Le projet MultiDES : premières réflexions sur la synonymie et étude de cas de synonymie translangue

Chris Chris A. Smith

▶ **To cite this version:**

Chris Chris A. Smith. Le projet MultiDES : premières réflexions sur la synonymie et étude de cas de synonymie translangue. Syntaxe et Sémantique, 2020, Synonymie, polysémie et questions de sémantique lexicale, 1 (21), pp.55-86. 10.3917/ss.021.0055 . hal-03158928

HAL Id: hal-03158928
https://hal.science/hal-03158928v1

Submitted on 12 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le projet MultiDES : premières réflexions sur la synonymie et étude de cas de synonymie translangue

Chris Smith

Université de Caen Normandie
Centre de recherches inter-langues sur la signification en contexte (CRISCO)
chris.smith@unicaen.fr

Résumé : Cet article propose une étude de cas de synonymie translangue, dans le cadre du projet d'extension à l'anglais du *Dictionnaire électronique des synonymes* (*DES*) du Centre de recherches inter-langues sur la signification en contexte (CRISCO). Le choix d'une paire d'adjectifs de couleur *vert* et *green*, ayant un sens primaire universel, permet de faire un premier test des espaces sémantiques combinés des deux adjectifs vedettes. Cette étude sera donc construite en quatre parties. La première partie fait le point sur le *DES* existant et le projet MultiDES. Une deuxième partie se concentre sur la synonymie comme accès au sens lexical. La troisième partie introduit l'étude du cas de la synonymie translangue à partir des adjectifs de couleur *vert* et *green*. Les synonymes de l'anglais sont générés et les relations synonymiques calculées et représentées dans un graphe d'adjacence combiné afin de visualiser les zones de recoupement interlangue ainsi que les zones de disparité. Un espace sémantique est alors proposé pour visualiser comment celui-ci représente les distances sémantiques entre les synonymes translangues, confirmant les spécialisations sémantiques distinctes des lexèmes VERT et GREEN. Enfin, dans une quatrième partie, les résultats obtenus sont mis à l'épreuve sur un corpus, à l'aide d'une analyse sémantique distributionnelle. Une étude des collexèmes des adjectifs *vert-vigoureux* et *green-vigorous* permet de tester si les proximités sémantiques sont équivalentes. Les résultats confirment l'existence d'importantes variations d'usage de paires de synonymes selon les corpus.

Abstract: *This paper provides a case study of crosslinguistic synonymy for the MultiDES project, which aims to extend the French* Dictionnaire électronique des synonymes *(*DES*) to English. The synonyms selected for study are the adjectives of colour* vert *and* green*, which present the advantage of being universal primitive adjectives, and therefore being relatively unproblematic in terms of translation. The paper is divided into four parts: the first section provides an overview of the existing* DES *dictionary and the MultiDES project. The second section focuses on the issues surrounding synonymy and the methods of calculating synonymy relations. The thirst section presents the data, the synonyms of* VERT *in French and the synonyms of* GREEN *in English, as well as the distribution of their senses across synonyms. A visualisation of the combined semantic spaces provides confirmation*

of the different uses of both adjectives in English and French. In a final fourth section, the results are put to the test using a distributional semantic analysis of two synonym relations, green-vigorous *and the corresponding* vert-vigoureux*, with the aim to verify the collocate scores of each pair. The results, unsurprisingly, show variations in the usage of synonym pairs in a corpus.*

1. Introduction

Le *Dictionnaire électronique des synonymes* (désormais *DES*) est une base de données de synonymes du français élaborée dans les années 1990. Il est à l'origine le produit de la numérisation des synonymes issus de sept dictionnaires compilés par l'INALF (Institut national de la langue française) pour l'élaboration du *Trésor de la langue française*. La base de données est construite à partir de deux dictionnaires du français du XX^{e} siècle, le *Grand Larousse de la langue française* (Guilbert *et al.* 1971-1989) et *Le Grand Robert de la langue française* (Rey 1985), ainsi que cinq dictionnaires de synonymes des XIX^{e} et XX^{e} siècles : Guizot (1809), Lafaye (1858), Bénac (1956), Bailly (1947), et Bertaud du Chazaud (1979) ; voir François *et al.* 2005. Comme le souligne Doualan (2013a), ces sept dictionnaires sont majoritairement distinctifs (ils donnent des indications sur les emplois différenciés des synonymes), à l'exception du plus récent dictionnaire des synonymes de 1989 qui est cumulatif (il ne contient aucune information sur les distinctions d'emploi), tout comme le *DES* lui-même.

