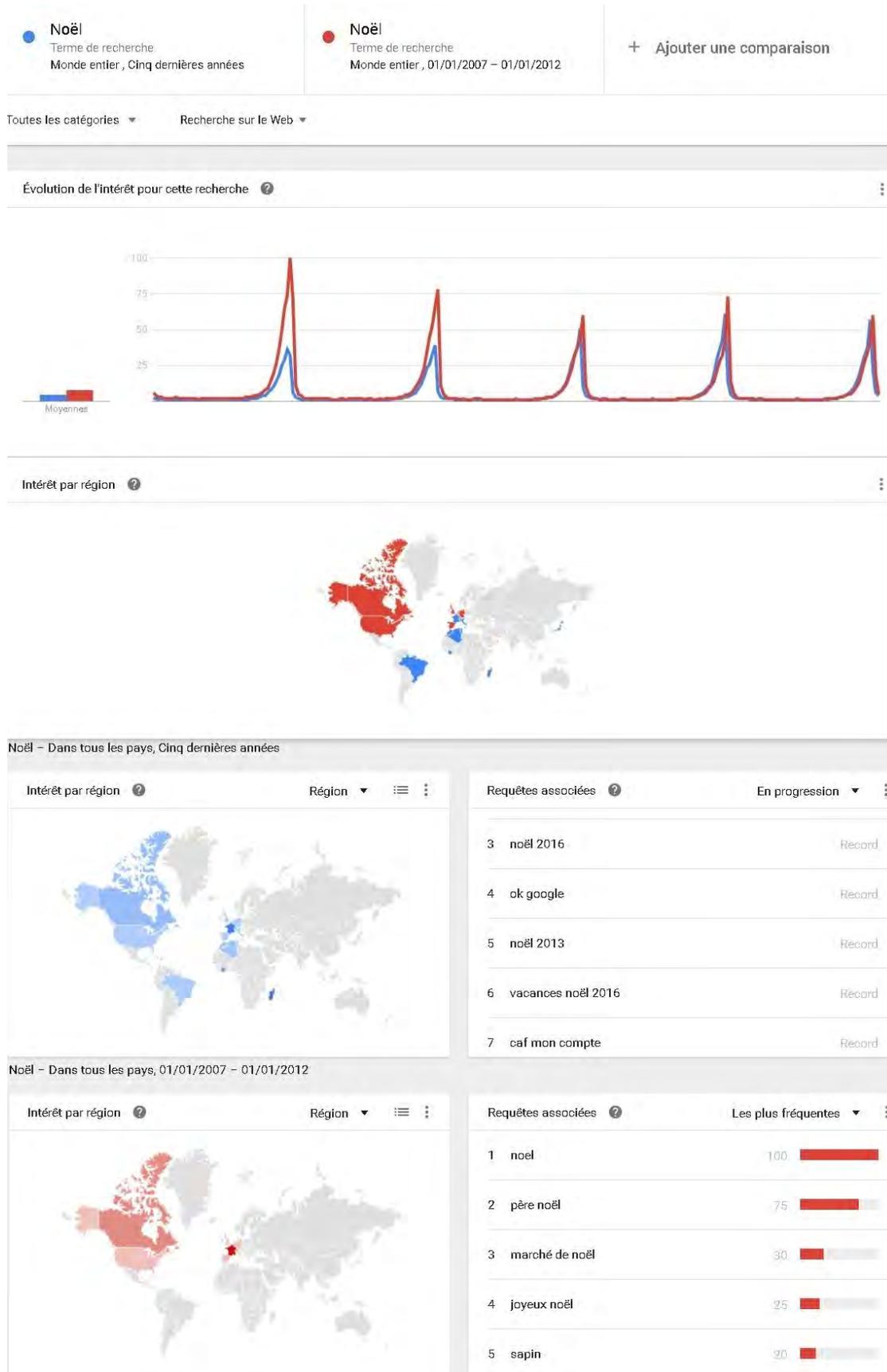


Figure 18 : comparaison de la popularité du terme de recherche « Noël » entre 2007 et 2012 et les cinq dernières années (34).

L'évolution de l'intérêt de la recherche dans le temps est toujours indiquée par une courbe avec pour chaque période recherchée une couleur différente attribué.

Il est à noter que l'obtention de courbe pour l'évolution de l'intérêt n'est possible que si on a choisi des périodes équivalentes à comparer en termes de durée. On va pouvoir avoir des courbes si on compare l'évolution sur deux années différentes. Cela ne sera pas possible si on compare l'évolution dans le temps pour une année et pour les cinq dernières années car l'échelle n'est plus la même.

Si on obtient une courbe, on aura aussi à gauche, la moyenne de popularité de recherche du groupe de terme pour la période. On pourra ainsi voir si le groupe de termes a été plus fréquemment cherché une année comparativement à une autre.

On retrouve ensuite la répartition dans l'espace géographique où une première grande carte synthétise dans quels endroits le groupe de terme a été le plus fortement recherché en fonction des périodes choisies.

Il faut savoir que si les mêmes zones géographiques sont concernées, les couleurs vont se superposer. La couleur dominante sera celle de la période où la popularité de recherche aura été la plus forte. Il faudra passer la souris sur chaque pays coloré pour voir l'indice de popularité de chaque pays en fonction de la période.

On obtient aussi pour chaque période une carte indiquant l'indice de popularité du terme dans chaque lieu concerné durant la période choisie.

Enfin, pour chaque période, on obtient aussi les requêtes associées les plus fréquemment à ce groupe de terme mais aussi les requêtes associées en progression, qui peuvent mettre en évidence un épiphénomène ou au contraire une tendance de fond en lien avec le groupe de terme recherché.

2.5.5.4 Comparaison de termes en langues différentes

Il peut arriver que l'on souhaite comparer le même groupe de termes dans des langues différentes dont certaines ne seront pas composées de caractère latins. On pense par exemple aux idéogrammes de type japonais ou chinois ou au sanscrit.

Si on recherche le mot « fleur » qui s'écrit « 花 » en japonais et qu'on veut comparer sa popularité de recherche avec celle en Italie, il est très probable que si nous ne tapons que

l'idéogramme « 花 » on aura un résultat quasi nul pour l'Italie. En effet, la population italienne aura tendance à rechercher le terme « fiore » pour « fleur » en italien.

Pour réaliser une comparaison entre des groupes de termes dont certains ne sont pas composés de caractères latins, il faudra taper les termes dans chacune des langues en les séparant par un signe plus (27). Dans notre exemple, ce serait : « fiore+« 花 ».

Ceci va nous être utile dans notre étude pour balayer le maximum de pays concernés par la carie.

2.5.5.5 Comparaison sujet / terme

Google Trends® différencie les termes et les sujets (35).

Un terme ou groupe de terme est un ou plusieurs mots qui lorsqu'on le recherche va faire ressortir tous les résultats où le(s) mot(s) sont contenus sans discrimination aucune.

Le sujet, quant à lui, va être proposé par Google® et va regrouper un concept identique dans toutes les langues. Si on cherche « Noël » comme un sujet et non plus comme terme, il s'agit du sujet « Noël » en tant que fête ».

Ainsi, les résultats comptabiliseront la popularité de la recherche de « Noël » pour tout ce qui a trait à Noël comme concept de fête et non pas toutes les recherches qui contiennent le terme Noël.

On va reconnaître les sujets proposés quand Google Trends®, au lieu d'indiquer « terme de recherche » sous le terme, va le classer dans une catégorie disponible.

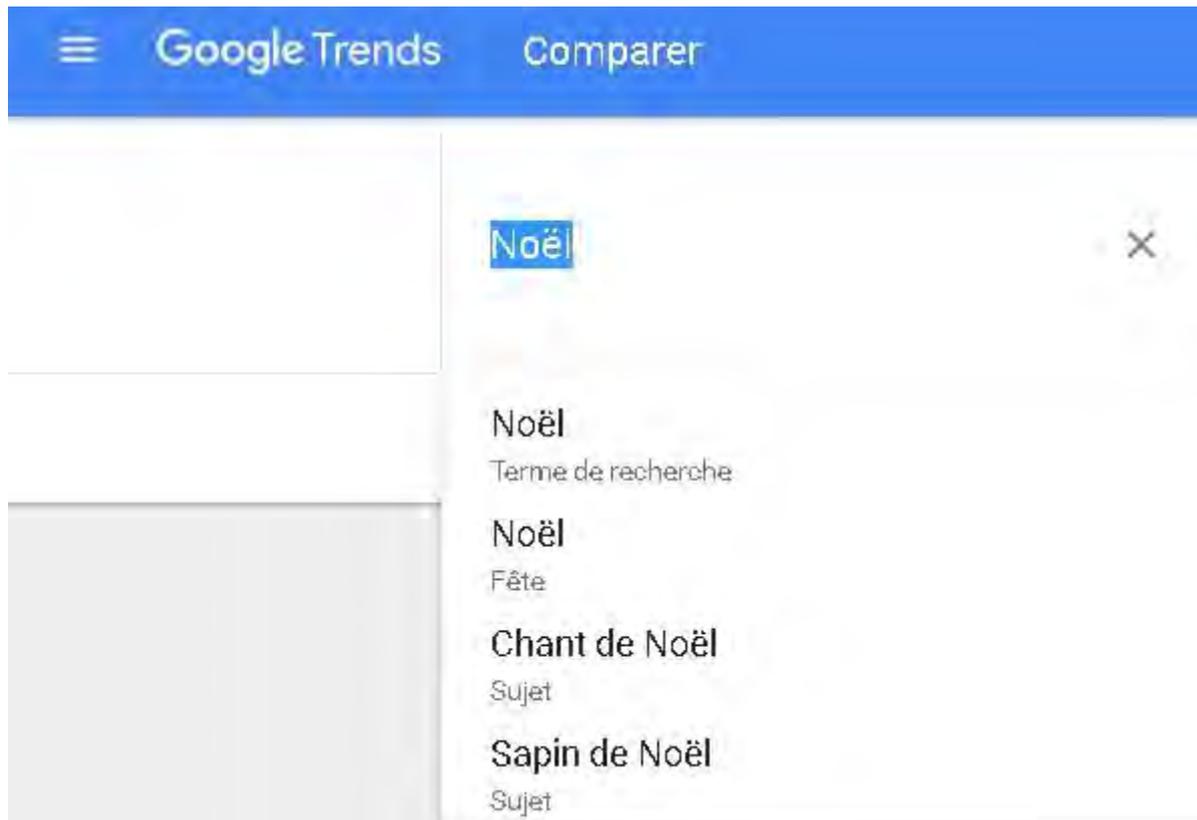


Figure 19 : « Noël » à la fois terme classique ou sujet (24).

3. Big Data et santé

Depuis quelques années, on observe une augmentation des études s'intéressant au Big Data dans le domaine de la santé. Les chercheurs ont voulu savoir si ce gigantesque volume d'informations disponible permettait d'obtenir des instantanés voir des prédictions sur certaines maladies. Tout cela dans une volonté d'anticiper plus rapidement les situations sanitaires qui entraînent une forte mobilisation tels les épidémies de grippe ou tout autre type de maladie saisonnière (36).

Si l'anticipation peut se faire plus rapidement et de manière plus globale, l'analyse des informations obtenues via le Big Data, avec des outils appropriés, permettraient de devenir une aide dans le domaine de la santé publique (37).

3.1. Définition du Big Data

Le Big Data est un terme générique pour définir un très grand volume d'informations complexe. Ces informations concernent des domaines de connaissances très variés tel que la santé, l'environnement, la finance ou les sciences de l'information (38).

Ce terme a d'abord été utilisé dans le domaine des sciences pour qualifier une quantité d'informations nécessitant des « supercomputers », ordinateurs conçus spécifiquement pour obtenir une grande puissance de calcul (utilisation pour les prévisions météorologiques par exemple) (39).

Aujourd'hui, cette définition peut sembler dépassée tant le phénomène de Big Data concerne non plus seulement les sciences mais trouve des applications dans des domaines variées (39).

Dans leur article paru en 2012, pour D. Boyd et K. Crawford, le Big Data n'est plus seulement une quantité d'informations qui demande à être analysée mais aussi un phénomène technologique, culturel concernant l'acquisition de savoirs (39).

3.2. Le monde médical et le Big Data

3.2.1. Un intérêt accru pour le Big Data

PubMed est reconnu comme le principal moteur de recherche bibliographique dans les domaines de la biologie et de la médecine.

Si on recherche le nombre d'articles publiés sur PubMed contenant le terme « Big Data » dans leur titre depuis les 5 dernières années, soit entre 2010 et 2015, on observe une forte augmentation des publications. Ceci semble indiquer un intérêt pour le sujet de la part des chercheurs.

Pour obtenir ce graphique (**figure 20**), nous avons fait rechercher à PubMed tous les articles contenant dans leur titre « Big Data ». Le terme est mis entre guillemets dans la barre de recherche pour que les titres contenant seulement « big » ou « data » ou un titre où les deux termes ne sont pas accolés, ne soient pas comptabilisés. Nous avons centré la période de recherche sur les 5 dernières années.

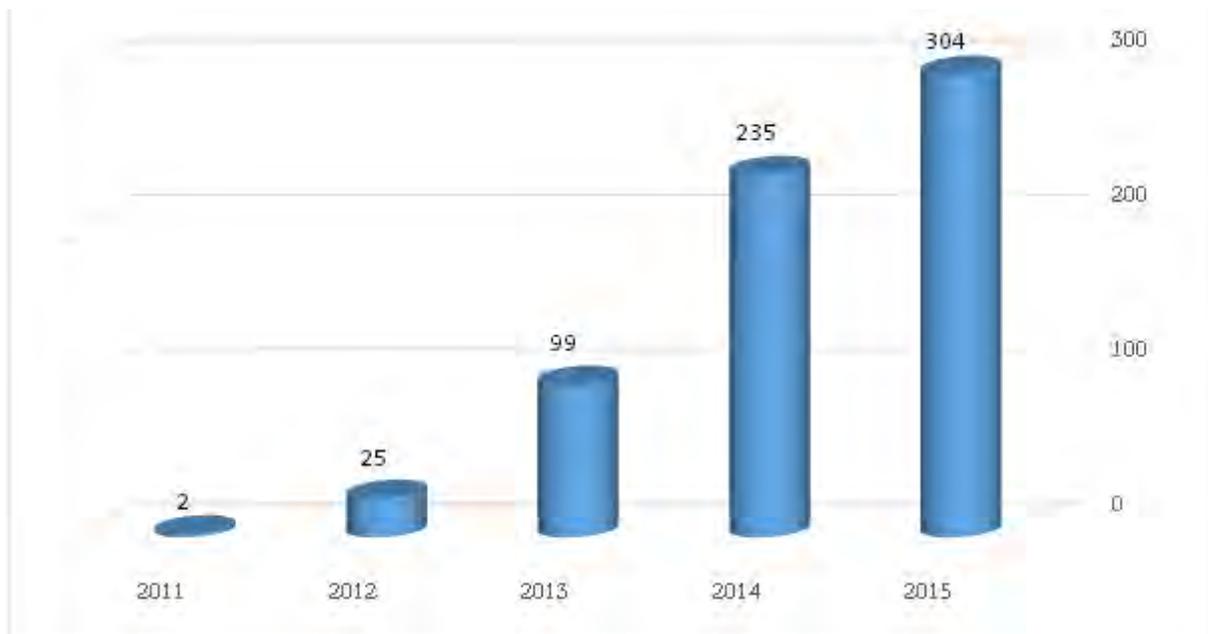


Figure 20 : nombre d'articles publiés sur PubMed concernant le Big Data au cours des 5 dernières années.

On observe que depuis 5 ans les chercheurs dans le domaine de la biologie et de la santé publient chaque année de plus en plus d'articles faisant mention du Big Data. Il y a donc fort à

parier que ce sujet éveille leur curiosité et prend de plus en plus d'ampleur dans le domaine médical.

3.2.2. Applications en santé publique : des potentialités non négligeables

Plusieurs types d'études portant sur le « Big Data » sont disponibles et permettent de percevoir les vastes domaines d'application qui sont proposés par cet afflux inédit et inégalé de données.

3.2.2.1 Des outils plus performants

Certaines études telle celle de Jeannot, parue en 2008, propose d'utiliser la base du moteur de recherche de Google® et de la personnaliser afin de proposer l'indexation des sites médicaux reconnus de qualité (40). Ce sont les sites médicaux adhérant entre autre à la charte « Health on the Net » ou PubMed Central par exemple. Le but est de répondre aux besoins du praticien et du patient par rapport aux informations fiables au sujet de la santé.

3.2.2.2 Des connaissances accrues sur les comportements des populations

La plupart des études concernant le « Big Data » sont des études descriptives. Elles essaient de décrire la fréquence et la répartition de phénomènes de santé ou de déterminants de santé dans les populations, en fonction de caractéristiques humaines, spatiales, temporelles (41).

A partir de la grande quantité d'informations que fournit le « Big Data », ces études vont explorer différents versants qui peuvent rentrer en jeu dans des pathologies et voir si on peut trouver une corrélation.

En 2013, Frijters et al. ont réalisé une étude dont le but est d'explorer le lien entre les conditions macro-économiques et l'alcoolisme à l'aide des recherches Google® (42).

Des termes tels que « réunion des alcooliques anonymes », « effets de l'alcool », « abus d'alcool » et « cure contre l'alcoolisme » sont analysés temporellement et géographiquement avec Google Trends® (nommé Google® Insight for Search aux Etats-Unis).

Les volumes de recherche obtenus sont mis en lien avec les statistiques du chômage obtenues auprès des administrations des états du pays.

D'après cette étude il semblerait bien qu'il y ait une augmentation des recherches sur internet au sujet des risques et des traitements possibles de l'alcoolisme, corrélée à une période de récession économique.

Les auteurs soulignent néanmoins que la robustesse de l'étude n'est pas parfaite. Leur méthode ne permet pas de savoir si ce sont les personnes directement concernées par un problème avec l'alcool qui cherchent ou plutôt un proche. Ils n'ont pas non plus accès à des données sur le nombre de personnes fréquentant les Alcooliques Anonymes ou les accidents de la route causés par l'alcool.

Ils indiquent qu'une étude longitudinale suivant individuellement des individus voire leur famille, en connaissant leur consommation d'alcool, leurs recherches internet, leurs conditions socio-économiques et les conduites à risque adoptées, permettraient d'affirmer avec plus de poids le lien entre les conditions macro-économiques et l'alcoolisme au travers des recherches Google®.

On voit donc ici les limites du Big Data qui malgré la gigantesque quantité de données, peut aider à pointer du doigt les liens entre un problème de santé et un facteur de risque mais qui ne peut s'affranchir totalement des études traditionnelles pour être valide. L'avantage réside dans la plus grande facilité à trouver des informations sur une maladie, l'alcoolisme, qui est socialement stigmatisée et qui est plus difficilement assumée et donc rend plus difficile la mise en place d'études.

Aussi en 2013, Song et al. conduisent une étude avec l'aide de Google Trends® sur l'impact des facteurs sociaux et psychologiques concernant le suicide dans les recherches internet en Corée (43).

La Corée a un taux de suicide de 33,5 pour 100 000 contre 12,8 en moyenne pour les autres pays de l'OCDE (43).

L'étude a consisté à analyser le volume de recherche par mois entre 2004 et 2010 pour les termes « stress », « suicide », « exercice » et « alcoolisation ». Les volumes sont obtenus avec Google Trends®. Ces derniers sont corrélés avec le nombre de suicides sur cette période.

L'analyse a utilisée des modèles mathématiques multidimensionnels.

Les résultats ont montré que depuis 2007 le taux de suicide et les recherches Google® en lien avec le suicide avaient augmenté. Pour les auteurs, les modèles pointent notamment un lien significatif entre les recherches Internet liées au stress et le nombre de suicides.

Les auteurs évoquent un lien entre le stress et le suicide et proposent d'utiliser le « Big Data en temps réel » comme indicateur du risque de suicide. On voit ici le « Big Data » comme un outil pour essayer si ce n'est de prévenir un phénomène, d'avoir une idée de son ampleur.

Ceci est tributaire de la validité des modèles d'analyses proposés par les auteurs et de la robustesse de l'étude.

3.2.2.3 Une aide pour les politiques de santé publique

Devant la grande quantité de données disponibles avec l'ère du Big Data, un des domaines de la santé le plus propice à en profiter est celui de la santé publique. Certaines études proposent ainsi de voir si des campagnes d'informations et de dépistage trouvent un écho dans la population.

En 2014, Dehkordy et al.(44) réalisent une étude en s'aidant de Google Trends® pour voir si la mise en place dans divers états américains d'une politique d'information des patientes subissant une mammographie entraîne une variation du nombre de recherches sur Internet concernant une « opacité détectée sur la mammographie ».

Dans leur étude, le terme « dense breast » qui peut être traduit par « opacité mammaire » est cherché dans Google Trends®. La popularité de cette recherche dans le temps et dans l'espace est obtenue.

Les auteurs regardent ensuite l'évolution de la popularité du terme au moment où les politiques d'information sont mises en place dans chacun des états.

L'étude montre des pics de recherche dans les états au moment où la législation évolue.

Pour les auteurs, il apparaît que Google Trends® peut aider à évaluer la pertinence des moyens mis en place dans le domaine de la santé publique.

En 2016, Ayers et al. publient une étude où ils s'interrogent sur l'intérêt que peut avoir le Big Data dans l'amélioration des campagnes de sensibilisation. Ici, ils s'intéressent à la journée nationale d'arrêt du tabac « Great American Smokeout » ou « GASCO » qui se tient le troisième jeudi de novembre, chaque année (45).

Ils pointent la difficulté d'évaluer l'impact de ces campagnes de sensibilisation.

Leur travail a pour objectif de proposer un nouveau système d'évaluation des campagnes de prévention. La journée nationale anti-tabac américaine permet de réaliser une étude de cas où est observée les articles de presse, les Tweets, les recherches Google® et Wikipédia en rapport avec le sevrage tabagique ainsi que les appels aux plateformes téléphonique de soutien à l'arrêt du tabac.

Pendant la période 2009-2014, le modèle proposé par les auteurs analyse et isole les pics de recherche des termes en rapport avec le sevrage tabagique sur les divers supports surveillés. Il permet de comparer les résultats obtenus le jour du « GASCO » avec les tendances de recherche, supposées grâce à un algorithme, des autres jours.

Les auteurs notent une augmentation de 61% des actualités et de 13% des Tweets en lien avec l'arrêt du tabac par rapport à une journée sans le « GASCO ». De même, ils obtiennent pour les recherches Google[®] et Wikipédia, une augmentation des recherches en rapport avec le sevrage tabagique de 25% et 22%. Enfin, les plateformes d'appel téléphonique ont 42% d'appels en plus.

Ainsi, il semblerait que cette journée de sensibilisation à l'arrêt du tabac porte ces fruits avec une augmentation sensible des recherches, appels ou actualités en lien avec le sevrage tabagique.

Les auteurs évoquent donc la possibilité d'utiliser le Big Data pour évaluer rapidement et à moindre coût les effets des campagnes de sensibilisation.

3.2.2.4 Prédire grâce au Big Data

Pour certains, la quantité de données disponibles devraient permettre, après analyse des informations, d'extrapoler les résultats obtenus pour prévoir des crises sanitaires comme des épisodes d'épidémie par exemple (36,46).

On pourrait imaginer qu'en analysant le nombre de recherche sur Internet de certains symptômes plus ou moins spécifiques à une maladie cela serait le signe d'une augmentation de l'incidence de ladite pathologie. Il faudrait bien évidemment, se munir de modèles d'analyse complexes, robustes et fiables.

La grippe est un bon exemple de maladie infectieuse pour laquelle les études essaient de voir si le Big Data peut être utilisé comme un outil prédictif afin d'anticiper une épidémie et préparer une lutte efficace.

En 2011, Malik et al. ont publié une étude où ils affirmaient que l'utilisation de « Google[®]Flu Trends » et des données du service de régulation des urgences permettaient de prévoir la pandémie de la grippe H1N1 en 2009, à Manitoba (46).

Google[®] Flu Trends a été lancé en 2008. Cet outil comptabilise les requêtes effectuées sur Google concernant la grippe (« symptômes grippe » ou « traitement grippe » par exemple). En les analysant à l'aide de modèles mathématiques et d'algorithmes, le but est de simuler le nombre de cas de grippe probable et ainsi de prévoir des pics et donc des épidémies (47).

Dans la publication de Malik et al. (46), il est comptabilisé le nombre de recherches en ligne liées à un syndrome grippal grâce à Google[®] Flu Trends, le nombre de visites au service d'urgence pour des sensations d'état grippal et enfin, le nombre de visites pour un état un état grippal avéré. Des modèles ont été réalisés pour ces trois indicateurs. Ils ont été comparés et corrélés avec le nombre de cas avérés de grippe H1N1, fournis par les laboratoires de Manitoba, via la présence du virus H1N1 A dans les échantillons des patients soumis à une PCR.

Les auteurs signalent qu'en comparant chacun des indicateurs, ceux issus de Google[®] Flu Trends et du département des urgences, anticipent avec 1 à 2 semaines d'avance l'épidémie de grippe H1N1, par rapport à celui issu des analyses en laboratoire.

Cette étude évoque donc l'idée d'une utilisation des données massives issues de Google[®] Flu Trends comme une possibilité dans l'anticipation des crises sanitaires.

Une autre étude d'Araz et al. publié en 2014 propose d'utiliser Google[®] Flu Trends pour prévoir le nombre de visites aux urgences concernant un syndrome grippal à Omaha, dans le Nebraska (36). Les auteurs concluent que Google[®] Flu Trends leur a permis d'améliorer leur capacité à prévoir les arrivées aux urgences pour un syndrome grippal.

Pour eux, ceci pourrait permettre aux hôpitaux d'être plus efficient dans la gestion de leurs ressources et d'optimiser l'utilisation du matériel et réduire les coûts.

Pour essayer d'évaluer les besoins spécifiques des diverses régions du monde en soins bucco-dentaires, Harorli et al. ont mené en 2014 une étude sur les requêtes les plus populaires concernant les problèmes bucco-dentaires entre 2004 et 2014 (48).

L'intérêt de ce type d'étude serait d'arriver à cerner le plus précisément possible les problèmes de santé bucco-dentaires selon les régions du monde. Ainsi, il serait possible de prioriser les différents types de politiques de santé publique : orientation plutôt vers la santé parodontale ou pour la lutte contre la carie ou les cancers de la cavité orale.

A première vue, l'utilisation du Big Data semble prometteuse pour anticiper certaines crises sanitaires. On pense évidemment aux substantielles économies que pourraient permettre un tel système en permettant de gérer au plus juste certaines dépenses de santé.

3.2.3. *Intérêts de Google® Trends en épidémiologie bucco-dentaire*

Google® Trends est un outil précieux pour analyser la masse de données que représente les recherches tapées dans le moteur de recherche Google®.

On pourrait imaginer une surveillance « en temps réel » des requêtes concernant la santé bucco-dentaire et voir s'il existe des moments de l'année où les gens se sentent plus concernées.

L'évaluation des programmes de prévention et d'information en santé bucco-dentaire pourraient bénéficier de cette évaluation rapide de leurs effets.

C'est par exemple le travail de Murray et al. qui ont observé l'influence de la mise en place de journées de sensibilisation aux cancers de la cavité buccale en Irlande chaque année au mois de septembre (49). Un pic de recherche est observé durant le mois de septembre depuis la mise en place de la journée.

Comme l'explique Canceill dans sa thèse portant sur l'apport de Google® Trends aux études épidémiologiques, cet outil permet de comparer jusqu'à cinq termes de recherche en simultanés, et d'analyser également les zones de recherche, afin de comprendre l'état d'esprit de la population à un temps donné.

On possède donc une base intéressante pour la réalisation d'études épidémiologiques observationnelles (50).

3.2.3.1 Utiliser Google® Trends à bon escient : suivre une check-list

Nuti et al., dans leur revue de la littérature sur les études utilisant Google® Trends, proposent également une « checklist » pour les auteurs souhaitant en réaliser (51).

Il est souvent reproché aux études portant sur le Big Data, le manque de transparence sur leur protocole. Cette marche à suivre est une bonne proposition pour essayer de clarifier la manière dont les études utilisant Google® Trends ont été menées

Table 4. Checklist for Documentation of Google Trends.

Section/Topic	#	Checklist Item	Reported on Page #
Search Variables			
Access Date	1	Provide the date(s) when Google Trends was accessed and when the data was downloaded.	
Time Period	2	Identify all the time periods that were searched for in Google Trends, providing up to the Month and Day in detail.	
Query Category	3	Identify which query category was used for search; if not using a query category, designate that "all query categories were used", which is the default setting.	
Search Input			
Full Search Input	4	Provide the full search input(s) that were queried for in Google Trends, along with the appropriate documentation of search syntax (detailed in 4a and 4b). Ensure that the provision of the search input is clear, using brackets (as in the example below) or other delineators to separate the search input from the body text.	
Combination	4a	If more than one search term was used, document whether those terms were used in combination with a plus sign (+), or if terms were excluded with a minus sign (-). If terms were not used in combination, state so clearly.	
Quotation Marks	4b	If there was more than one word in any search term (ex. "lipid guideline"), document whether those words were queried with quotation marks or not.	
Rationale for Search Strategy			
For Search Input	5	Provide the reasoning behind the choice of search input.	
For Settings Chosen	6	Provide the reasoning for the settings/search variables chosen to specify the search.	
Hypothetical Example			
<p>On May 1, 2014, we queried Google Trends and downloaded the data for the following search input: ["cholesterol guideline" + "lipid guideline" + "cholesterol recommendation" + "statin recommendation"]. We searched within the United States from January 1, 2013 to May 1, 2014 using the "health" query category. We chose these search terms based on a survey of cardiovascular disease patients' most likely search terms for this topic. We chose January 1, 2013 as the start date to capture baseline interest in the year before the publication (November 2013), chose United States because it is the country of the guideline publication, and chose the "health" query category because we wanted to assess interest in the context of health.</p>			

Figure 21 : liste des vérifications à faire lors de l'utilisation de Google[®] Trends d'après Nuti et al. (51).

SECTION	DONNEES	N°	CONTENU
Options de recherche	Date d'accès	1	Date de connexion sur Google Trends® et de téléchargement des données
	Période	2	Identifier précisément toutes les périodes (mois, jours...) recherchés dans Google Trends®
	Catégorie	3	Identifier quelle a été la catégorie à laquelle devait se rapporter le terme de recherche. Si aucune catégorie spécifique n'a été choisie, il conviendra de préciser « toutes les catégories de recherche ont été utilisées » (réglage par défaut).
Termes recherchés	Ensemble des termes recherchés	4	Donner l'ensemble des termes recherchés dans Google Trends®, en détaillant – d'après les informations détaillées ci-dessous en 4a et 4b – l'exacte formulation de la requête. S'assurer que la recherche soit correctement identifiable si la procédure est rédigée comme un corps de texte.
	Combinaison des termes	4a	Si plus d'un terme est utilisé, renseigner sur la façon dont ces termes ont été soumis au moteur de recherche (avec un signe « + » pour une association, ou un signe « - » pour une exclusion). Si les termes n'ont pas été soumis ensemble, il convient de le préciser. Il s'agit donc de la formulation de la requête.
	Marques de citations	4b	Si le terme recherché est composé de plus d'un mot, détailler la façon dont la recherche a été effectué (avec des guillemets ou pas...).
Raisons, stratégies	Des termes recherchés	5	Expliquer le raisonnement ayant conduit au choix des termes recherchés.
	Des options de recherche	6	Expliquer le raisonnement ayant conduit au choix des critères de recherche.

Figure 22 : traduction par Canceill (50) de la checklist de Nuti et al. (51).

3.3. Les limites du Big Data

Le Big Data est un outil puissant pour essayer d'analyser et de répondre à des enjeux de société comme l'est le domaine de la santé publique. Il concerne tous les milieux et tous les types de population. Ceci rend ses domaines d'applications excessivement variés (39).

Cependant il est nécessaire de s'interroger si le BigData n'entraîne une incursion dans la vie privée que le grand public rejetterait massivement s'il en avait pleinement conscience.

3.3.1. *Pour qui, pour en faire quoi ?*

Une de ses forces qui peut devenir une faiblesse est que la plupart des données sont facilement accessibles et pas systématiquement par des personnes ayant les clés pour les analyser correctement.

La capacité d'analyser de manière efficiente cette quantité de données n'est pas donnée à n'importe qui. Ceci a pour effet qu'une quantité exponentielle d'informations sur la population peut être en possession d'organismes gouvernementaux ou privés. Ainsi on peut tout à fait imaginer un usage détourné ou frauduleux. Sans un encadrement strict de son utilisation, le Big Data pourrait entraîner une réduction des libertés publiques (38).

De plus, l'accès aux données peut parfois coûter cher. Il y pourrait y avoir besoin de moyens financiers auxquels tous les pôles de recherche ne peuvent prétendre. Cet accès peut même être impossible ou biaisé si ceux qui possèdent les données ne veulent pas céder ou que partiellement leur échantillon (38).

Il est important que la communauté scientifique, voire le législateur si nécessaire, envisagent jusqu'où il est souhaitable d'aller dans l'utilisation de ces données. Il faut bien garder à l'esprit la balance bénéfice-risque (38).

3.3.2. *Trop d'informations tuent l'information*

On a un changement dans la manière d'approcher la connaissance car on a accès très rapidement à une grande quantité de données mais sans tri. De plus, certains outils ou base de

données n'ont pas été forcément créées pour la recherche et peuvent rendre l'exploitation des résultats compliquée (39).

Ainsi, l'exemple de la surévaluation de la vague de grippe par Google[®] Flu Trends en 2013 est un exemple des limites atteintes par le Big Data. Il est évident qu'il est difficile de trouver des informations significatives et valides avec tout ce « bruit » autour. Dans ce cas, la sur médiatisation de la vague de grippe cette année-là a fait augmenter de manière significative la présence du terme « grippe » sur Internet (le « bruit »), et a faussé les calculs de l'algorithme de Google[®] Flu Trends (52).

Les standards de recherche en épidémiologie nécessitent des échantillons solides et des protocoles très précis, avec le moins de biais possibles (38).

Devant le grand nombre de données, on pourrait vouloir faire des liens qui n'existent pas entre deux variables mais qui se tissent, de manière subjective dans l'esprit du chercheur, car le volume de données le permet.

Par exemple, dans une étude, en utilisant les informations d'une base de données, Leinweber avait démontré qu'on trouvait de manière complètement fortuite, une forte corrélation entre un indice de la bourse de New-York (S&P 50) et la production de beurre au Bangladesh (53).

L'enjeu du Big Data, comme l'évoque Boyd et Crawford dans leur article « Critical questions for Big Data », est que malgré la grande quantité de données, il faut pouvoir être en mesure d'obtenir un échantillon valide, sans biais et en justifiant l'origine des données (39).

De même, Nuti et al. notent que Google Trends[®] est régulièrement utilisé pour étudier des phénomènes médicaux de manière très diverse. Les résultats obtenus peuvent être très intéressants mais la faible documentation sur les méthodes utilisées empêchent la reproductibilité de ces derniers. Si les protocoles de recherche utilisés avec Google Trends[®] étaient renseignés de manière plus transparente cela appuierait la cohérence et la fiabilité des études menés avec cet outil (51).

Ainsi, ce n'est pas parce qu'on obtient un plus grand échantillon que ceux d'une étude avec des dizaines voire des centaines ou des milliers de cas qu'il est de meilleure qualité, sans biais ou plus représentatif. La grandeur de l'échantillon n'est pas la garantie de sa validité (39).

En compilant les données obtenues, on les a de manière brute, parfois sorties de leur contexte alors que ce dernier pourrait changer la donne.

Boyd et Crawford expliquent notamment que, dans les études portant sur les réseaux sociaux, des gens pourraient sembler lier l'un à l'autre car ils sont « liés » sur le réseau social, alors qu'ils ont peut être seulement échangés quelques mots. Pourtant, ils se retrouvent liés au même niveau que leurs amis proches sur ce même réseau. Sans le contexte, on ne sait pas la réelle valeur du lien entre les personnes ce qui peut être pénalisant pour une étude en sciences humaines (39).

Le Big Data ouvre un immense champ des possibles pour la recherche, notamment dans le domaine de la santé publique.

Mais comme pour tout outil récent, nous en sommes aux balbutiements et devons faire attention à l'utiliser avec mesure, dans le respect des individus et en lui appliquant toute la rigueur dont nous faisons preuve avec les outils traditionnels.

3.3.2.1 Google® Flu Trends : un échec de l'utilisation du Big Data devant l'abondance d'informations

Google® Flu Trends ne publie plus de résultats depuis le 20 août 2015 (54). Seules les archives sont maintenant accessibles.

Un article paru dans Nature en 2009 de Declan Butler a interrogé divers acteurs de la surveillance des cas de grippe dans divers pays (52). La plupart s'accordent à dire que l'utilisation d'un seul moyen de surveillance que serait celui issu du Big Data ne suffit pas et qu'il faut toujours se servir des modèles « classiques », avec les cas objectivés par prélèvement bactérien.

Le système de Google® dépendant des requêtes des internautes, une simple augmentation de la couverture médiatique de la grippe pourrait exagérer le nombre de requêtes et donc l'estimation des cas.

En 2014, l'étude de Lazer et al. démontre que Google® Flu Trends a eu tendance surestimé de 50% les cas de grippe en 2013 (55).

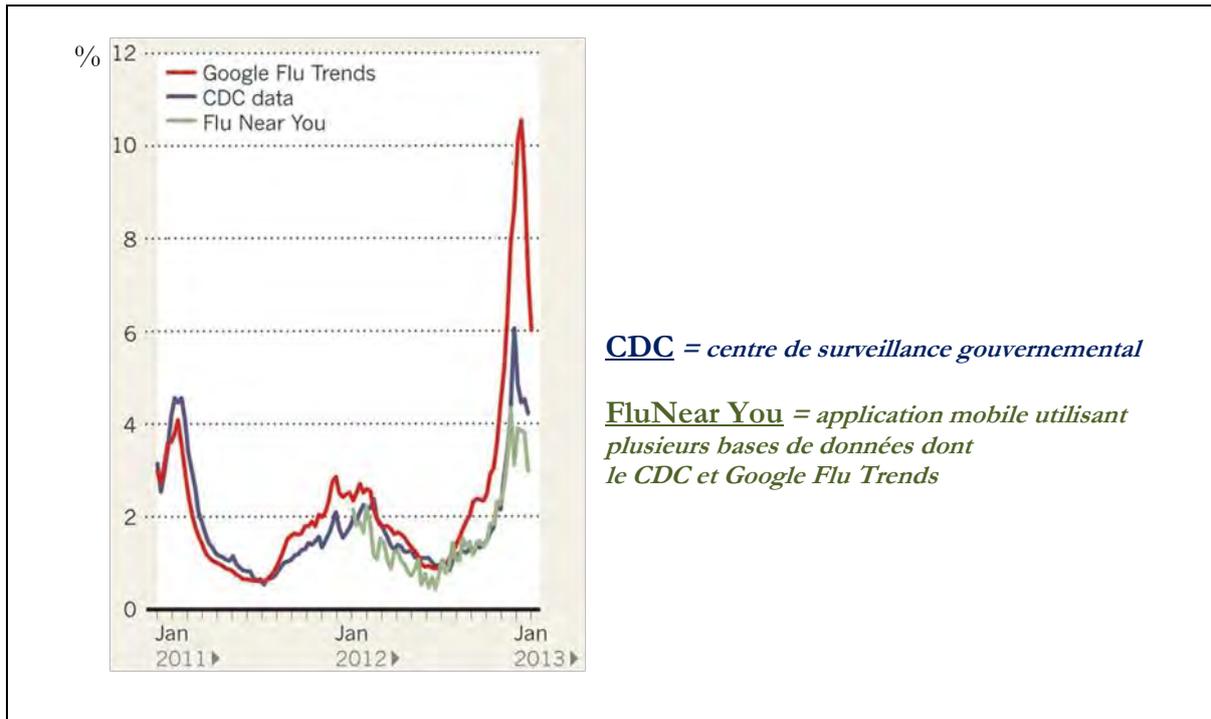


Figure 23 : pourcentage estimé de la population américaine atteinte de symptômes grippaux : différences entre plusieurs systèmes de mesure (52).

Ainsi, même si cet outil permettait d'avoir une estimation rapide et économique d'un risque d'épidémie de grippe, la nécessité d'améliorer très rapidement le modèle et d'avoir des chiffres fiables ne permet pas encore son utilisation à une échelle gouvernementale. L'avenir semble à une utilisation de modèles mêlant à la fois Big Data et modèles classiques comme le suggère l'étude de Preis et Moat (56). Ils démontrent que l'utilisation de ces deux systèmes à la fois permet d'améliorer la surveillance en temps réel de l'épidémie de grippe.

3.3.3. Limites propre à Google[®] Trends

Les travers que l'on peut reprocher à Google[®] Trends sont à la fois ceux qu'on peut reprocher en général à tous les outils d'exploitation du Big Data mais aussi des inconvénients propres à son fonctionnement.

Un des principaux problèmes posé par Google[®] Trends réside dans la manière dont le résultat est donné, ce dernier n'informant qu'à partir d'un volume de recherche relatif et non pas absolu.

Ceci a pour conséquence de ne pas indiquer le nombre exact de fois où le terme concerné a été tapé dans la barre de recherche (volume absolu). La comparaison se fait par rapport à lui-même (volume relatif) (50).

Seule la volonté de transparence de Google[®] concernant la méthode de calcul de ce volume pourrait permettre d'en faire un outil de recherche en épidémiologie plus fiable.

Enfin, comme le démontre Canceill dans sa thèse, une partie des résultats obtenus peut être invalidés par ce qu'on appelle des recherches croisées (50).

Le même terme que celui de l'étude va être tapé par les internautes mais il va désigner une tout autre chose.