Selon les concepteurs du *DES*, Sabine Ploux et Bernard Victorri (1998 : 162 ; nous soulignons), un synonyme peut être défini comme une forme lexicale substituable à un mot vedette :

> […] deux unités lexicales sont en relation de synonymie si toute occurrence de l'une peut être remplacée par une occurrence de l'autre *dans un certain nombre d'environnements* sans modifier notablement le sens de l'énoncé dans lequel elle se trouve.

Dans l'absolu, si un contexte de substituabilité existe entre deux lexèmes, le lien synonymique est acquis[1], même si la définition et les paramètres de la synonymie sont loin d'être consensuels (voir Masseron 2009

1. Pourtant cette approche hors contexte fondée sur des données lexicographiques, qui elles-mêmes envisagent des contextes possibles, représente une dichotomie fondamentale. L'opposition entre données paradigmatiques (langue) et données syntagmatiques (discours) est factice car toute ressource lexicographique prend désormais en compte des facteurs d'usage et de fréquence d'usage.

sur les « paradoxes » de la synonymie). On retombe ici sur le débat autour de la définition de la synonymie comme relation d'identité de sens, ou comme relation d'équivalence de sens, et ainsi de substitution partielle comme le décrit Doualan (2013b : 16). Comme le souligne Kleiber (2009 : 23), on sait que « l'interchangeabilité totale n'est pas possible et donc ne peut pas constituer un argument pour rejeter la synonymie-"identité de sens" ». L'absence de la synonymie « parfaite » ne signifie pas pour autant que la synonymie n'est pas une relation lexicale légitime (voir Cruse 1986 ; Hüllen 2004 ; Desagulier 2014 ; Murphy 2016). Spärck Jones (1986) avait déjà établi que la synonymie en traitement automatique devait passer par un contexte de substituabilité qui n'affecte pas l'interprétation[2]. On peut noter toutefois que le statut linguistique des liens synonymiques établis dans le *DES* demeure problématique dans la mesure où les liens synonymiques sont extraits de ressources lexicographiques variées dans le temps et sans accès aux contextes de jugement de synonymie. Les jugements de synonymie peuvent être plus ou moins inclusifs selon la définition que l'on se donne (contexte unique, contexte rare, ou plusieurs contextes, contextes fréquents).

L'objectif de cet article est de proposer une étude de cas de synonymie translangue entre le français et l'anglais, dans le cadre du projet MultiDES, qui repose sur l'extension du *DES* à la langue anglaise, en combinant les synonymes des deux langues. Il s'agit ici de générer les synonymes de deux vedettes bilangues, l'adjectif VERT et son équivalent anglais GREEN[3]. On prend en compte deux vedettes non problématiques dans leur sens propre littéral, des adjectifs de couleur au sens universel. Nous allons donc traiter ici d'un exemple de synonymie translangue : il sera question de discuter des résultats obtenus, des problèmes et difficultés rencontrés dans cette approche, qui consiste à intégrer les synonymes de l'anglais aux réseaux synonymiques du français. Au vu de la prolifération des ressources numériques variées et inégales sur les synonymes, comment faut-il procéder pour mettre en place la génération de synonymes anglais ? Plusieurs décisions méthodologiques se présentent pour faire face à la multiplication des synonymes et pour assurer une cohérence dans l'ajout des synonymes anglais.

2. « *Two or more word-uses are synonymous, and therefore form a row, if their word-signs are mutually replaceable at some position in some sentence, without any change in the ploy of the sentence* » (Spärck Jones 1986 : 71).
3. Nous utilisons les petites capitales pour les vedettes (point de départ de l'analyse de synonymie), la référence aux lexèmes anglais et français sera en italique, et la référence aux sens sera indiquée entre chevrons (<sens>).

Dans une première partie (section 2), nous faisons un tour d'horizon du *DES* existant ; puis, dans une deuxième partie (section 3), nous faisons le point sur la synonymie comme accès au sens lexical. Dans un troisième temps (section 4), nous introduisons une étude de cas de la synonymie translangue à partir des adjectifs de couleur VERT et GREEN, et enfin, dans une quatrième partie (section 5), nous proposons un test de la synonymie de *vert-vigoureux* et *green-vigorous* à partir d'une étude des collexèmes[4] (les cooccurrences de lexèmes dans le contexte) comparés des deux paires d'adjectifs en français et en anglais.

2. La base de données du *DES* et son extension à l'anglais (le projet MultiDES)

2.1. Dictionnaires et thesaurus en ligne

Le projet MultiDES consiste à étendre le dictionnaire de synonymes *DES* à la langue anglaise, en augmentant la base de données avec des mots de l'anglais. L'objectif est de tester la possibilité d'un espace sémantique translangue[5]. Une phase test a permis de générer les synonymes de l'anglais à partir de quelques dizaines de mots vedettes dans le *DES* actuel tels que MIROIR, PORTE, VERT, INTÉRESSANT.