Il le démontre via le cas de sa thèse où il a dû étudier la popularité du terme « amalgame », qui désigne dans ce cas, le matériau d'obturation dentaire. Or, la période où il a mené son étude était celle des attentats en France. Le gouvernement et la sphère médiatique évoquaient régulièrement la nécessité de ne pas faire d'« amalgame ». Canceill a donc observé des pics très importants sur la période des attentats et a pris le parti d'affiner sa recherche en recherchant « amalgame dentaire ».

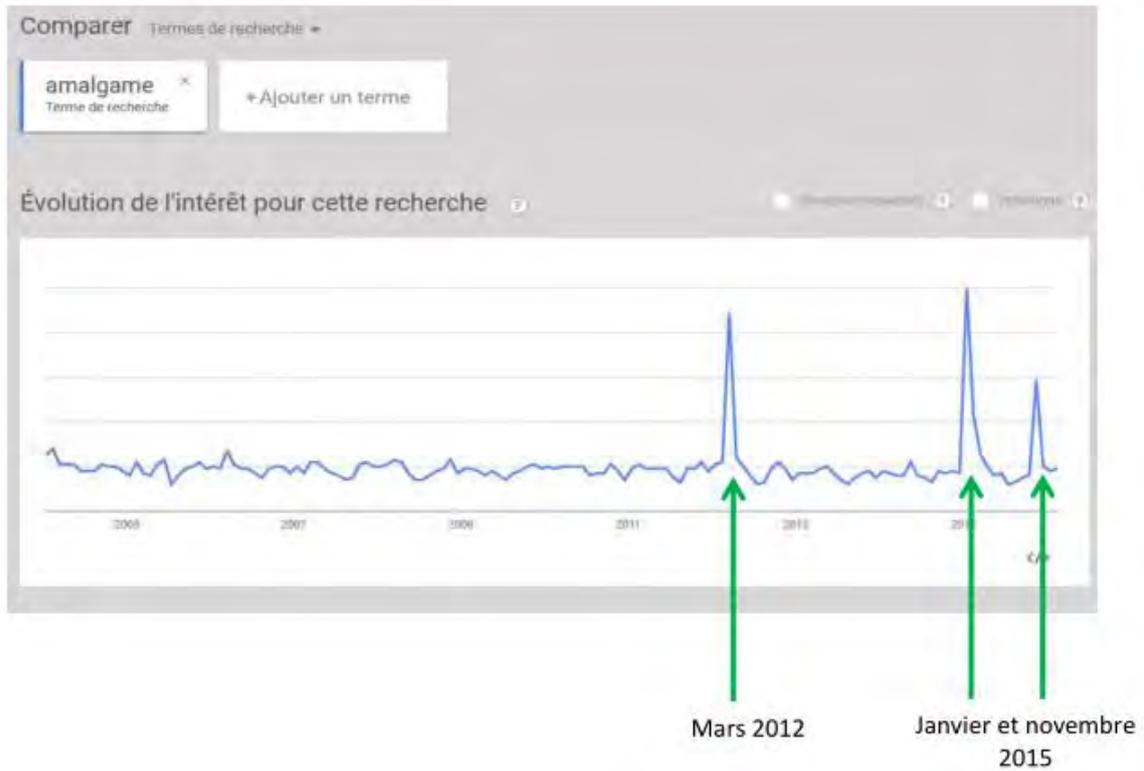


Figure 24 : courbe obtenue par Canceill dans Google® Trends en tapant "amalgame" sans guillemets le 06/02/2016 dans Google® Trends. (50.)

Est-ce que les catégories proposées par Google® permettraient de s'affranchir de ce type de biais ? Cela reste à démontrer. Le manque de transparence sur le fonctionnement de ces catégories empêche d'évaluer leur fiabilité.

3.4. Ethique et Big Data

Les données sont obtenues à partir d'informations que des citoyens vont renseigner. De nos jours, la plupart des données sont informatisées dans des bases de données qui peuvent être rapidement compilées, partagées et stockées.

L'article 27 de la Déclaration Universelle des Droits de l'Homme de 1948 reconnaît comme un droit fondamental de « participer au progrès scientifique et aux bienfaits qui en résultent » mais aussi « la protection des intérêts moraux et matériels découlant de toute production scientifique [...] ».

Comme dans tout progrès, la multiplicité des possibilités expose forcément à des dérives, parfois involontaires. Il est donc important de prévenir les mauvais usages et la mauvaise interprétation des données (57).

Par exemple, on peut imaginer que si des études essaient de lier une pathologie à un type de population, on court le risque d'identifier des communautés comme porteuses d'une pathologie et donc de discriminer une catégorie d'individus alors que tous les individus ne sont pas touchés.

Une technologie en elle-même, n'est jamais bonne ou mauvaise, ce sont les conséquences humaines, sociales et environnementales causées par son utilisation qui le sont potentiellement. Ainsi ce n'est pas parce que des informations sont accessibles qu'il est éthique de les utiliser.

Même si les données semblent publiques, est-on certain que leur propriétaire serait d'accord pour qu'elles soient utilisées ? Et est-on certain de pouvoir garantir leur anonymat total ?

Les chercheurs doivent garder à l'esprit que la plupart des personnes ne se doutent pas que ce qu'ils recherchent ou publient sur internet peut être utilisé. Bien qu'il soit impossible d'obtenir le consentement de toutes les personnes concernées par les études, il faut toujours essayer de garder à l'esprit la dimension éthique de toute démarche (39,57).

4. Estimation de la prévalence carieuse mondiale à l'aide de Google Trends®

4.1. Introduction

La carie est, avec la maladie parodontale, la principale pathologie dentaire en terme de prévalence (10).

En tentant de comprendre comment les populations se comportent face à des pathologies, le but est de proposer des politiques de santé publique les plus personnalisées possibles en fonction du pays où elles s'appliquent.

Comme nous l'avons vu dans la première partie, le temps de mise en place des programmes de prévention de santé bucco-dentaire est long et nécessite une logistique importante pour le recueil des informations propres à chaque pays ainsi que la coordination des actions à l'échelle mondiale.

Malgré tous les biais et les limites que réserve l'utilisation du Big Data dans le domaine de la santé publique, nous avons voulu savoir dans quelle mesure Google®Trends pouvait être un outil intéressant pour estimer la prévalence de la carie au niveau mondial.

Le but de cette étude est d'établir à l'aide des informations obtenues sur Google®Trends, une carte du monde concernant la requête du terme « carie » et la comparer à la carte de l'OMS de l'indice CAOD.

4.2. Matériel et méthode

Pour mener à bien cette étude, nous avons besoin de savoir si la carte que nous réalisons à l'aide des données issues de Google[®] Trends se rapproche de cartes issues de méthodes dites « classiques ».

4.2.1. *La carte de l'OMS*

La carte de référence dans le domaine de la carie dentaire est la carte de l'OMS publié lors de son dernier rapport sur la santé bucco-dentaire, paru en 2003 (10).

Il s'agit d'une carte du monde où est représenté l'indice CAOD moyen des adultes entre 35 et 44 ans et des enfants de 12 ans.

Il existe très peu d'informations disponibles sur les conditions des réalisations de cette carte. Nous savons seulement que les pays évalués sur la carte se sont soumis à des études permettant d'évaluer l'indice CAOD.

Nous ne connaissons malheureusement pas les protocoles exacts de ces études. Il est à noter aussi que cette carte, publiée en 2003, est réalisée à partir de recueils de données parfois bien antérieures (58).

4.2.1.1 L'indice CAOD

L'indice CAOD utilisé pour réaliser la carte a été mis au point par Klein et Palmer en 1940. C'est un indice de sévérité de l'atteinte carieuse. Il dénombre les dents permanentes cariées (C), absentes pour cause de carie (A) et obturées (O) chez un individu. Le score maximum est de 28 lorsque les troisièmes molaires ne sont pas prises en compte (2).

Plus l'indice CAOD est élevé, plus cela signifie que la population est atteinte de carie.

Les cartes de l'OMS nous permettent d'obtenir des listes de la prévalence carieuse par pays pour les enfants de 12 ans et les adultes de 35 à 44 ans (**Tableau 1 et 2**). Ce chiffre est une valeur absolue. Il est à noter que certains archipels d'îles-états trop peu précis sur la carte n'ont pu être identifiés avec certitude donc n'ont pas été mentionné dans cette liste.

Tableau 1 : Liste des pays en fonction de leur indice CAOD chez les adultes de 35 à 44 ans d'après la carte de l'OMS (10).

Indice CAOD	Pays
Très bas <5,0	Bangladesh, Bénin, Burkina Faso, Chine (dont Macao et Hong Kong), Congo, Corée du Nord, Ethiopie, Ghana, Jordanie, Laos, Namibie, Népal, Niger, Nigéria, Pakistan, Rwanda, Swaziland, Tanzanie, Togo, République démocratique du Congo, Zambie, Zimbabwe
Bas 5,0-8,9	Afghanistan, Arabie Saoudite, Bahreïn, Cambodge, Comores, Côte d'Ivoire, Egypte, Emirats Arabes Unis, Gambie, Inde, Indonésie, Irak, Kazakhstan Kenya, Koweït, Lesotho, Malawi, Maldives, Moldavie, République centrafricaine, Sierra Leone, Singapour, Soudan, Taïwan, Thaïlande, Turkménistan, Vietnam, Yémen
Modéré 9,0-13,9	Albanie, Algérie, Argentine, Arménie, Biélorussie, Chypre, Corée du Sud, Cuba, Espagne, Etats-Unis, Fidji, Géorgie, Iles Marshall, Iran, Israël, Italie, Jamaïque, Japon, Madagascar, Malaisie, Malte, Maroc, Mongolie, Ouzbékistan, Panama, Papouasie-Nouvelle-Guinée, Portugal, Roumanie, Russie, Sri Lanka, Syrie, Turquie, Ukraine
Haut >13,9	Allemagne, Australie, Autriche, Belgique, Brésil, Bosnie-Herzégovine, Bulgarie, Canada, Chili, Colombie, Croatie, Danemark, El Salvador, Estonie, Finlande, France, Grèce, Groenland, Hongrie, Irlande, Islande, Lettonie, Liban, Macédoine, Norvège, Nouvelle-Zélande, Pays-Bas, Pérou, Philippines, Pologne, République tchèque, Royaume-Uni, Serbie, Slovaquie, Slovénie, Suède, Suisse

Tableau 2 : Liste des pays en fonction de leur indice CAOD chez les enfants de 12 ans d'après la carte de l'OMS (10)

Indice CAOD	Pays
Très bas <5,0	Anguilla, Australie, Bangladesh, Belize, Bénin, Botswana, Burkina-Faso, Burundi, Chine (Chine (dont Macao et Hong Kong), Congo, Danemark, Djibouti, Ethiopie, Finlande, Ghana, Groenland, Guinée-Bissau, Haïti, Iles Marshall, Jamaïque, Kenya, Lesotho, Libéria, Lybie, Malawi, Mozambique, Népal, Nigéria, Ouganda, Pakistan, Pays-Bas, République démocratique du Congo, Royaume-Uni, Rwanda, Singapour, Somalie, Suède, Suisse, Swaziland, Taïwan, Tanzanie, Togo, Tunisie
Bas 5,0-8,9	Afrique du Sud, Albanie, Allemagne, Algérie, Angola, Arabie Saoudite, Arménie, Autriche, Bahreïn, Belgique, Birmanie, Bhoutan, Cambodge, Cameroun, Canada,, Chypre, Colombie, Comores, Costa Rica, Côte d'Ivoire, Dominique, Egypte, El Salvador, Emirats Arabes Unis, Espagne, Etats-Unis, France
Bas 5,0-8,9	Gambie, Géorgie, Grèce, Guatemala, Guyana, Inde, Indonésie, Irak, Iran, Irlande, Islande, Israël, Italie, Japon, Kazakhstan, Koweït, Laos, Malaisie, Maldives, Mali, Malte, Maroc, Mauritanie, Moldavie, Mongolie, Namibie, Niger, Norvège, Nouvelle-Zélande, Oman, Ouzbékistan, Papouasie-Nouvelle-Guinée, Portugal, Sénégal, Sierra Leone, Slovénie, Soudan, Sri Lanka, Vanuatu, Syrie, Tadjikistan, Thaïlande, Tonga, Turkménistan, Uruguay, Venezuela, Vietnam, Zambie, Zimbabwe
Modéré 9,0-13,9	Afghanistan, Argentine, Barbade, Biélorussie, Brésil, Bulgarie, Chili, Corée du Nord, Corée du Sud, Croatie, Cuba, Equateur, Estonie, Fidji, Gabon, Grenade, Honduras, Hongrie, Jordanie, Kirghizistan, Lettonie, Lituanie, Luxembourg, Macédoine, Madagascar, Mexique, Nicaragua, Panama, Paraguay, Pérou, Pologne, République Centrafricaine, République Dominicaine, République Tchèque, Roumanie, Russie, Sainte-Lucie, Saint-Vincent et les Grenadines, Serbie, Slovaquie, Suriname, Turquie, Ukraine, Vanuatu, Yémen
Haut >13,9	Bolivie, Bosnie-Herzégovine, Brunei-Darussalam, Liban, Maurice, Philippines, Trinité-et-Tobago

4.2.2. Méthode

Notre étude a pour but d'établir la popularité de la recherche du terme « carie » à travers le monde et d'établir une carte à partir des données obtenues afin de la comparer à celle de l'OMS.

4.2.2.1 Premier préalable : traduction du terme « carie »

Une première étape est d'utiliser le maximum de langues possibles pour avoir la carte mondiale la plus exhaustive possible. Il nous faut avoir le plus de traductions possibles du terme « carie ».

Pour cela nous allons utiliser un autre outil de Google®, Google® Traduction. Il est disponible à l'adresse : <https://translate.google.fr/>.

Il s'agit d'un logiciel de Google® qui permet de traduire des termes d'une langue à l'autre.

Actuellement, Google® Traduction propose une traduction dans 103 langues différentes. Au moment de notre étude, toutes les langues de la **figure 25** sont disponibles sauf l'amharique, le corse, le frison, le gaélique écossais, l'hawaïen, le kirghize, le kurde, le luxembourgeois, le pachto, le samoan, le shona, le sindhi et le xhosa, rajouté le 19 février 2016 (59).

Détecter la langue	Cebuano	Galicien	Japonais	Maltais	Serbe	Turc
Afrikaans	Chichewa	Gallois	Javanais	Maori	Sesotho	Ukrainien
Albanais	Chinois	Géorgien	Kannada	Marathi	Shona	Urdu
Allemand	Cingalais	Grec	Kazakh	Mongol	Sindhî	Vietnamien
Amharique	Coréen	Gujarati	Khmer	Néerlandais	Slovaque	Xhosa
Anglais	Corse	Haoussa	Kirghiz	Népalais	Slovène	Yiddish
Arabe	Créole haïtien	Hawaïen	Kurde	Norvégien	Somali	Yorouba
Arménien	Croate	Hébreu	Laotien	Ouzbek	Soundanais	Zoulou
Azéri	Danois	Hindi	Latin	Pachtô	Suédois	
Basque	Espagnol	Hmong	Letton	Panjabi	Swahili	
Bengali	Espéranto	Hongrois	Lituanien	Persan	Tadjik	
Biélorusse	Estonien	Igbo	Luxembourgeois	Polonais	Tagalog	
Birman	Finnois	Indonésien	Macédonien	Portugais	Tamoul	
Bosniaque	Français	Irlandais	Malaisien	Roumain	Tchèque	
Bulgare	Frison	Islandais	Malayalam	Russe	Telugu	
Catalan	Gaélique (Écosse)	Italien	Malgache	Samoan	Thaï	

Figure 25 : listes des langues disponibles dans Google® Traduction (60).

Concernant l'utilisation de Google® Traduction comme le montre la **figure 26**, dans le tableau de gauche on tape le terme à traduire en sélectionnant sa langue d'origine (case la plus grisée), ici le français. Dans le tableau de droite la traduction s'affiche automatiquement dans la langue sélectionnée (case la plus grisée).

Sous les tableaux indiquant les traductions, nous trouvons des propositions de définitions avec des exemples d'utilisation du mot et des synonymes possibles de ce terme dans la langue traduite. Ainsi, on peut vérifier qu'on obtient la traduction du terme dans son bon contexte et de choisir celle qui convient le mieux à la situation.

D'autres options telles que la détection automatique de la langue, l'écoute possible de la prononciation du mot ou l'utilisation d'un clavier différent voire de l'écriture manuscrite ne nous ont pas été utiles.

The screenshot shows the Google Translate interface. At the top, there are language selection buttons for Français, Anglais, and Allemand, and a 'Détecter la langue' button. The source language is set to Français and the target to Espagnol. The input text is 'partir' and the output is 'go'. Below the input field, there are icons for voice input and keyboard input. Below the output field, there are icons for star, list, voice, and share, along with a 'Suggérer une modification' button.

Définitions de partir

verbe

Quitter un lieu.
"Partiras-tu à la campagne cet été? Elle est partie en voyage ."

Prendre le départ, se mettre en marche.
"« Rien ne sert de courir : il faut partir à point »"

Se déclencher, en parlant d'une arme à feu; jaillir, en parlant d'un projectile.
"Faire partir un pétard, une fusée dans un feu d'artifice ."

2 autres définitions

Voir aussi

à partir de, partir en vacances, partir en voyage, à partir du moment où, faire partir, laisser partir, partir en courant, partir pour, prêt à partir, partir de rien

Traductions de partir

verbe

■ leave	laisser, quitter, partir, abandonner, déposer, oublier
■ go out	sortir, partir, se retirer, terminer, émigrer, être distribué
■ go	aller, partir, passer, aller à, marcher, devenir
■ go away	partir, échouer, s'en aller
■ start	commencer, démarrer, partir, débiter, entamer, entreprendre
■ depart	partir, quitter, mourir, s'écarter de
■ take off	décoller, enlever, retirer, ôter, partir, décrocher
■ get away	partir, quitter, filer, prendre la fuite, démarrer, prendre son
■ go off	partir, exploser, sonner, éclater, ravager, perdre le goût de
■ head off	partir, forcer à reculer
■ retire	se retirer, prendre sa retraite, abandonner, quitter, reculer,
■ move away	partir
■ march off	partir
■ run along	courir, partir
■ pop off	claquer, partir, quitter, casser sa pipe, mourir subitement,
■ make tracks	partir

Voir aussi

carié, carier

Figure 26 : exemple d'utilisation de Google[®] Traduction avec le terme « partir » (60).

Dans notre étude, nous choisissons de traduire le terme « carie ». Grâce aux définitions, nous observons qu'il s'agit bien de la carie dentaire. Nous procédons donc à la traduction et obtenons 59 termes différents pour la traduction de caries. La liste complète de la traduction de « carie » est disponible en **annexe n°2**.

Un même terme peut désigner la carie dans des langues différentes. Par exemple, « karies » est la traduction de carie en albanais mais aussi en allemand, finnois, indonésien, malaisien, norvégien et slovène).

Au contraire, un terme peut désigner « carie » pour une seule langue tel que « ฟันผุ » qui en est la traduction de « carie » pour la langue Thaï.

Enfin, nous avons décidé de rajouter dans notre liste les termes « dental caries », « dental cavity » et « dental cavities » car ce sont des appellations courantes pour désigner la carie en anglais.

4.2.2.2 Deuxième préalable : choix des options de recherche : période et catégorie

Nous avons accédé à Google[®] Trends entre mars et octobre 2015.

De même, nous avons vu que Google[®] Trends permettait de centrer la recherche sur des domaines prédéfinies appelés catégories.

Cependant il n'existe aucune information sur la manière dont ils sont définis. Devant cette opacité nous choisissons de ne pas utiliser les catégories et de considérer tous les domaines.

4.2.2.3 Troisième préalable : choix de la combinaison des termes et des marques de citations

Pour l'analyse des termes, quand nous en mettons au moins deux, nous les associons avec le « + » sauf pour « dental cavity », « dental cavities » et « dental caries ». C'est la méthode préconisée par Google[®] Trends pour l'analyse de termes de langues différentes.

Nous utilisons les guillemets pour la constitution d'un premier échantillon (« Echantillon A ») et la vérification de l'absence de « niche linguistique » afin de bien centrer la recherche sur le terme « carie ». Lorsque nous réalisons le croisement avec le pool de référence nous voulons être le plus exhaustif dans les recherches des internautes donc nous ôtons les guillemets.

4.2.2.4 Première étape : recherche du terme dans Google® Trends

Une fois que nous avons obtenus la traduction du terme « carie » dans toutes les langues disponibles sur Google® Traduction, nous les recherchons chacun leur tour dans Google® Trends.

Nous allons obtenir plusieurs cas de figures.

- Pour certains termes, Google® Trends ne sera pas en mesure de nous donner d'informations sur sa popularité dû aux faibles volumes de recherches effectuées.

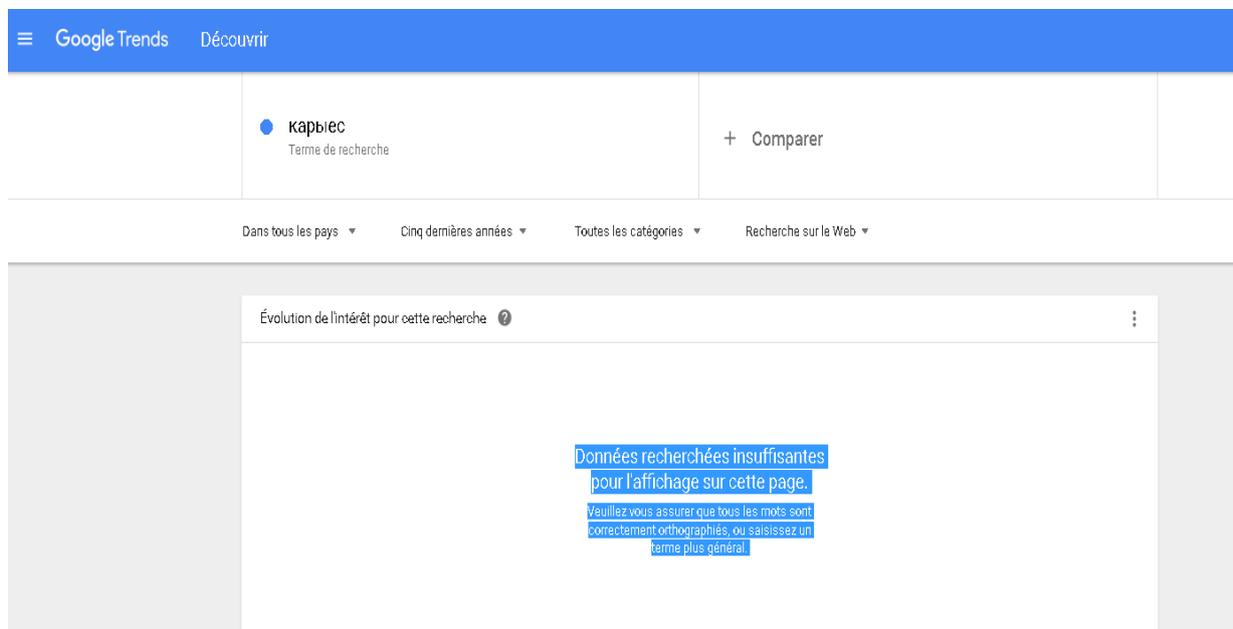


Figure 27 : « карыес » (biélorusse) ne renvoie pas de résultats dans Google® Trends (61).

Nous choisissons de ne pas retenir ces termes pour effectuer nos recherches car nous savons d'emblée que Google® ne pourra pas nous donner d'informations concernant leur popularité à leur sujet et ainsi nous permettre de compléter la carte du monde.

➤ Un autre cas de figure s'est présenté, celui de l'obtention effective d'informations via Google[®] Trends, mais avec une discordance entre les lieux où le terme est populaire et la langue du terme.

Par exemple le terme « cari » est la traduction de « carie » en galicien, une langue du nord-ouest de l'Espagne. Hors quand il est tapé dans Google[®] Trends, on s'aperçoit que sur la carte les zones de popularité sont principalement en Malaisie.

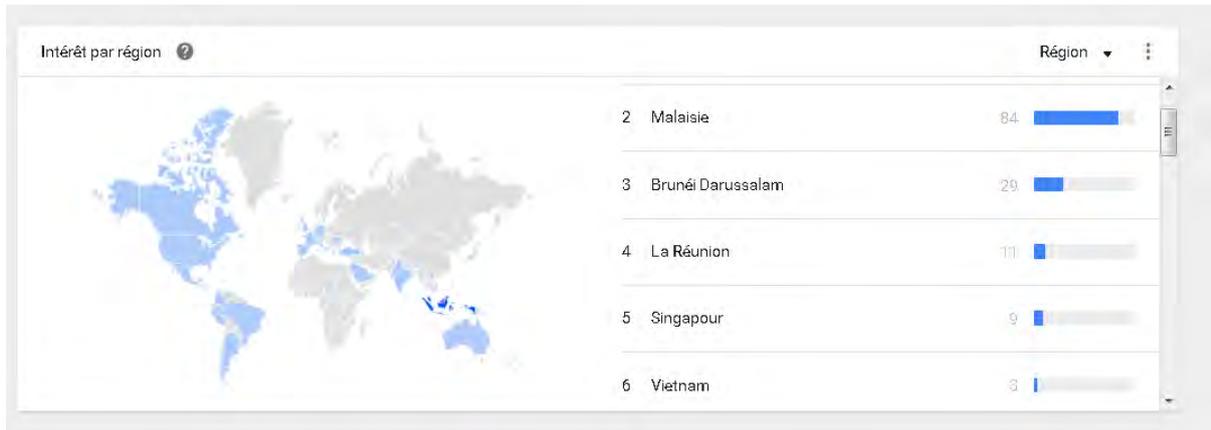


Figure 28 : régions où la recherche de « cari » est populaire : incohérence entre la région de langue et les lieux de popularité (62).

Cette différence entre la langue et le lieu de popularité interpelle. Nous notons chaque discordance dans le but de les expliquer. Ainsi :

- Kari : traduction de « carie » en créole haïtien mais qui se retrouvent majoritairement populaire en Finlande, Norvège, Malaisie, Albanie et Kosovo
- Kario : traduction de « carie » en esperanto mais populaire en Lituanie, Equateur et Chili
- Cari : traduction de « carie » en galicien et populaire notamment en Malaisie et Brunei Darussalam
- Suus : traduction de « carie » en somali mais populaire surtout en Estonie, Pays-Bas et Pologne

➤ Enfin le dernier cas de figure est celui où le terme recherché permet à Google® Trends d'obtenir assez de données pour établir les cartes et courbes ainsi qu'une cohérence entre la région où la langue est parlée et les pays où la langue est populaire.

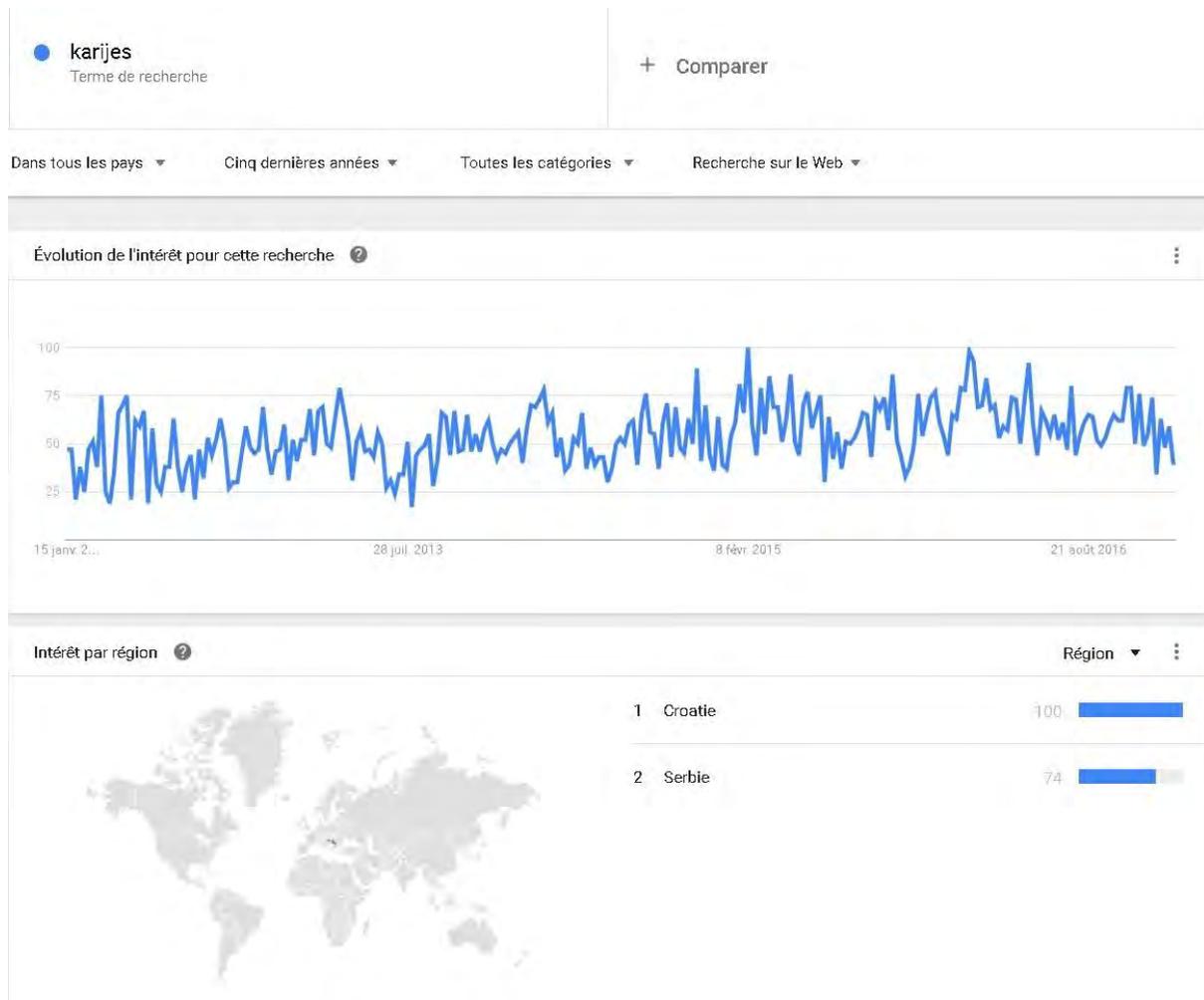


Figure 29 : « karijes » renvoie suffisamment d'informations et sans incohérence (63).

Une fois toutes les traductions passées au crible de Google® Trends, nous pouvons éliminer ceux qui ne renvoient rien. Nous en éliminons 36 au total. Les termes ôtés au cours de cette étape sont listés dans **l'annexe n°3**.

Nous obtenons ainsi **l'échantillon A composé de 22 termes.**

4.2.2.5 Deuxième étape : création d'un« pool de référence »

Notre objectif final est de créer la carte du monde la plus exhaustive possible.

Dans un premier temps nous avons traduit le terme dans le maximum de langues disponibles pour toucher le plus de pays. Notre premier tri a permis de garder seulement les termes renvoyant des données exploitables.

L'étape de création d'un pool de référence est nécessaire car, lors de la réalisation de l'étude en mars 2015, nous avions un nombre de caractères limités dans la barre de recherche.

Ainsi nous avons dû tenir compte de la répartition des diverses langues à travers le monde.

En effet, certaines langues sont parlées par de nombreuses personnes à travers la planète telles que le chinois, le français, le russe, l'anglais, l'espagnol, le portugais, l'arabe ou l'hindi. Malgré tout, cela ne signifie pas forcément que la langue la plus parlée en termes de locuteurs soit la plus influente.

Le linguiste Georges Weber a défini un système dans les années 1990 un système pour évaluer l'influence des langues en évaluant plusieurs domaines tels que le pouvoir économique du pays utilisant cette langue ou le prestige socio-culturel découlant du langage (64).

Ainsi, on obtient la liste suivante des langues les plus influentes :

- | | |
|-------------|--------------|
| 1. Anglais | 6. Chinois |
| 2. Français | 7. Allemand |
| 3. Espagnol | 8. Japonais |
| 4. Russe | 9. Portugais |
| 5. Arabe | 10. Hindi |

Dans notre cas, la création d'un pool de référence permet de conserver avec certitude les langues les plus influentes et qui se retrouvent sur plusieurs continents.

Nous sélectionnons donc le français, l'anglais, l'espagnol, l'arabe et le portugais. L'allemand, le japonais, le chinois, le russe et l'hindi, malgré leur influence notoire, restent davantage confinées géographiquement comme l'indique la **figure 30** (65).



Figure 30 : répartition des langues dans le monde (64)

Ainsi pour le pool de référence initial, nous allons sélectionner les termes désignant la carie dentaire dans les langues choisies ce qui donne :

carie+ caries + dental cavity + cárie + نخر الأسنان

Pool de référence initial

4.2.2.6 Troisième étape : identifier les « niches linguistiques » ou la guerre aux recherches croisées

Les traductions de « carie » peuvent désigner autre chose que la carie dentaire. Il peut s'agir d'un homonyme avec le nom de famille d'une personnalité ou la désignation de toute autre chose que la carie dentaire dans une autre langue que celle de la traduction. On se retrouve dans ce que Canceill appelle des recherches croisées (50) où on ne sait pas ce que l'internaute a réellement cherché : la carie dentaire ou autre chose s'écrivant de manière identique. C'est ce qu'on appelle une « niche linguistique ». Nous devons donc procéder à l'élimination de ces niches linguistiques et homonymes qui fausseraient nos résultats.

Il peut aussi arriver que le terme soit particulièrement populaire alors que la langue recherchée ne fait pas partie du pool de référence initial. Il faut donc associer ce terme au pool de référence initial afin de le conserver dans la réalisation de la carte.

Elimination via le fichier .csv fourni par Google® Trends

Lors de la 1^{ère} étape, nous avons exporté pour chacun des 22 termes le fichier .csv des résultats comme le permet Google® Trends.

Ce fichier contient entre autre la popularité de la recherche (sur 100) pour chaque semaine durant la période définie comme le montre la **figure 31**. Ce sont ces données chiffrées qui permettent de matérialiser la courbe de l'évolution de l'intérêt d'une recherche dans le temps.

	A	B	C	D
1	Recherche sur Recherche sur le Web : caries,			
2	Dans tous les pays, De 2004 à ce jour			
3	,			
4	Intérêt dans le temps,			
5	Semaine, caries			
6	2004-01-04 - 2004-01-10, 55			
7	2004-01-11 - 2004-01-17, 51			
8	2004-01-18 - 2004-01-24, 53			
9	2004-01-25 - 2004-01-31, 60			
10	2004-02-01 - 2004-02-07, 71			
11	2004-02-08 - 2004-02-14, 48			

Figure 31 : popularité de la recherche (sur 100) du terme « caries » pour chaque semaine grâce au fichier .csv fourni par Google Trends

Pour repérer les niches linguistiques nous avons besoin d'un autre type de données fournies par ce fichier .csv, la popularité de la recherche (sur 100) en fonction des pays. Ce sont ces informations (**figure 32**) qui permettent de créer la carte affichée dans Google® Trends.

593	,		
594	Régions principales pour caries,		
595	Région, caries		
596	Pérou, 100		
597	Venezuela, 86		
598	Paraguay, 82		
599	Mexique, 74		
600	Honduras, 73		
601	Bolivie, 72		

Figure 32 : popularité de la recherche (sur 100) du terme « caries » dans différents pays grâce au fichier .csv fourni par Google Trends

Concernant le terme espagnol « caries », on voit que le maximum de popularité est atteint avec le Pérou qui obtient une valeur de 100 sur 100. Il va s'agir de notre « mètre étalon ».

Afin de débusquer les niches linguistiques, nous allons recouper de nouveaux chacun des 21 autres termes de l'échantillon A avec le terme « caries » et voir si dans le fichier .csv que nous fournit Google®, le Pérou reste toujours à un résultat de 100 sur 100.

En effet, à la vue de la très forte prévalence carieuse supposée en Amérique du Sud (majoritairement hispanophone) si lors de notre recherche de « caries » + « autre traduction », le Pérou n'arrive pas à 100, il se peut que l'autre terme traduit désigne autre chose que la carie dentaire.

La **figure 33** présente le résultat attendu pour valider l'absence de niches linguistiques. On retrouve toujours le Pérou à 100 sur 100 de popularité malgré l'intégration d'un nouveau terme dans la requête.

	A	B	C
595	,		
596	,		
597	Régions principales pour "caries"+"cariës",		
598	Région,"caries""+"cariës"		
599	Pérou,100		
600	Venezuela,86		
601	Paraguay,81		
602	Mexique,72		
603	Honduras,68		

Figure 33 : résultat de la recherche « carie »+ « cariës » qui valide l'absence de niche linguistique avec le Pérou qui apparaît toujours à 100.

Suite à ces investigations, seuls deux termes ne renvoient pas le Pérou à 100 sur 100 dans le fichier .csv.

Il s'agit en partie des termes où nous avons repérés une incohérence entre la langue et les régions populaires affichés dans la carte fourni par Google® Trends lors de la constitution de l'échantillon A :

- « kari » traduction en créole haïtien
- « cari » traduction en galicien

Au contraire, quand on recherche « caries »+ « cari » ou « caries »+ « kari », les résultats de popularité par pays ne renvoient plus le Pérou comme pays où la requête est la plus populaire comme l'indique la **figure 34**.

597 Régions principales pour "caries"+"kari",	597 Régions principales pour "caries"+"cari",
598 Région,"caries""+"kari"	598 Région,"caries""+"cari"
599 Finlande,100	599 Indonésie,100
600 Norvège,59	600 Malaisie,95
601 Malaisie,31	601 Brunéi Darussalam,25
602 Albanie,25	602 La Réunion,9
603 Kosovo,22	603 Singapour,8
604 Brunéi Darussalam,20	604 Pérou,4
605 Kenya,16	605 Venezuela,3
606 Lettonie,14	606 Bolivie,3
607 République tchèque,10	607 Chili,3

Figure 34 : résultats obtenus lors des requêtes « caries »+ « kari » et « caries »+ « cari ». Le Pérou n'est plus le pays le plus populaire.

La popularité des pays se trouve chamboulé par l'introduction d'un nouveau terme. On peut donc supposer que ce terme ne désigne pas forcément la carie dentaire mais **une** autre chose qui se trouve être plus populaire. Pour s'en assurer, nous recherchons les termes « kari » et « cari » dans le moteur de recherche Google[®] et regardons les résultats renvoyés, notamment les images.

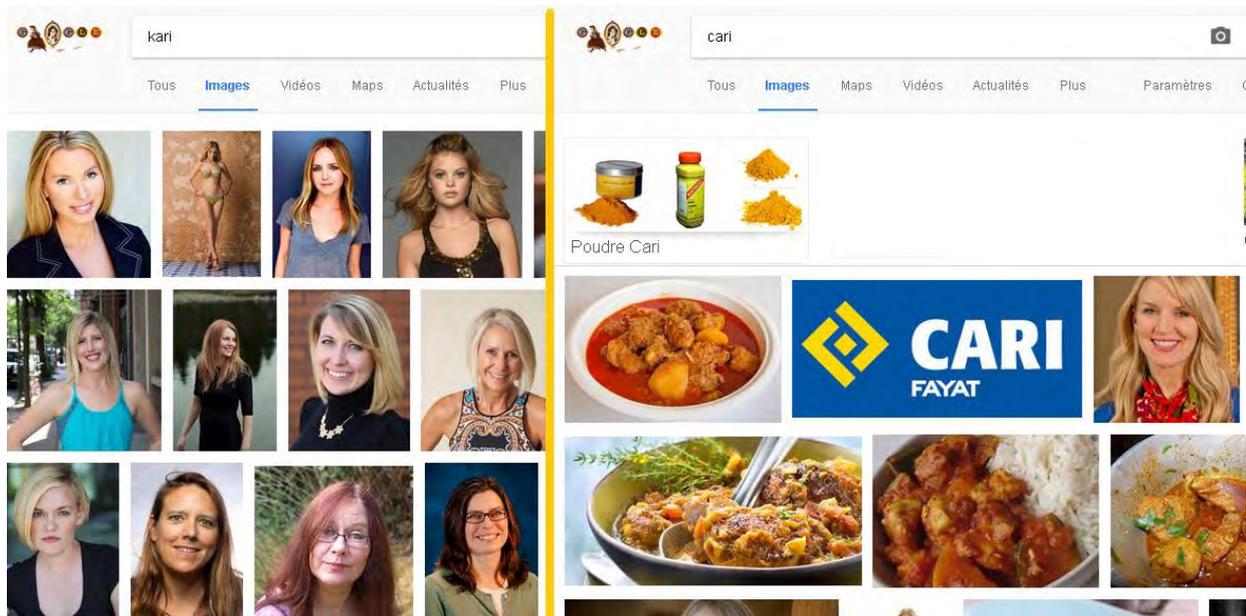


Figure 35 : images populaires renvoyées par Google lors de la recherche du terme « kari » et « cari » (66, 67).

Comme le démontre la **figure 35**, les termes « kari » et « cari » désignent des personnalités ou un plat.

Nous savons donc que nous devons éliminer ces deux termes qui fausseraient nos résultats en nous indiquant la popularité d'une requête ne concernant pas la carie dentaire.

Enfin, voulant nous assurer que les termes restants désignent bien la carie dentaire et ayant remarqué des incohérences entre la popularité géographique et la langue d'origine des termes « suus » et « kario », nous passons les termes restant au crible de la recherche Google[®] classique pour voir les images renvoyées. Ceci nous permet rapidement de voir si le sujet traité concerne bien la carie dentaire comme l'indique la **figure 36**.