Les synonymes résultent d'une compilation de ressources lexicographiques, comme pour le *DES*, sauf que cette compilation a été manuelle pour cette phase test. Une autre distinction est la grande différence des sources : la base de données du *DES* est fondée sur la numérisation de dictionnaires papier de synonymes des XIX^{e} et XX^{e} siècles reflétant des états de langue et d'usage qui sont anciens, alors que la base de données des synonymes anglais a été générée en s'appuyant sur des ressources lexicographiques électroniques fondées elles-mêmes sur des corpus contemporains. La génération des synonymes anglais de *green* s'est construite sur les thesaurus *Collins*, *Merriam-Webster* et *Oxford Living Dictionaries* (désormais *Lexico*). Toutes ces ressources ont la caractéristique d'être consultables en ligne en accès libre, sont mises à jour régulièrement, et sont souvent adossées à des informations contextuelles d'emploi. Il existe ainsi de multiples ressources de synonymie et de mots cooccurrents parmi les sources lexicographiques

4. Terme emprunté à l'anglais, qui provient de Stefanowitsch & Gries 2003 : 226 (« *we refer to words significantly associated with a particular slot in a construction as collexemes* ») et Stefanowitsch & Gries 2004.
5. Cette entreprise a été menée entre le français et le mandarin par Gaillard *et al.* 2010b, et entre le français et l'anglais par Ploux *et al.* 2010 (l'*Atlas sémantique*).

électroniques. À la grande différence des conditions de création du *DES* en 1990, nous disposons actuellement de multiples ressources électroniques contemporaines de mots synonymes. Ce foisonnement est à la fois une aubaine, mais également une difficulté, puisque le modèle de génération de synonymes (et la définition de la synonymie, comme nous l'avons évoqué précédemment) varie considérablement d'une ressource à l'autre. Par exemple, le thesaurus du *Merriam-Webster* est plus généreux dans l'inclusion de synonymes et ajoute à la catégorie synonymes et antonymes la catégorie de « *near antonyms* » et celle de « *related words* ».

Le protocole de génération manuelle des synonymes de l'anglais a ensuite également inclus la vérification de liens synonymiques translangues par un dictionnaire bilingue *Sensagent* et *Glosbe*, puis la vérification sur corpus pour des cas litigieux – Oxford English Corpus (OEC), British National Corpus (BNC) et Corpus of Contemporary American English (COCA).

2.2. Deux vedettes « équivalentes » parallèles

Des décisions méthodologiques ont ainsi dû être prises pour faire face à la multiplication des synonymes et pour assurer une cohérence dans l'ajout des synonymes anglais. Le premier choix méthodologique consiste à traduire le mot vedette puis à chercher les synonymes du mot vedette anglais équivalent, comme MIROIR, traduit par MIRROR. Cette méthodologie aboutit à une sélection en amont du fait de la traduction du français vers l'anglais d'un lexème. Or, un mot vedette a forcément plusieurs sens et choisir un seul terme est donc nécessairement sélectif. Notons qu'il s'agit bien ici de prendre comme point de départ un lexème et non un concept : il ne s'agit donc pas d'un thesaurus multilingue ou bilingue qui partirait d'un concept pour donner les multiples items lexicaux capables de désigner cette idée dans deux langues. La méthodologie qui a ainsi prévalu prend comme point de départ le mot vedette du français, pour sélectionner la traduction du sens propre de cette vedette, et générer le réseau de synonymes anglais du mot vedette traduit. Cela suppose bien évidemment que la traduction « équivalente » de ce mot vedette ne soit pas problématique. Une vérification des liens synonymiques français-anglais est ensuite assurée dans des dictionnaires bilingues, pour obtenir un réseau translangue qui serait capable de montrer les zones de recoupement des relations de sens mais aussi les zones de divergence.

Nous allons traiter ici d'un exemple de synonymie translangue, afin de discuter des résultats obtenus, des problèmes et difficultés rencontrés dans cette approche potentiellement controversée ou polémique, qui

consiste à intégrer les synonymes de l'anglais au réseau synonymique du français. Cet exemple traite d'une lexie de type primaire (perçue comme moins problématique que des concepts plus élaborés), correspondant à un adjectif de type universel comme *vert* et *green*, pour lequel un nombre important de sens métaphoriques et métonymiques forme des réseaux polysémiques.