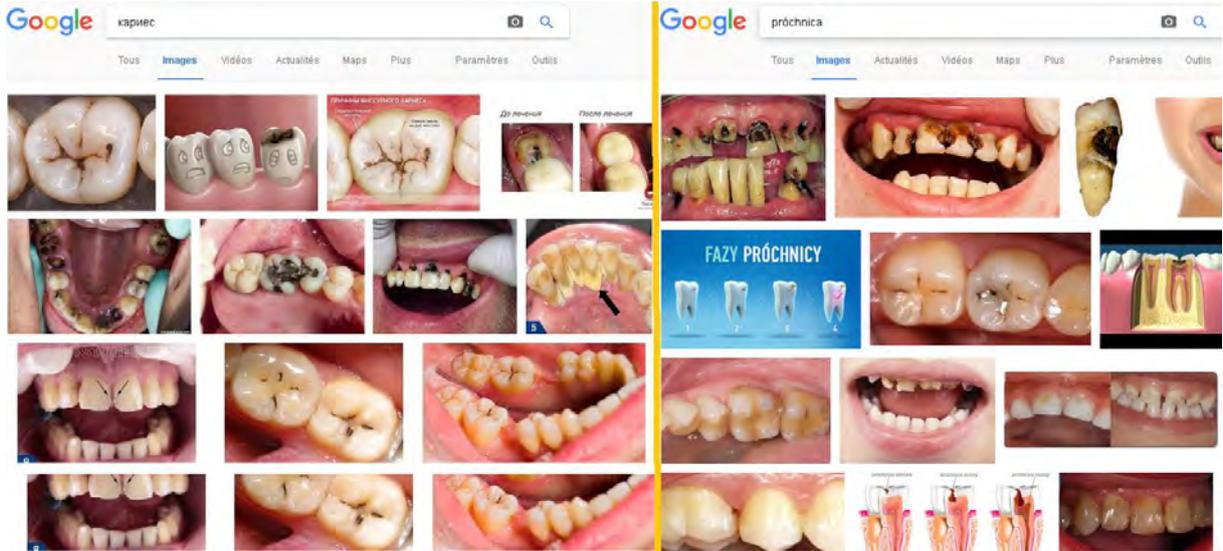


Figure 36 : iconographie obtenue dans Google Images pour les termes « кариес » et « próchnica » (68,69.)

Le passage de chacun des vingt termes restant par la recherche d'images via Google[®] nous confirme ce que nous supputions à propos des termes « suus » et « kario » à la vue de l'incohérence popularité géographique et langue d'origine.

« Suus » est entre autre une marque commerciale et une locution ; et « Kario » une personnalité et une boisson ainsi que le montre la **figure 37**.

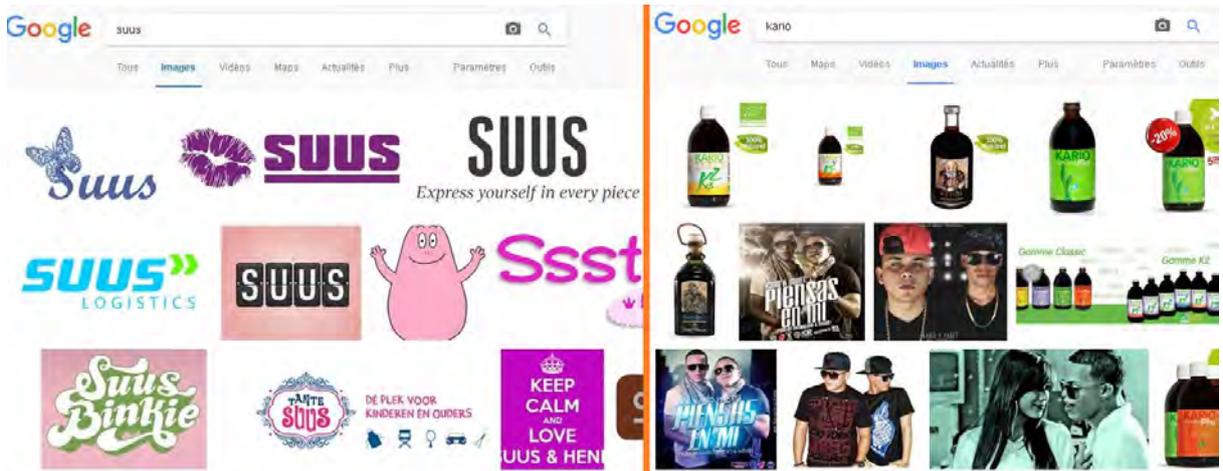


Figure 37 : exemple d'iconographie obtenue dans Google Images pour les termes « suus » et « kario » (70, 71)

Ainsi, nous savons que nous devons éliminer aussi les termes « suus » et « kario » de notre échantillon. Nous obtenons 18 termes.

Inclusion au pool de référence : le Japon

Enfin, nous sommes confrontés à la situation où la popularité dans le fichier .csv de la recherche « caries »+ « 虫歯 », le Pérou n'apparaît pas à 100 (figure 38). Mais contrairement aux autres termes, ici, il s'agit bien de la carie dentaire comme l'atteste la figure 39.

597	Régions principales pour "虫歯"+"caries",		
598	Régions pour "虫歯"+"caries",		
599	Japon,100		
600	Pérou,52		
601	Venezuela,45		
602	Paraguay,42		
603	Honduras,38		
604	Mexique,38		
605	Bolivie,35		

Figure 38 : résultats obtenus lors de la requête « caries »+ « 虫歯 ». Le Japon apparaît à 100 sur 100 de popularité.

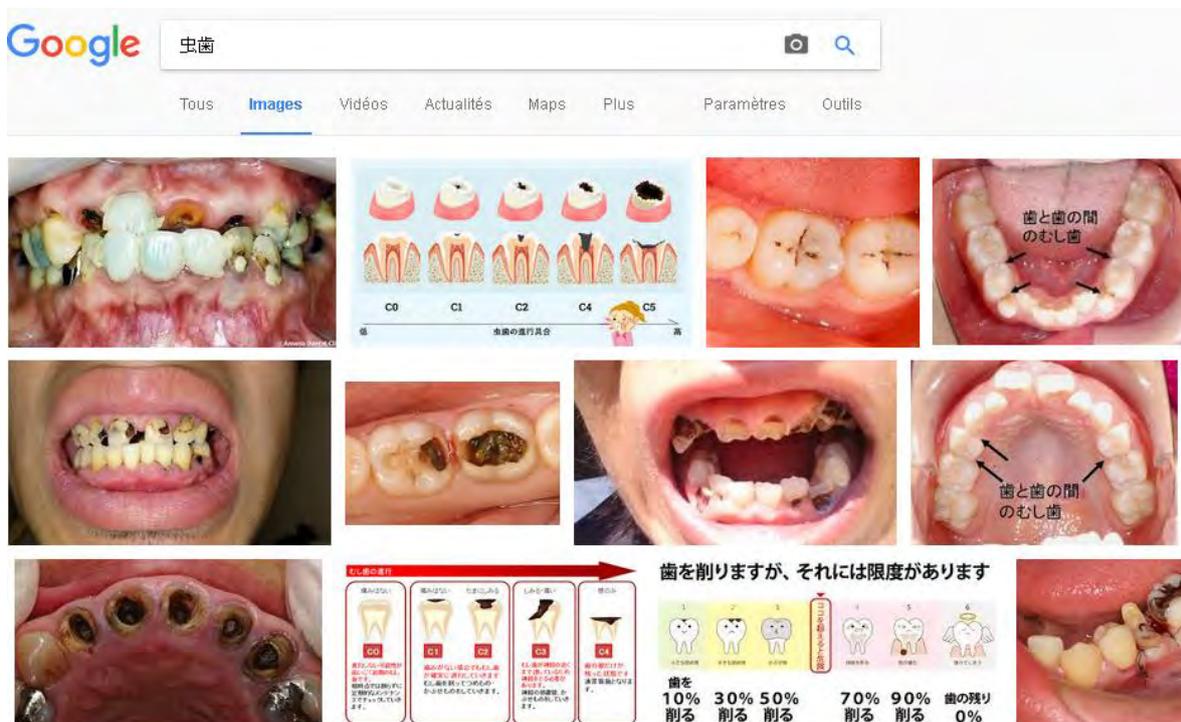


Figure 39 : exemple d'iconographie obtenue dans Google Images pour le terme« 虫歯 » (72).

Nous ne savons pas expliquer pourquoi le terme « carie dentaire » est aussi populaire, au point de supplanter « caries » (carie en espagnol) et donc de reléguer le Pérou en seconde position.

De même, concernant la popularité au cours du temps, nous observons un pic de popularité en août 2012 sur la **figure 40**, dont la raison nous est inconnue.

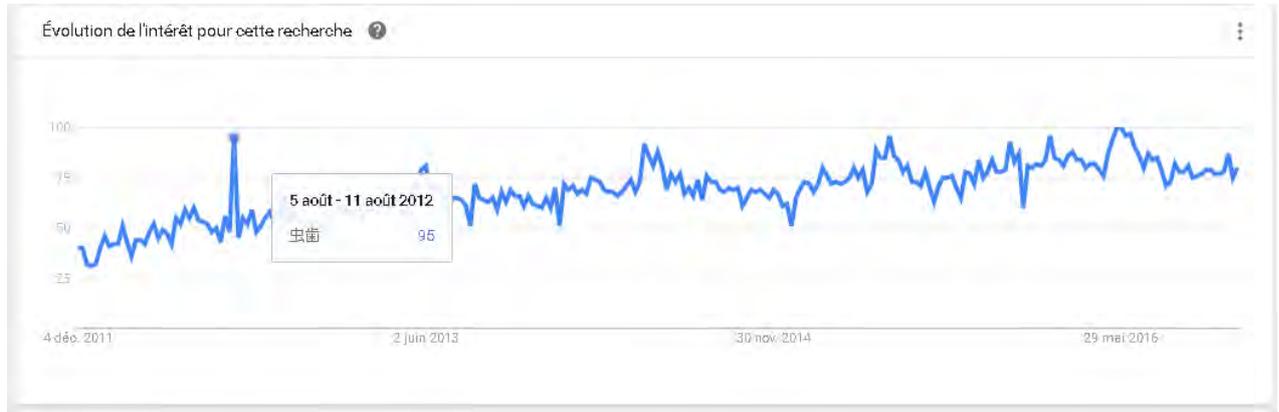


Figure 40 : pic de popularité de recherche de « 虫歯 » au cours du mois d'août 2012 (73).

Ainsi, devant ce résultat, il est nécessaire d'ajouter au pool de référence initial le terme « 虫歯 » pour ne pas fausser nos résultats.

A la suite de ces étapes, nous obtenons notre pool de référence final composés de 7 termes :

carie+caries+dentalcavity+cárie+虫歯 + نخر الأسنان

Pool de référence final

Les autres termes qui seront croisés avec ce pool permettront d'obtenir les données nécessaires à la réalisation de la carte.

Au total, ces 18 termes constituent notre **échantillon B**

4.2.2.7 Quatrième étape : obtention des données permettant de réaliser la carte

Grâce aux étapes précédentes, nous avons obtenu un pool de référence cohérent et nous pensons qu'aucun des termes utilisés ne présentent de niches linguistiques.

Cette dernière étape va consister à croiser chacun des termes exclus du pool de référence avec ce dernier.

Nous allons ainsi obtenir la valeur la plus précise de la popularité de recherche du terme « carie » pour chaque pays. Le pool de référence permettant de tenir compte du poids de chaque langue. Nous savons aussi que nous avons éliminé au maximum les interférences liées à des recherches ne concernant pas la carie dentaire.

La **figure 41** montre un exemple de requête où nous croisons un terme avec le pool de référence.

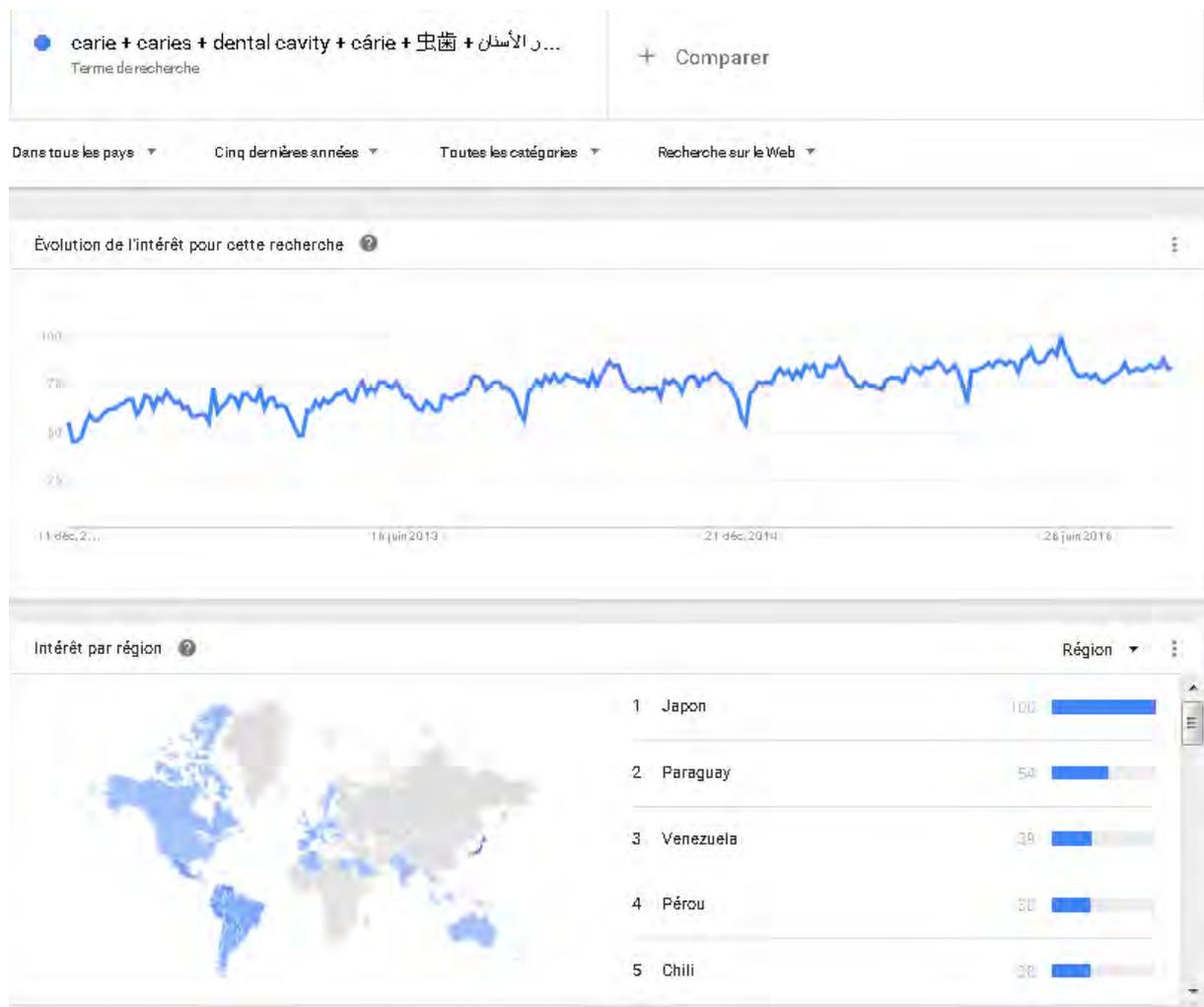


Figure 41 : résultat obtenu lors du croisement entre le terme « próchnica » et le pool de référence (74)

Ainsi, nous obtenons pour chacun des termes croisés avec le pool de référence un fichier .csv, exploitable dans un tableur, et qui nous indique la popularité sur 100, de chaque pays concerné par la recherche comme le montre la **figure 42**.

2	Région, próchnica+carie + caries + dental cavity + cárie + 虫歯 + نخر الأسنان.				
3	Egypte,12				
4	Emirats arabes unis,13				
5	Equateur,69				
6	Etats-Unis,21				
7	Afrique du Sud,13				
8	Algérie,23				
9	Allemagne,6				
10	Arabie saoudite,9				
11	Argentine,43				
12	Australie,15				
13	Belgique,17				
14	Bolivie,71				
15	Brésil,57				
16	Canada,25				
17	Chili,78				
18	Colombie,64				
19	Costa Rica,47				
20	Danemark,10				
21	El Salvador,60				
22	Espagne,38				
23	France,34				
24	Guatemala,58				

Figure 42 : fichier .csv exploitable obtenu à la suite de la requête « próchnica+carie + caries + dental cavity + cárie + 虫歯 + نخر الأسنان » A côté de chaque pays est indiquée sa popularité sur 100

Nous reportons dans un tableau, pour chaque requête, la popularité de chacun des pays listés dans le fichier .csv fourni par Google® Trends.

Certains pays vont apparaître seulement une fois, lors d'une requête précise, d'autres seront cités quasiment à chaque requête, notamment ceux dont la langue principale fait partie du pool de référence.

	Pool+próchnica	Pool+zubní kaz	Pool+ฟันผุ	Pool+кариес
Philippines	17	17	17	17
Pologne	28	5	5	4
Porto Rico	44	44	45	45
Portugal	31	32	32	31
République dominicaine	63	63	64	65
République tchèque		24		8
Roumanie	16	16	16	15
Royaume-Uni	14	14	14	12
Russie				1
Serbie				
Singapour	11	11	11	10
Slovénie				
Suède	5	5	5	5
Suisse	14	14	14	14
Taïwan				4
Thaïlande	7	6	70	6
Tunisie	27	27	27	26
Turquie	2	2	2	2
Ukraine				6
Uruguay	47	49	50	51
Venezuela	91	92	94	88

Figure 43 : aperçu du tableau obtenu grâce aux fichiers .csv où est indiquée la popularité des pays en fonction des différentes requêtes.

La **figure 43** donne un aperçu du tableau réalisé une fois toutes les requêtes réalisées. Il reprend la popularité de chacun des pays en fonction des requêtes. En vert, il s'agit de la popularité maximale atteinte. Si la case n'indique aucun nombre, le pays concerné n'est pas apparu dans le fichier .csv lors de la requête.

L'intégralité de ce tableau est disponible dans **l'annexe n°4**.

Toutes ces valeurs de popularité vont nous servir à établir la carte grâce à la valeur maximale obtenue par chaque pays.

Nous postulons que grâce au croisement du pool de référence avec les autres langues nous obtenons le volume de recherche le plus juste pour chaque pays.

Nous obtenons une liste de 72 pays pour lesquels nous disposons de la popularité et qui sont les suivants

:

- | | |
|-------------------------|----------------------------|
| 1. Afrique du Sud | 38. Japon |
| 2. Algérie | 39. Kazakhstan |
| 3. Allemagne | 40. Malaisie |
| 4. Arabie saoudite | 41. Maroc |
| 5. Argentine | 42. Mexique |
| 6. Australie | 43. Nicaragua |
| 7. Autriche | 44. Nigéria |
| 8. Belgique | 45. Norvège |
| 9. Biélorussie | 46. Nouvelle-Zélande |
| 10. Bolivie | 47. Pakistan |
| 11. Brésil | 48. Panama |
| 12. Bulgarie | 49. Paraguay |
| 13. Canada | 50. Pays-Bas |
| 14. Chili | 51. Pérou |
| 15. Chine | 52. Philippines |
| 16. Colombie | 53. Pologne |
| 17. Corée du Sud | 54. Porto Rico |
| 18. Costa Rica | 55. Portugal |
| 19. Croatie | 56. République dominicaine |
| 20. Danemark | 57. République tchèque |
| 21. Égypte | 58. Roumanie |
| 22. El Salvador | 59. Royaume-Uni |
| 23. Émirats arabes unis | 60. Russie |
| 24. Équateur | 61. Serbie |
| 25. Espagne | 62. Singapour |
| 26. États-Unis | 63. Slovénie |
| 27. Finlande | 64. Suède |
| 28. France | 65. Suisse |
| 29. Guatemala | 66. Taïwan |
| 30. Honduras | 67. Thaïlande |
| 31. Hong Kong | 68. Tunisie |
| 32. Hongrie | 69. Turquie |
| 33. Inde | 70. Ukraine |
| 34. Indonésie | 71. Uruguay |
| 35. Iran | 72. Venezuela |
| 36. Irlande | |
| 37. Italie | |

4.2.2.8 La check-list de Nuti et al. : rendre notre méthode transparente et reproductible

Dans un souci de transparence et afin de donner du poids à notre étude nous utilisons la check-list de Nuti et al. (51) traduite par Canceill (50). Elle permet de voir de manière claire la manière dont nous avons effectué nos recherches.

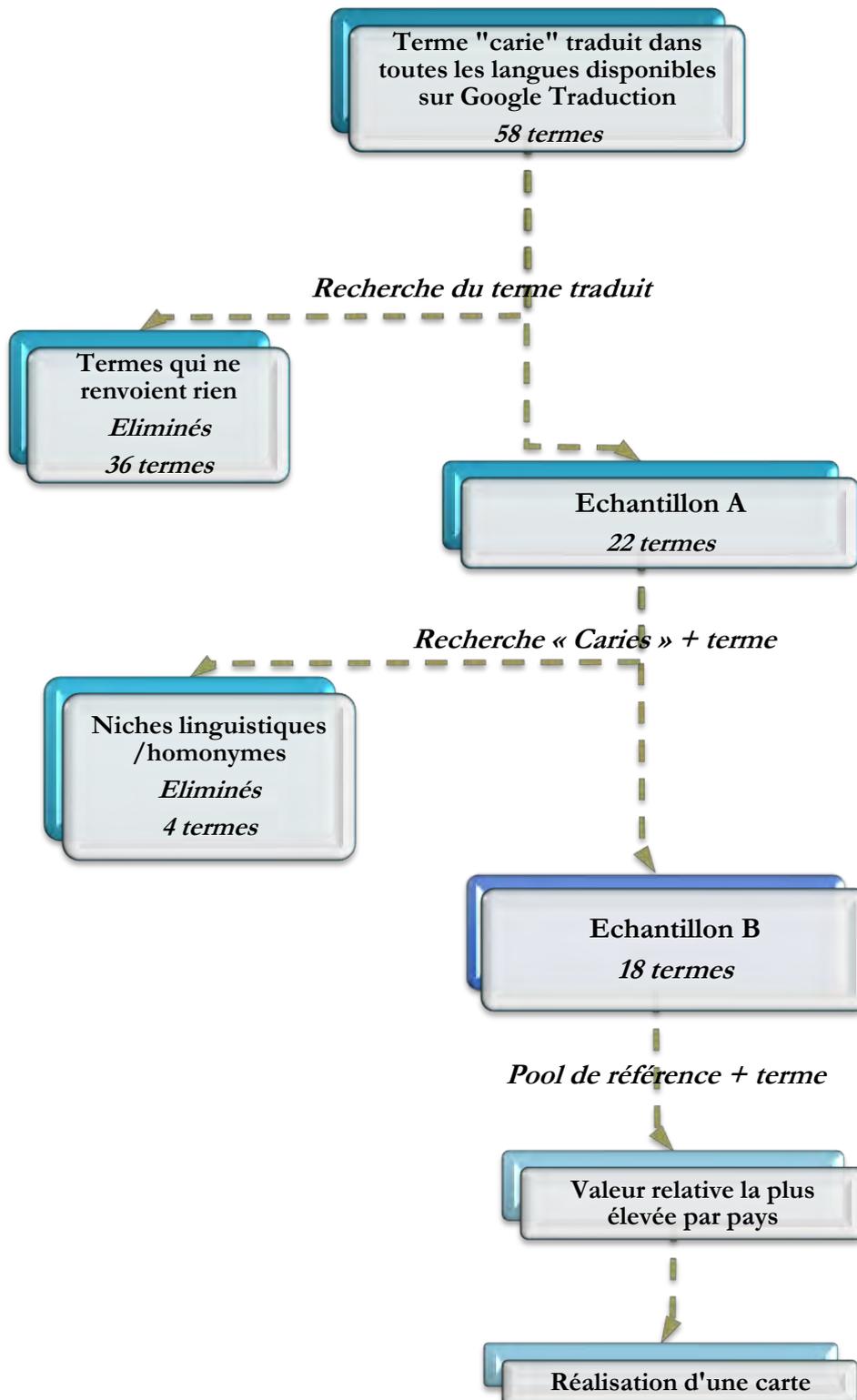
Tableau 3 : check-list de la procédure, d'après celle proposée par Nuti et coll. en 2014 (43)

SECTION	DONNEES	N°	CONTENU
Options de recherche	Dates d'accès	1	Entre le 30 mars et le 20 octobre 2015
	Période	2	Du 4 janvier 2004 au 08mars 2015
	Catégorie	3	Toutes les catégories de recherche
Termes recherchés	Ensemble des termes recherchés	4	<p><u>Pour l'échantillon A :</u></p> <ul style="list-style-type: none"> • carie, • caries, • dental caries, • dental cavity, • dental cavities, • نخر الأسنان • cariës, • karies, • Πυλκρωψυττορητιόν, • kariyes, • txantxarra, • অস্থিরক্ষয়রোগ, • карыес, • karijes, • кариес, • càries, • 齲, • 齲, • දිරායාම, • 카리에스, • kari, • kario, • kaaries, • cari, • pydredd, • კარიესის, • τερηδόνα,

Termes recherchés	Ensemble des termes recherchés	4	<ul style="list-style-type: none"> • અસ્થિક્ષય, • गणपुग • क्षय, • csontszú, • tannáta, • 虫齒, • ಕುಳಿಗಟು, • karioze, • éduonis, • Кариесо, • ദന്തക്ഷയം • snien, • किङणे, • цоорох, • kariesga, • દેહિગાઝઠું • کرم خورگی دندان. • , próchnica, • cária, • каријес, • zubnýkaz , • suus, • кариеси, • பற்சொத்தை, • zubníkaz, • క్షయాలను, • ฟันผุ, • yenirce, • кариєс, • chŭngxɯŋgmɯc • ,סג'אק <p><u>Pour l'absence de niche :</u></p> <ul style="list-style-type: none"> • caries+cari • caries+cária • caries+càries • caries+cariës • caries+cariës • caries+cariës • caries+cariës • caries+cariës • caries+kario • caries+próchnica • caries+suus • caries+zubníkaz • caries+кариес • caries+кариєс • caries+ฟันผุ
-------------------	--------------------------------	---	---

Termes recherchés	Ensemble des termes recherchés	4	<ul style="list-style-type: none"> • caries+ 虫齒 • caries+ 齲 <p><u>Pour l'échantillon B :</u></p> <ul style="list-style-type: none"> • Pool de référence = carie+ caries+ dentalcavity + cárie+ 虫齒+ نخر الأسنان • Pool+dental caries • Pool+dentalcavities • Pool+karies • Pool+karijes • Pool+ฟันผุ • Pool+ 虫齒 • Pool+càries • Pool+cariës • Pool+próchnica • Pool+zubníkaz • Pool+καριες • Pool+кариес • Pool+齲
	Combinaison des termes	4a	Les termes sont tous associés par le symbole « + » sauf « dental cavity », « dental caries » et « dental cavities »
	Marques de citations	4b	<p>Les guillemets sont utilisés pour tous les termes lors de la constitution de l'échantillon A et la recherche de niche linguistique.</p> <p>Ils sont abandonnés lors du croisement des termes avec le pool de référence</p>
Raisons, stratégies	Des termes recherchés	5	<p>Vérification de la présence de données exploitables pour l'échantillon A</p> <p>Croisement des termes pour vérifier l'absence de niches linguistiques et d'homonymes</p> <p>Croisement du pool de référence avec les termes n'en faisant pas partie pour obtenir la popularité maximale de chaque pays</p>

Raisons, stratégies	Des options de recherche	6	<p>Cantonnement de la recherche de 2004 au 8 mars 2015 pour s'affranchir de la différence de mode de calcul entre les données récentes (< 7jours)</p> <p>Pas d'utilisation des catégories de recherche cantonnées à la santé car fonctionnement trop opaque</p> <p>Pas de limites géographiques car volonté d'obtenir une cartographie mondiale</p>
---------------------	--------------------------	---	--

4.2.2.9 Diagramme de flux de l'étude

4.3. Résultats

Le tableau permet d'obtenir la valeur de la popularité pour 72 pays.

Pour avoir une lecture aisée de nos résultats, nous avons besoin d'élaborer une carte issue des données obtenues avec Google® Trends.

Pour cela, nous nous sommes rapprochés de l'Atelier de Cartographie de l'Université du Mirail à Toulouse, qui a établi la carte la plus appropriée pour illustrer au mieux ces valeurs relatives obtenues.

4.3.1. Méthode de réalisation de la carte

Pour réaliser la carte, il faut procéder à une discrétisation des données afin que ces dernières puissent être ensuite représentées sur une carte.

Comme l'indique Cauvin, Reymond et Serradj dans Discrétisation et représentation géographique (75) : « La discrétisation est l'opération qui permet de découper en classes une série de variables qualitatives ou de variables quantitatives. [...] L'opération de discrétisation doit satisfaire à la fois aux exigences de la représentation cartographiques et à celles des principes statistiques. Elle doit conserver les caractéristiques essentielles présentées par les données, perdre le moins d'information possible, mais aussi respecter les règles de la perception visuelle afin de transmettre une information géographique efficace et de qualité. »

L'Atelier de Cartographie a choisi comme moyen de discrétisation la méthode de Jenks (**Figure 44**).

- Méthode de Jenks ou d'analyse de variance.

Cette méthode propose, sur la base d'une analyse de variance, de minimiser la variance intra-classe, et de maximiser la variance inter-classe.

- On ordonne les données par valeur croissante
- On cherche tous les groupes possibles pouvant former k classes.
- Pour chaque configuration, on calcule la variance à l'intérieur du groupe, et entre les groupes.
- On compare l'ensemble des valeurs et on prend la configuration qui minimise la variance dans le groupe et maximise la variance entre les groupes.

Avantage:

Cette méthode est la meilleure du point de vue statistique parce qu'elle crée des groupes homogènes et très hétérogènes entre eux.

Les limites de classes sont des valeurs réelles de la distribution.

Inconvénient:

Peu de stabilité en augmentant le nombre de classes.

Ne peut s'appliquer qu'à des valeurs faibles (problème d'algorithme)

Figure 44 : avantages et inconvénients de la méthode de Jenks (76).

4.3.2. Cartes

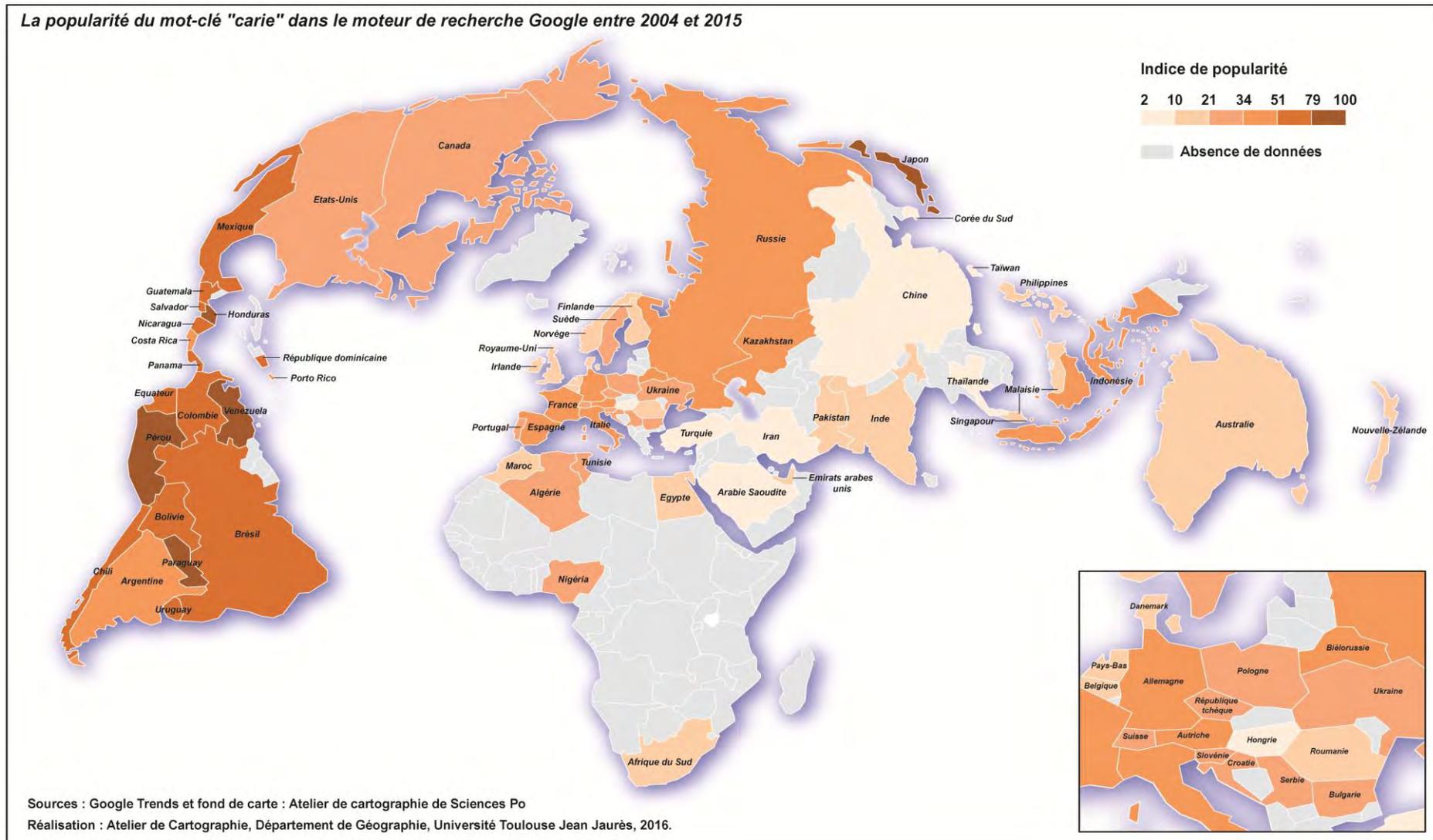
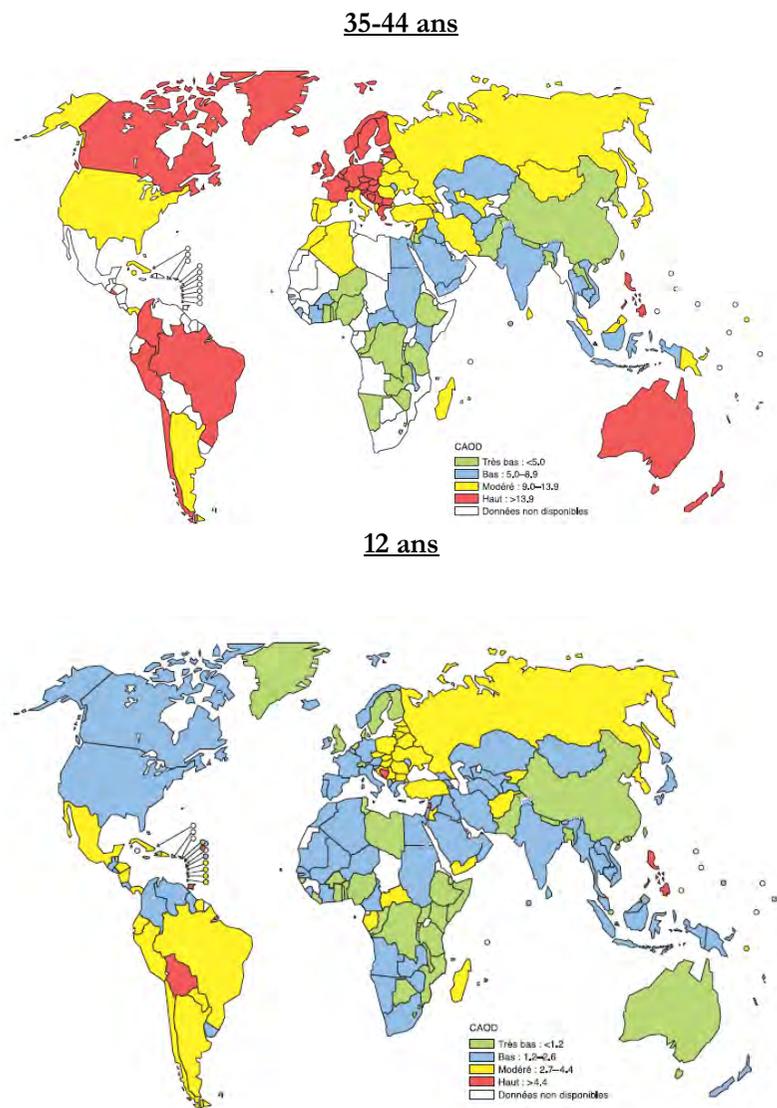
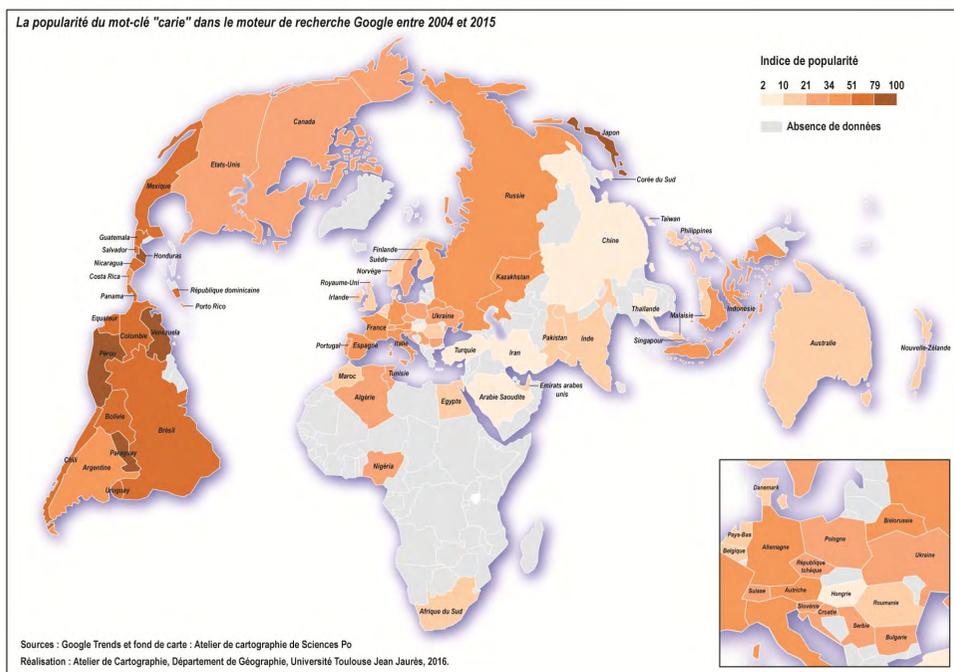
4.3.2.1 Carte Google® Trends

Figure 45 : carte Google® Trends.

4.3.2.2 Correspondance entre la carte des requêtes Google Trends® et les cartes de l’OMS

Figure 46 : comparaison visuelle entre la carte Google Trends® et les cartes de l’OMS (10).



Visuellement, la carte OMS des enfants laisse apparaître une plus grande correspondance avec la carte Google Trends que celle des adultes.

4.3.2.3 Comparaison statistique des cartes

Les valeurs représentées sur la carte issue de Google® Trends sont des valeurs relatives et non absolues comme les cartes de l'OMS :

- En effet, il s'agit de la popularité de recherche dans des pays les uns par rapport aux autres.

- Au contraire, les valeurs obtenues par l'OMS sont absolues. Elles estiment une moyenne de nombre de caries par personne et pays via l'indice CAOD.

Pour pouvoir les comparer avec des outils statistiques nous avons la nécessité de dichotomiser les données en langage binaire. Selon leur position par rapport à la médiane, chaque donnée se verra assignée au groupe 0 ou au groupe 1.

Pour les cartes OMS, suite à notre classement des pays selon leur indice CAOD moyen, nous avons dichotomisé en deux groupes constitués des pays pour lesquels l'indice est qualifié :

- de très bas à bas : assignation au groupe 0
- de modéré à haut : assignation au groupe 1

CAOD	DICHOTOMIE
Très bas : <5.0	0
Bas : 5.0–8.9	0
Modéré : 9.0–13.9	1
Haut : >13.9	1
Données non disponibles	

Figure 47 : correspondance entre l'indice CAOD et la dichotomisation.

De la même manière, nous avons dichotomisé la distribution des valeurs issues de Google Trends® à partir de la médiane : les 50% des valeurs les plus hautes ont été affiliées au groupe 1 et les 50% des valeurs les plus basses à 0.

4.3.2.4 Résultats et concordance des cartes

Le test exact de Fisher montre une association significative ($p=0,03$) pour la correspondance entre les cartes OMS enfants et Google Trends.

Ce n'est pas le cas pour les cartes OMS adultes et Google Trends ($p=0,48$).

Ce test statistique vient donc confirmer ce que nous observions visuellement.

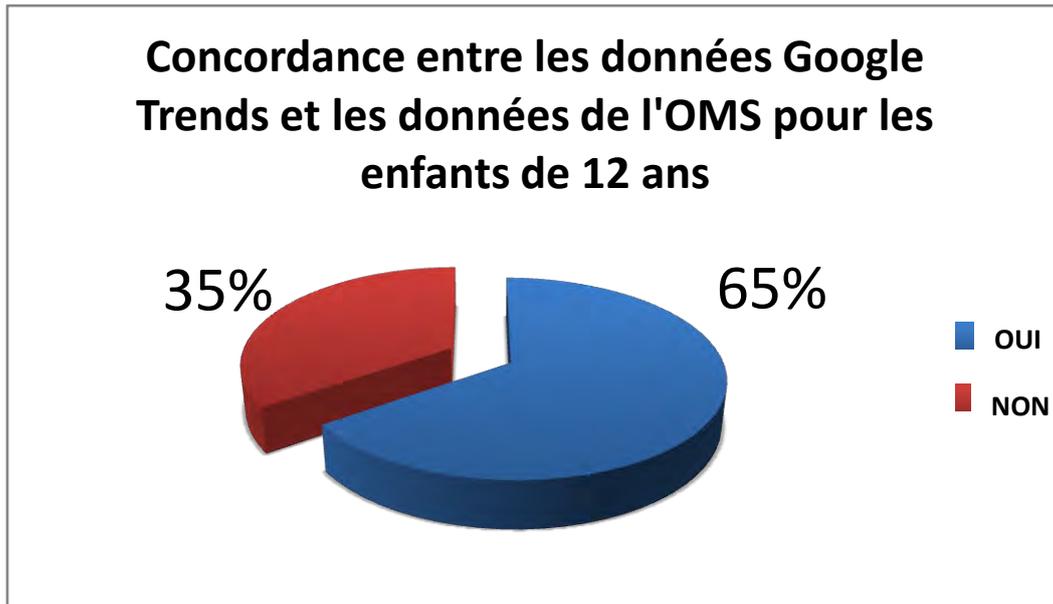


Figure 48 : pourcentage de concordance (en bleu) entre les données Google Trends et les données de l'OMS pour les enfants de 12 ans.

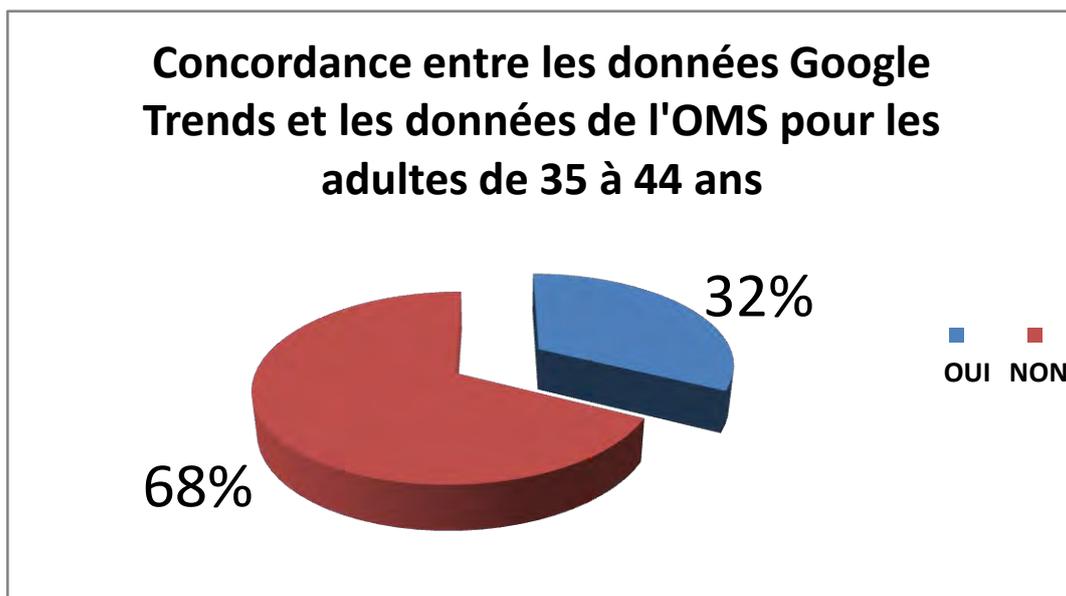


Figure 49 : pourcentage de concordance (en bleu) entre les données Google Trends et les données de l'OMS pour les adultes de 35 à 44 ans.

4.4. Discussion

Ce travail de thèse apporte plusieurs perspectives intéressantes en termes de santé publique bucco-dentaire. Notre étude nous a permis de trouver une significativité entre les données obtenues via Google® Trends et la carte OMS concernant l'indice CAOD des enfants de 12 ans. Cependant, concernant les adultes de 35 à 44 ans, nous voyons clairement que les données obtenues via Google® Trends ne sont pas superposables

Ainsi, le constat que ce que nous pouvons qualifier de « préoccupation globale d'une population vis-à-vis de la maladie carieuse » semble être assez bien corrélé à l'estimation de la prévalence carieuse chez l'enfant. Nous n'avons pas trouvé une telle corrélation chez l'adulte. Ceci peut s'expliquer par un effet temporel différentiel entre la nature de la variable « requête dans un moteur de recherche » (correspondant à une « préoccupation » instantanée) et celle de l'indice CAO (cumulatif à l'échelle de toute une vie). On peut supposer que l'indice CAO moyen des enfants d'une population reflète une situation transversale, instantanée. En revanche, il est probable que le CAO adulte moyen d'un pays soit la conséquence longitudinale de nombreux phénomènes politiques, médico-socio-économiques, démographiques et culturels qui se succèdent dans le temps. Le CAO moyen des enfants de 12 ans est, lui, vraisemblablement, davantage interprétable comme le reflet des balances entre les facteurs de risque et les facteurs de protection à l'échelle d'une population, à un moment donné.

Ensuite, il semble qu'au niveau mondial, la préoccupation la plus élevée vis-à-vis de la carie dentaire se trouve en Amérique du Sud, ce qui correspond aux données épidémiologiques de la prévalence carieuse des enfants de certains pays de ce continent. La forte augmentation, assez récente, de la préoccupation vis-à-vis de la maladie carieuse au Japon est difficilement interprétable, malgré nos efforts (nous avons contacté à ce sujet deux responsables japonais spécialisés en épidémiologie bucco-dentaire et travaillant avec l'OMS, mais n'avons pas obtenu de réponses à nos messages).

Il existe de nombreuses limites à ce travail, elles sont communes à tous les travaux de ce type et ont été décrites dans la partie relative aux analyses des données massives en santé. Par exemple, on observe sur la carte Google® Trends que beaucoup de pays désignés en voie de développement sont absents. Cela est certainement dû à la difficulté des populations de ces zones géographiques à accéder à une connexion internet.

Utilisé comme un outil de recherche en santé, Google® Trends présente des défauts qui sont à la fois ceux qu'on peut reprocher en général à tous les outils d'exploitation du Big Data, mais aussi propres à son fonctionnement qui manque de transparence. Il ne permet pas de monter des modèles complexes d'analyse des résultats. Seule la volonté de transparence de Google® concernant la méthode de calcul de ce volume pourra permettre d'en faire un outil de recherche en épidémiologie plus fiable

Conclusion

L'utilisation des données de requête peut s'avérer être un complément utile des méthodes de surveillance de santé publique traditionnelles. En effet, le coût des études d'observation pour estimer la prévalence de la maladie carieuse n'est pas négligeable (ceci explique d'ailleurs la faible quantité de données fiables et récentes dans beaucoup de pays, y compris en France). L'analyse des données de requête permet l'exploration rapide de données médicales certes distales de toute considération diagnostique, mais particulièrement standardisées, localisables et quantifiables. Il faut rester très prudent sur l'interprétation de ces données, qui ne peuvent en rien se substituer aux mesures de terrain. Il faut toujours garder à l'esprit les règles et la rigueur qui s'appliquent à l'épidémiologie classique. Toutefois, elles ouvrent des perspectives très intéressantes.

L'exploitation des données massives en santé est promise à un avenir certain. Nous ne sommes qu'à l'aube de toutes les possibilités que les analyses des données massives dans le champ de la santé peuvent apporter. L'essor des objets connectés, des capteurs divers (y compris au niveau bucco-dentaire) laisse entrevoir une myriade de perspectives, tout aussi prometteuses qu'inquiétantes. Il existe un risque de dépersonnalisation de la médecine, d'ultra-surveillance des individus et de perte de libertés individuelles qu'il faudra toujours prendre en compte dans ce type de recherche.

Vu, le directeur de thèse
Docteur J-N. VERGNES



le 12/01/2017

Vu, le président du jury
Professeur C. NABET



Le 12/01/2017

Table des illustrations

Figure 1 : logo de l'UFSBD (3).	19
Figure 2 : logo de l'Organisation Mondiale de la Santé (7).	20
Figure 3 : logo de la Fédération Dentaire Internationale (11).	22
Figure 4 : modèle conceptuel de santé orale des enfants proposé par Fisher –Ovvens et al. (13).	25
Figure 5 : indice CAOD moyen dans le monde chez les enfants de 12 ans (10).	31
Figure 6 : indice CAOD moyen dans le monde chez les adultes de 35- 44 ans (10).	32
Figure 7 : parts de marché des moteurs de recherche en 2015 (17).	34
Figure 8 : page d'accueil de Google [®] Trends (23).	38
Figure 9 : évolution dans le temps de la popularité de recherche du terme « Noël » (24).	39
Figure 10 : évolution dans l'espace géographique de la popularité de recherche du terme « Noël » au cours des cinq dernières années (24).	41
Figure 11 : évolution dans l'espace géographique de la popularité de recherche du terme « Noël » au cours des cinq dernières années au niveau de la France et de la région Occitanie (24).	42
Figure 12 : sujets et requêtes associés au terme Noël lors des recherches au cours des 5 dernières années (24).	42
Figure 13 : différences de résultats obtenus selon le niveau de précisions de la recherche (28, 29).	44
Figure 14 : filtrage global disponible sous la barre de recherche et filtres de période ou de lieu appliqués à un groupe de terme en cliquant sur les trois points à droite de ce dernier (30).	45
Figure 15 : comparaison de l'évolution de la popularité de recherche entre « Noël » en bleu et « Pâques » en rouge au cours des 5 dernières années (32).	47
Figure 16 : cartes et requêtes associées obtenues lors de la comparaison des termes « Noël » et « Pâques » (32).	49
Figure 17 : comparaison de la popularité de recherche du terme « Noël » entre la France et le monde (33).	51
Figure 18 : comparaison de la popularité du terme de recherche « Noël » entre 2007 et 2012 et les cinq dernières années (34).	53
Figure 19 : « Noël » à la fois terme classique ou sujet (24).	56
Figure 20 : nombre d'articles publiés sur PubMed concernant le Big Data au cours des 5 dernières années.	58
Figure 21 : liste des vérifications à faire lors de l'utilisation de Google [®] Trends d'après Nuti et al. (51).	65
Figure 22 : traduction par Canceill (50) de la checklist de Nuti et al. (51).	66
Figure 23 : pourcentage estimé de la population américaine atteinte de symptômes grippaux : différences entre plusieurs systèmes de mesure (52).	70
Figure 24 : courbe obtenue par Canceill dans Google [®] Trends en tapant "amalgame" sans guillemets le 06/02/2016 dans Google [®] Trends. (50.)	72
Figure 25 : listes des langues disponibles dans Google [®] Traduction (60).	79
Figure 26 : exemple d'utilisation de Google [®] Traduction avec le terme « partir » (60).	80
Figure 27 : « карыес » (biélorusse) ne renvoie pas de résultats dans Google [®] Trends (61).	82
Figure 28 : régions où la recherche de « cari » est populaire : incohérence entre la région de langue et les lieux de popularité (62).	83
Figure 29 : « karijes » renvoie suffisamment d'informations et sans incohérence (63).	84
Figure 30 : répartition des langues dans le monde (64)	86
Figure 31 : popularité de la recherche (sur 100) du terme « caries » pour chaque semaine grâce au fichier .csv fourni par Google Trends	87
Figure 32 : popularité de la recherche (sur 100) du terme « caries » dans différents pays grâce au fichier .csv fourni par Google Trends	88

Figure 33 : résultat de la recherche « carie »+ « cariès » qui valide l'absence de niche linguistique avec le Pérou qui apparaît toujours à 100. _____	89
Figure 34 : résultats obtenus lors des requêtes « caries »+ « kari » et « caries »+ « cari ». Le Pérou n'est plus le pays le plus populaire. _____	89
Figure 35 : images populaires renvoyées par Google lors de la recherche du terme « kari » et « cari » (66, 67). _____	90
Figure 36 : iconographie obtenue dans Google Images pour les termes « kapuec » et « próchnica » (68,69.) _____	91
Figure 37 : exemple d'iconographie obtenue dans Google Images pour les termes « suus » et « kario » (70, 71) _____	91
Figure 38 : résultats obtenus lors de la requête « caries »+ « 虫歯 ». Le Japon apparaît à 100 sur 100 de popularité. _____	92
Figure 39 : exemple d'iconographie obtenue dans Google Images pour le terme « 虫歯 » (72). _____	92
Figure 40 : pic de popularité de recherche de « 虫歯 » au cours du mois d'août 2012 (73). _____	93
Figure 41 : résultat obtenu lors du croisement entre le terme « próchnica » et le pool de référence (74) _____	94
Figure 42 : fichier .csv exploitable obtenu à la suite de la requête « próchnica+carie + caries + dental cavity + cárie + 虫歯 + زخرالاً سدمات » A côté de chaque pays est indiquée sa popularité sur 100 _____	95
Figure 43 : aperçu du tableau obtenu grâce aux fichiers .csv où est indiquée la popularité des pays en fonction des différentes requêtes. _____	96
Figure 44 : avantages et inconvénients de la méthode de Jenks (76). _____	104
Figure 45 : carte Google® Trends. _____	105
Figure 46 : comparaison visuelle entre la carte Google Trends® et les cartes de l'OMS (10). _____	106
Figure 47 : correspondance entre l'indice CAOD et la dichotomisation. _____	107
Figure 48 : pourcentage de concordance (en bleu) entre les données Google Trends et les données _____	108
Figure 49 : pourcentage de concordance (en bleu) entre les données Google Trends et les données _____	108

Table des tableaux

Tableau 1 : Liste des pays en fonction de leur indice CAOD chez les adultes de 35 à 44 ans d'après la carte de l'OMS (10). _____	76
Tableau 2 : Liste des pays en fonction de leur indice CAOD chez les enfants de 12 ans d'après la carte de l'OMS (10). _____	77
Tableau 3 : check-list de la procédure, d'après celle proposée par Nuti et coll. en 2014 (43). _____	98

Bibliographie

1. Petersen PE. World Health Organization global policy for improvement of oral health-World Health Assembly 2007. *Int Dent J.* 2008;58(3):115–121.
2. Jeanblanc G, Poullie A-I, Rumeau-Pichon C, Despeyroux S, Lombry Y, Pages F, et al. Stratégies de prévention de la carie dentaire. Haute Autorité de Santé; 2010 mars p. 180.
3. Brochures de présentation | UFSBD [Internet]. [cité 22 juill 2016]. Disponible sur: <http://www.ufsbd.fr/espace-public/espace-presse/brochures-de-presentation/>
4. Gigon G. Le moteur de recherche internet GOOGLE, un allié pour le médecin généraliste dans son cabinet médical de ville ? [Thèse d'exercice]. [France]: Université Pierre et Marie Curie (Paris). UFR de médecine Pierre et Marie Curie; 2013.
5. Direction Générale de la Santé B de la santé des populations. Synthèse du plan bucco-dentaire [Internet]. 2006 p. 7. Disponible sur: http://social-sante.gouv.fr/IMG/pdf/Synthese_du_plan_bucco-dentaire.pdf
6. Dates clés | UFSBD [Internet]. [cité 19 mai 2016]. Disponible sur: <http://www.ufsbd.fr/espace-public/qui-sommes-nous/dates-cles/>
7. OMS | Qui nous sommes [Internet]. WHO. [cité 25 mai 2016]. Disponible sur: <http://www.who.int/about/who-we-are/fr/>
8. OMS | À propos de l'OMS [Internet]. WHO. [cité 25 mai 2016]. Disponible sur: <http://www.who.int/about/fr/>
9. OMS | Ce que nous faisons [Internet]. WHO. [cité 25 mai 2016]. Disponible sur: <http://www.who.int/about/what-we-do/fr/>
10. Petersen PE. Rapport sur la santé bucco-dentaire dans le monde. [cité 20 mai 2016]; Disponible sur: http://wwwlive.who.int/entity/oral_health/media/en/orh_report03_fr.pdf
11. Glick M, Monteiro Da Silva O, Seeberger G, Xu T, Pucca G, Williams D, et al. La Vision 2020 de la FDI : une prospection sur l'avenir de la santé bucco-dentaire [Internet]. FDI; 2012 déc p. 28. Disponible sur: http://apps.who.int/iris/bitstream/10665/39310/1/WHO_OFFSET_103_fre.pdf
12. Carie Dentaire-Dictionnaire médical de l'Académie de Médecine [Internet]. [cité 30 mai 2016]. Disponible sur: <http://dictionnaire.academie-medecine.fr/?q=carie+dentaire>
13. Fisher-Owens SA, Gansky SA, Platt LJ, Weintraub JA, Soobader M-J, Bramlett MD, et al. Influences on children's oral health: a conceptual model. *Pediatrics.* sept 2007;120(3):e510-520.
14. Direction Générale de la Santé B de la santé des populations « MC1 ». Plan national de prévention bucco-dentaire 2006-2009 [Internet]. Ministère de la Santé; 2006 p. 2. Disponible sur: http://social-sante.gouv.fr/IMG/pdf/Plan_national_de_prevention_bucco-dentaire_2006-2010.pdf

15. Google. In: Wikipédia [Internet]. 2016 [cité 3 janv 2017]. Disponible sur: <https://fr.wikipedia.org/w/index.php?title=Google&oldid=132980177>
16. Bort J, May 13 2013, 12 785, 2. Nearly 500 Million Searches A Day Are For Things Google Has Never Heard Of [Internet]. Business Insider. [cité 5 sept 2016]. Disponible sur: <http://www.businessinsider.com/500m-things-google-has-never-heard-of-2013-5>
17. 11:14 LRJM à jour le 06/01/16. Parts de marché des moteurs de recherche dans le monde [Internet]. [cité 5 sept 2016]. Disponible sur: <http://www.journaldunet.com/ebusiness/le-net/1087491-parts-de-marche-des-moteurs-de-recherche-dans-le-monde/>
18. À propos de Google - Produits [Internet]. [cité 5 sept 2016]. Disponible sur: <https://www.google.fr/intl/fr/about/products/>
19. ► La Liste des 200 Services et Applications Google [Guide Complet] [Internet]. ① Webmarketing Conseil - Attirez votre Client Idéal. 2016 [cité 5 sept 2016]. Disponible sur: <http://www.webmarketing-conseil.fr/google/>
20. Google Trends [Internet]. [cité 21 mars 2016]. Disponible sur: <http://www.webrankinfo.com/google/trends.htm>
21. Origine des données de l'outil Trends - Aide Google Tendances des recherches [Internet]. [cité 3 janv 2017]. Disponible sur: https://support.google.com/trends/answer/4355213?hl=fr&ref_topic=6248052
22. Ajustement des données Google Trends - Aide Google Tendances des recherches [Internet]. [cité 3 janv 2017]. Disponible sur: https://support.google.com/trends/answer/4365533?hl=fr&ref_topic=6248052
23. Google Trends-Page d'accueil [Internet]. Disponible sur: <https://www.google.fr/trends/?hl=fr>
24. Google Trends-Noël [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=no%C3%AB>
25. Localiser les recherches associées - Aide Google Tendances des recherches [Internet]. [cité 22 juill 2016]. Disponible sur: https://support.google.com/trends/answer/4355000?hl=fr&ref_topic=4365530
26. Dépannage - Aide Google Tendances des recherches [Internet]. [cité 4 sept 2016]. Disponible sur: https://support.google.com/trends/answer/4359078?hl=fr&ref_topic=4365599
27. Conseils de recherche pour Google Trends - Aide Google Tendances des recherches [Internet]. [cité 4 sept 2016]. Disponible sur: https://support.google.com/trends/answer/4359582?hl=fr&ref_topic=4365530
28. Google Trends - Fleur Lotus [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=fleur%20lotus>
29. Google Trends- Fleur+Lotus [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=fleur%20%2B%20lotus>

30. Google Trends - brosse à dents versus dentifrice [Internet]. Disponible sur: <https://www.google.fr/trends/explore?geo=FR&q=brosse%20%20dents,dentifrice>
31. Affiner les résultats des tendances par catégorie - Aide Google Tendances des recherches [Internet]. [cité 4 sept 2016]. Disponible sur: https://support.google.com/trends/answer/4359597?hl=fr&ref_topic=4365530
32. Google Trends-Noël versus Pâques [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=No%20%20P%C3%A2ques>
33. Google Trends - Noël : France versus Monde [Internet]. Disponible sur: <https://www.google.fr/trends/explore?date=today%20-y,today%20-y&geo=FR,&q=No%20%20ABl,No%20%20ABl>
34. Google Trends - Noël : 2007 à 2012 versus 2012 à 2016 [Internet]. Disponible sur: <https://www.google.fr/trends/explore?date=today%20-y,2007-01-01%202012-01-01&geo=,&q=No%20%20ABl,No%20%20ABl>
35. Comparer les tendances relatives aux termes de recherche - Aide Google Tendances des recherches [Internet]. [cité 12 janv 2017]. Disponible sur: https://support.google.com/trends/answer/4359550?hl=fr&ref_topic=4365530
36. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. *Am J Emerg Med.* sept 2014;32(9):1016-23.
37. Davidson MW, Haim DA, Radin JM. Using networks to combine « big data » and traditional surveillance to improve influenza predictions. *Sci Rep.* 2015;5:8154.
38. Khoury MJ, Ioannidis JPA. Medicine. Big data meets public health. *Science.* 28 nov 2014;346(6213):1054-5.
39. Boyd D, Crawford K. Critical Questions for Big Data. *Inf Commun Soc.* 1 juin 2012;15(5):662-79.
40. Jeannot J-G. Dr Google : moteur-de-recherche-médical. org, un moteur de recherche Google pour la médecine. *Rev Med Suisse.* 2008;4(158):1280-4.
41. Boini S, Serge Briançon. Chapitre 2 - L'épidémiologie - Qu'est-ce que l'épidémiologie ? [Internet]. 2009 [cité 19 sept 2016]. Disponible sur: http://www.sante-pub.u-nancy.fr/presentation/ILLUSTRATIONS/UE1_Chap2/co/Epidemiologie.html
42. Frijters P, Johnston D-W, Lordan G, Shields M-A, =Centre of Health Economics. Monash University. AUS. Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the US. *Soc Sci Med.* 2013;84:61-8.
43. Song TM, Song J, An J-Y, Hayman LL, Woo J-M. Psychological and Social Factors Affecting Internet Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends. *Yonsei Med J.* 2014;55(1):254.

44. Fazeli Dehkordy S, Carlos RC, Hall KS, Dalton VK. Novel Data Sources for Women's Health Research: Mapping Breast Screening Online Information Seeking Through Google Trends. *Acad Radiol.* sept 2014;21(9):1172-6.
45. Ayers JW, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, et al. Leveraging Big Data to Improve Health Awareness Campaigns: A Novel Evaluation of the Great American Smokeout. *JMIR Public Health Surveill.* juin 2016;2(1):e16.
46. Mohammad Tufail Malik, Gumel A, Thompson L-H, Strome T, Mahmud S-M, =Department of Mathematics. University of Manitoba. Winnipeg. MB. CAN, et al. « Google Flu Trends » and Emergency Department Triage Data Predicted the 2009 Pandemic H1 N1 Waves in Manitoba. *Rev Can SANTE PUBLIQUE Can J PUBLIC Health.* 2011;102(4):294-7.
47. Google Flu Trends. In: Wikipedia, the free encyclopedia [Internet]. 2016 [cité 28 sept 2016]. Disponible sur: https://en.wikipedia.org/w/index.php?title=Google_Flu_Trends&oldid=726837636
48. Harorli OT, Harorli H. Evaluation of internet search trends of some common oral problems, 2004 to 2014. *Community Dent Health.* sept 2014;31(3):188-92.
49. Murray G, O'Rourke C, Hogan J, Fenton JE. Detecting internet search activity for mouth cancer in Ireland. *Br J Oral Maxillofac Surg.* févr 2016;54(2):163-5.
50. Canceill T. Apport de l'outil Google Trends(r) dans les études épidémiologiques : exemple de l'amalgame [Internet] [exercice]. Université Toulouse III - Paul Sabatier; 2016 [cité 10 oct 2016]. Disponible sur: <http://thesesante.ups-tlse.fr/1287/>
51. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The Use of Google Trends in Health Care Research: A Systematic Review. *PLoS ONE* [Internet]. 22 oct 2014 [cité 6 sept 2016];9(10). Disponible sur: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215636/>
52. Butler D. When Google got flu wrong. *Nat News.* 14 févr 2013;494(7436):155.
53. Leinweber DJ. Stupid Data Miner Tricks: Overfitting the S&P 500. *J Invest. Spring* 2007;16(1):15-22.
54. Google arrête de prévoir (mal) les épidémies de grippe [Internet]. *Sciences et avenir.* [cité 1 oct 2016]. Disponible sur: <http://www.sciencesetavenir.fr/sante/e-sante/20150908.OBS5459/google-arrete-de-prevoir-mal-les-epidemies-de-grippe.html>
55. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science.* 14 mars 2014;343(6176):1203-5.
56. Preis T, Moat HS. Adaptive nowcasting of influenza outbreaks using Google searches. *R Soc Open Sci* [Internet]. 29 oct 2014 [cité 2 oct 2016];1(2). Disponible sur: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4448892/>
57. Bourne PE. Confronting the ethical challenges of big data in public health. *PLoS Comput Biol.* févr 2015;11(2):e1004073.

58. WHO | The objectives of the WHO Global Oral Health Programme (ORH) [Internet]. WHO. [cité 20 mai 2016]. Disponible sur: http://www.who.int/oral_health/objectives/en/
59. Google Traduction. In: Wikipédia [Internet]. 2016 [cité 2 nov 2016]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Google_Traduction&oldid=131195280
60. Google Traduction [Internet]. [cité 2 nov 2016]. Disponible sur: <https://translate.google.fr/>
61. Google Trends - қарые [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=%D0%BA%D0%B0%D1%80%D1%8B%D0%B5>
62. Google Trends-cari [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=cari>
63. Google Trends-karijes [Internet]. Disponible sur: <https://www.google.fr/trends/explore?q=karijes>
64. 10 Most Influential Languages in the World [Internet]. ALTA Language Services. 2008 [cité 4 déc 2016]. Disponible sur: <https://www.altalang.com/beyond-words/2008/06/05/10-most-influential-languages-in-the-world/>
65. Sbeyrand. Top 5 des langues les plus parlées dans le monde [Internet]. kiWiloo. 2015 [cité 4 déc 2016]. Disponible sur: <https://www.kiwiloo.fr/fr/top-5-des-langues-les-plus-parlees-dans-le-monde/>
66. Google Images - kari [Internet]. [cité 13 janv 2017]. Disponible sur: https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbn=isch&q=kari
67. Google Images - cari [Internet]. [cité 13 janv 2017]. Disponible sur: https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbn=isch&q=cari
68. Google Images - қариец [Internet]. [cité 13 janv 2017]. Disponible sur: https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbn=isch&q=%D0%BA%D0%B0%D1%80%D0%B8%D0%B5%D1%81
69. Google Images - próchnica [Internet]. [cité 13 janv 2017]. Disponible sur: https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbn=isch&q=pr%C3%B3chnica
70. Google Images - suus [Internet]. [cité 13 janv 2017]. Disponible sur: https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbn=isch&q=suus
71. Google Images - kario [Internet]. [cité 13 janv 2017]. Disponible sur: <https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox->

b&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbm=isch&q=kario

72. Google Images - 虫歯 [Internet]. [cité 13 janv 2017]. Disponible sur:
https://www.google.fr/search?q=r%C3%A9partition+langue+monde&client=firefox-b&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiysJLU9b3RAhVEyRQKHbHvDIIQ_AUICCGB#tbm=isch&q=%E8%99%AB%E6%AD%AF
73. Google Trends - 虫歯 [Internet]. Disponible sur:
<https://www.google.fr/trends/explore?q=%E8%99%AB%E6%AD%AF>
74. Google Trends - carie + caries + dentalcavity + carie + 虫歯 + الأسنانن خر + próchnica [Internet]. Google Trends. [cité 13 janv 2017]. Disponible sur:
<https://www.google.fr/trends/explore?q=carie%20%2B%20caries%20%2B%20dentalcavity%20%2B%20c%C3%A1rie%20%2B%20%E8%99%AB%E6%AD%AF%20%2B%20%D9%86%D8%AE%D8%B1%20%D8%A7%D9%84%D8%A3%D8%B3%D9%86%D8%A7%D9%86.%20%2B%20pr%C3%B3chnica>
75. Cauvin C., Reymond H., Serradj A. Discrétisation et représentation cartographique. Montpellier: GIP RECLUS; 1987. (Reclus modes d'emploi).
76. Institut de Recherche pour le Développement. Avantages et inconvénients de la méthode de Jenks.
77. Butty P. Liste brute des 1426 catégories de Google Trends [Internet]. Pierrick Butty. 2014 [cité 4 sept 2016]. Disponible sur: <http://www.pierrick-buty.com/liste-brute-google-trends/>

**Annexe 1 : exemple de catégories et sous-catégories disponibles
dans Google® Trends (77)**

Santé

Abus de substances illégales ou toxiques
 Dépistage de l'alcoolisme et de la toxicomanie
 Stéroïdes et produits dopants
 Tabac et arrêt du tabac
 Traitement de l'alcoolisme et de la toxicomanie
 Actualités de la santé
 Politique sanitaire
 Education sanitaire et formation médicale
 Equipements et appareils médicaux
 Technologies assistives
 Equipement et accessoires de mobilité
 Fondations pour la santé et la recherche médicale
 Littérature et ressources médicales
 Illustrations et photos médicales
 Maladies
 Allergies
 Arthrite
 Blessures
 Cancer
 Diabète
 Gestion de la douleur
 Maux de tête et migraines
 Maladies de la peau
 Maladies endocriniennes
 Diabète
 Maladies thyroïdiennes
 Maladies génétiques
 Maladies infectieuses
 Maladies sexuellement transmissibles
 SIDA et VIH
 Parasites et parasitoses
 Rhume et grippe
 Vaccins et immunisations
 Maladies respiratoires
 Asthme
 Obésité
 ORL

Santé

Rhume et grippe
 SIDA et VIH
 Troubles alimentaires
 Troubles du sommeil
 Troubles neurologiques
 Maladie d'Alzheimer
 Médecines douces et naturelles
 Acupuncture et médecine chinoise
 Nettoyage du corps et désintoxication
 Nutrition
 Régimes spéciaux et restrictifs
 Problèmes de cholestérol
 Vitamines et compléments
 Ophtalmologie
 Lunettes et lentilles de contact
 Pédiatrie
 Personnel infirmier
 Aide à domicile et soins de longue durée
 Pharmacie
 Traitements et médicaments
 Santé féminine
 Gynécologue obstétricien
 Grossesse et maternité
 Santé génésique
 Amélioration des performances sexuelles
 Contraception
 Education et conseils sexuels
 Gynécologue obstétricien
 Grossesse et maternité
 Maladies sexuellement transmissibles
 SIDA et VIH
 Stérilité
 Troubles de l'érection
 Santé masculine
 Troubles de l'érection

Santé

Santé mentale

- Dépression
- Difficultés d'apprentissage et troubles du développement
 - Trouble déficitaire de l'attention et hyperactivité
- Stress et anxiété

Santé publique

- Empoisonnements et overdoses
- Politique sanitaire
- Santé et sécurité au travail
- Vaccins et immunisations

Services et établissements médicaux

- Cabinets médicaux
- Hôpitaux et cliniques
- Physiothérapie
- Procédures médicales
 - Chirurgie
 - Chirurgie esthétique
 - Tests et examens médicaux
 - Vaccins et immunisations

Soins bucco-dentaires

- Vieillesse et soins gériatriques

- Maladie d'Alzheimer

Annexe 3 : liste finale des termes retenus pour la constitution de l'échantillon et raisons de rejets des autres termes

<u>Liste finale : échantillon B</u>	<u>Niches linguistiques</u>	<u>Pas de données renvoyées dans Google Trends</u>	
Carie	Сари	кариес	кариес
Caries	Cari	龋	கறிசெத்த
Dental caries	Kari	kariva	கரியோலம்
Dental cavity	Cario	кариоза	Yenice
Dental cavities		अधिर कुलगा	chung suong
نخر الأسنان		кариес	龋
Caries		ಕರೀಸ	கரியா
Karies		龋	
Kariya		카리에스	
Kariya		Кариес	
Kariya		కరీయ	
龋		కరీయ	
虫歯		కరీయ	
zubni kaz		кариес	
龋		కరీయ	
кариес		кариес	
prochnica		кариес	

Annexe 4 : tableau complet des pays et de leur popularité en fonction des requêtes

Liste des pays et popularité

1/9

Région	Pool	Pool+dental caries	Pool+dental cavities	Pool+cariès	Pool+karies
Afrique du Sud	12	12	12	12	12
Algérie	22	22	22	23	22
Allemagne	5	6	5	5	40
Arabie saoudite	9	9	9	9	9
Argentine	43	44	43	44	43
Australie	15	15	15	15	15
Autriche					35
Belgique	17	17	16	17	17
Biélorussie					
Bolivie	72	71	69	71	71
Brésil	57	58	57	58	58
Bulgarie					14
Canada	24	24	25	25	25
Chili	77	77	76	78	77
Chine					
Colombie	65	64	64	65	64
Corée du Sud					
Costa Rica	48	46	47	50	46
Croatie					
Danemark	10	10	10	10	12
Égypte	12	12	12	13	12
El Salvador	60	58	59	60	60
Émirats arabes unis	13	13	13	13	14
Équateur	70	70	70	70	69
Espagne	38	38	38	38	38

 popularité maximale

Liste des pays et popularité

2/9

Région	Pool+karijes	Pool+кариес	Pool+càries	Pool+龋	Pool+虫歯
Afrique du Sud	12	12	12	12	6
Algérie	22	23	21	23	11
Allemagne	6	6	5	5	3
Arabie saoudite	9	9	9	9	5
Argentine	43	44	43	44	22
Australie	15	15	14	15	8
Autriche					
Belgique	17	17	16	17	8
Biélorussie		34			
Bolivie	71	72	71	71	36
Brésil	58	58	56	58	28
Bulgarie		25			
Canada	25	25	24	25	12
Chili	77	77	76	77	38
Chine				7	
Colombie	64	66	64	65	32
Corée du Sud					
Costa Rica	47	46	47	45	24
Croatie	27				
Danemark	10	10	10	10	5
Égypte	12	13	12	12	6
El Salvador	62	63	62	62	30
Émirats arabes unis	13	13	13	13	6
Équateur	70	70	69	70	35
Espagne	38	38	38	38	19



popularité maximale

Liste des pays et popularité
3/9

Région	Pool+próchnica	Pool+zubní kaz	Pool+ฟันผุ	Pool+кариес
Afrique du Sud	13	12	12	12
Algérie	23	22	22	22
Allemagne	6	5	6	5
Arabie saoudite	9	9	9	9
Argentine	43	43	43	40
Australie	15	15	15	13
Autriche				
Belgique	17	16	17	16
Biélorussie				
Bolivie	71	70	72	72
Brésil	57	57	58	50
Bulgarie				8
Canada	25	25	25	22
Chili	78	76	78	73
Chine				2
Colombie	64	65	66	61
Corée du Sud				6
Costa Rica	47	47	46	48
Croatie				
Danemark	10	10		9
Égypte	12	12	13	11
El Salvador	60	61	60	64
Émirats arabes unis	13	13	13	12
Équateur	69	70	70	70
Espagne	38	38	38	34



popularité maximale

Liste des pays et popularité
4/9

	Pool	Pool+dental caries	Pool+dental cavities	Pool+cariës	Pool+karies
États-Unis	21	20	22	21	20
Finlande					20
France	34	34	34	34	34
Guatemala	58	56	56	57	56
Honduras	82	83	81	81	85
Hong Kong					
Hongrie	7	8	8	8	8
Inde	18	18	18	18	18
Indonésie	18	18	18	18	39
Iran					
Irlande	12	12	12	12	12
Italie	41	41	41	41	41
Japon	2	2	2	2	2
Kazakhstan					
Malaisie	13	13	13	14	16
Maroc	19	19	20	20	19
Mexique	75	74	75	74	75
Nicaragua					
Nigéria	29	30	30	29	30
Norvège					13
Nouvelle-Zélande	13	14	14	14	14
Pakistan	16	16	16	16	16
Panama	63	59	58	63	61
Paraguay	93	90	90	93	91
Pays-Bas	12	12	12	14	13
Pérou	100	100	100	100	100

 popularité maximale

Liste des pays et popularité

5/9

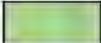
	Pool+karijes	Pool+кариес	Pool+càries	Pool+龋	Pool+虫歯
États-Unis	20	21	20	21	10
Finlande					
France	35	34	34	35	17
Guatemala	56	57	57	57	28
Honduras	82	84	80	83	
Hong Kong					
Hongrie	7	7	8	7	4
Inde	18	18	18	18	9
Indonésie	18	18	18	18	9
Iran					
Irlande	11	12	12	12	6
Italie	42	41	41	41	20
Japon	2	2	2	2	100
Kazakhstan		40			
Malaisie	14	13	13	13	7
Maroc	19	19	20	20	10
Mexique	76	76	75	75	37
Nicaragua					
Nigéria	30	30	29	29	15
Norvège					
Nouvelle-Zélande	14	14	13	13	7
Pakistan	16	16	16	16	8
Panama	60	62	60	62	29
Paraguay	94	94	88	96	46
Pays-Bas	12	12	12	12	6
Pérou	100	100	100	100	49

 popularité maximale

Liste des pays et popularité

6/9

	Pool+próchnica	Pool+zubní kaz	Pool+ฟันผุ	Pool+кариес
États-Unis	21	21	21	16
Finlande				
France	34	35	34	30
Guatemala	58	58	58	58
Honduras	81	80	83	88
Hong Kong				9
Hongrie	8	8	8	7
Inde	18	18	18	16
Indonésie	18	18	18	
Iran				6
Irlande	12	12	12	11
Italie	42	41	42	37
Japon	2	2	2	2
Kazakhstan				
Malaisie	14	14	14	13
Maroc	20	19	20	19
Mexique	75	75	76	67
Nicaragua				73
Nigéria	30	30	30	30
Norvège				
Nouvelle-Zélande	13	14	14	13
Pakistan	16	16	17	15
Panama	61	60	63	64
Paraguay	91	94	91	100
Pays-Bas	12	12	12	11
Pérou	100	100	100	95

 popularité maximale

Liste des pays et popularité

7/9

	Pool	Pool+dental caries	Pool+dental cavities	Pool+cariës	Pool+karies
Philippines	17	17	17	17	17
Pologne	4	5	4	4	5
Porto Rico	42	45	43	44	44
Portugal	33	32	32	32	32
République dominicaine	64	65	66	66	63
République tchèque					
Roumanie	16	16	16	16	16
Royaume-Uni	14	14	14	14	14
Russie					
Serbie					
Singapour	11	10	11	11	11
Slovénie					26
Suède	5	5	6	5	23
Suisse	14	14	14	14	30
Taiwan					
Thaïlande	7	6	6	7	7
Tunisie	27	27	26	26	27
Turquie	2	2	2	2	2
Ukraine					
Uruguay	46	50	47	50	46
Venezuela	92	92	92	92	91

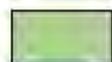


popularité maximale

Liste des pays et popularité

8/9

	Pool+karijes	Pool+кариес	Pool+càries	Pool+齲	Pool+虫齒
Philippines	17	17	17	17	9
Pologne	5	5	5	5	2
Porto Rico	45	45	45	44	22
Portugal	32	32	33	33	16
République dominicaine	64	65	63	63	32
République tchèque					
Roumanie	16	16	16	15	8
Royaume-Uni	14	14	14	14	7
Russie		40			
Serbie	22				
Singapour	11	11	11	11	6
Slovénie					
Suède	5	6	5	5	3
Suisse	14	14	14	14	7
Taiwan					
Thaïlande	7	7	6	7	4
Tunisie		28	26	26	14
Turquie	2	2	2	2	1
Ukraine		26			
Uruguay	48	51	48	48	25
Venezuela	93	93	91	93	46



popularité maximale

Liste des pays et popularité

9/9

	Pool+próchnica	Pool+zubní kaz	Pool+ฟันผุ	Pool+кариес
Philippines	17	17	17	17
Pologne	28	5	5	4
Porto Ricó	44	44	45	45
Portugal	31	32	32	31
République dominicaine	63	63	64	65
République tchèque		24		8
Roumanie	16	16	16	15
Royaume-Uni	14	14	14	12
Russie				1
Serbie				
Singapour	11	11	11	10
Slovénie				
Suède	5	5	5	5
Suisse	14	14	14	14
Taiïwan				4
Thaïlande	7	6	70	6
Tunisie	27	27	27	26
Turquie	2	2	2	2
Ukraine				6
Uruguay	47	49	50	51
Venezuela	91	92	94	88

 popularité maximale

**INTERETS DU « BIG DATA » POUR L'ESTIMATION DE LA
PREVALENCE CARIEUSE MONDIALE : APPORT DE GOOGLE®
TRENDS**

RESUME:

Cette thèse propose d'évaluer l'apport que les données massives ou « Big Data » pourraient avoir dans la surveillance populationnelle de la santé bucco-dentaire. Google® Trends est un outil permettant de connaître comment sont effectuées les requêtes dans Google® par la population mondiale.

A l'heure de l'omniprésence d'Internet et devant l'enjeu de santé mondial que représente la carie dentaire, il est judicieux de savoir si l'utilisation de cet outil moderne d'analyses des données massives peut devenir un complément des méthodes d'observation épidémiologique à moindre coût.

Une cartographie des requêtes mondiales sur le terme « carie », traduit en 17 langues, montre une corrélation intéressante entre la préoccupation des populations face à cette maladie et la prévalence estimée par les études de l'OMS.

TITRE EN ANGLAIS :Big data analyses in the survey of caries prevalence :
interests of Google® Trends

DISCIPLINE ADMINISTRATIVE : Chirurgie dentaire

MOTS-CLES :big data, carie, Google®, épidémiologie, prévalence mondiale

INTITULE ET ADRESSE DE L'UFR OU DU LABORATOIRE :
Université Toulouse III-Paul Sabatier
Faculté de chirurgie dentaire 3 chemin des Maraîchers 31062 Toulouse Cedex

Directeur de thèse : Jean-Noel VERGNES