2.3. Les difficultés inhérentes au DES *et son caractère hétérogène*

Le *DES* s'est constitué sur des relations synonymiques tirées de ressources des XIX^e^ et XX^e^ siècles (voir *supra*), et, de ce fait, la base de données du dictionnaire contient de nombreux lexèmes archaïques, datés, rares, argotiques, voire injurieux. En outre, la base de données du *DES* est désormais activement enrichie par une génération automatique de liens synonymiques, mais aussi par des examens ponctuels d'ajouts sur suggestion des utilisateurs, faisant du *DES* un outil collaboratif. De ce fait, la multiplicité des ressources de synonymes, et surtout le caractère diachronique de ces calculs de liens synonymiques, aboutissent à une base de données de relations synonymiques hétéroclites validées ponctuellement à des instants t s'échelonnant sur plus d'un siècle, du XIX^e^ au XIX^e^ siècle. Doualan (2011 : 18) souligne cette incompatibilité entre la relation synonymique (forcément synchrone) et l'état diachronique du *DES* :

> De ce fait, une relation de synonymie établie à un moment t n'est plus forcément valide au moment t+1, ce pas temporel étant le temps moyen que met un sens pour changer. On comprend dès lors que la synonymie est une relation purement synchronique. Elle n'a plus aucune légitimité en diachronie. Cela réduit considérablement la validité de la notion de synonymie et c'est valable pour toutes les relations basées sur le sens puisque les langues n'échappent pas au changement sémantique.

Ensuite, comme nous avons pu l'évoquer, les relations synonymiques du *DES*, cumulatif et minimaliste, sont fournies sans éléments de justification de la restriction ou non de la relation synonymique. Cela aboutit à deux effets potentiellement problématiques. D'une part, il n'y a aucune gradation de fréquence d'usage, ce qui signifie que les synonymes rares ou peu fréquents ont le même poids que des synonymes fréquents. D'autre part, les relations synonymiques sont perçues hors discours (hors contexte) mais mêlent en vérité des types de synonymies différentes, celle de « langue » et celle de « parole »,

selon le raisonnement de Cigada (2009)[6]. L'argument proposé est que « courageux » peut avoir comme synonyme « hardi » ou encore « lion » dans certains contextes, bien que ces deux synonymes ne soient pas équivalents, dans la mesure où « lion » n'est synonyme de « courageux » qu'en discours (dans un contexte spécifique et à un niveau de *granularité*[7] plus ou moins fin). Il est bien évident que cette substitution potentielle dans un contexte donné relève de la métaphore ou de la métonymie, mais l'argument de la substituabilité pose effectivement ce problème.

2.4. La question de l'usage et de l'approche lexicographique : pistes

Comment concilier l'existant du *DES* avec un projet d'extension, et le constat qu'une relation synonymique est validée à un moment t, dans un usage et un registre de langue précis ? La relation synonymique n'existe pas hors texte et hors contexte. Le projet MultiDES pourrait ainsi profiter, d'une part, d'une approche de corpus pour l'extraction de synonymes et la vérification des liens et, d'autre part, de l'intégration de dates dans une approche diachronique si on veut rester fidèle à l'existant. Cela aboutirait à un ambitieux projet d'inclusion de contextes d'usage, mais aussi de datations. Cela permettrait aussi au MultiDES de se différencier des autres ressources de thesaurus et synonymes déjà existants, mais demanderait un travail considérable. Comme le souligne Doualan (2011 : 84) :

> Mais comment pallier les diverses interrogations soulevées par la diachronie ? Il faudrait pouvoir apporter quantités d'indications temporelles aux entrées du DES et leurs synonymes. Mais est-ce seulement possible ? De même, un autre point touchant à la diachronie concerne l'enrichissement du DES avec les nouveaux lexèmes que compte la langue. Cela mettrait le DES en accord avec les dernières évolutions de la langue. Mais il ne faut pas oublier que cela ne ferait qu'ajouter des strates synchroniques à toutes celles déjà présentes dans le DES. Et pourtant, un des atouts du DES en tant que ressource électronique est justement de pouvoir combattre sa propre obsolescence à moindre coût.

---

6. « La synonymie / *langue* est en effet une synonymie de sémèmes. La synonymie / *parole* est une synonymie de sèmes uniques, qui se trouvent dans des sémèmes divers et qui ne sont souvent pas homologables » (Cigada 2009 : 22).
7. Voir Vetulani & Vetulani 2015 sur les problèmes de granularité de sens dans WordNet.

3. L'enjeu du *DES* : la synonymie comme moyen d'accès au sens lexical (à la polysémie)

3.1. Les espaces sémantiques de synonymes

Le *DES* repose sur un outil dont l'objectif est de capturer une représentation géométrique du sens. Cette entreprise menée par Ploux & Victorri (1998) consiste à appliquer la théorie des graphes au sens lexicographique. Il s'agit donc de produire une représentation géométrique du sens à partir de données lexicographiques, en langue et non en discours. Comme le précisent François *et al.* (2005 : 157), il n'est nullement question de sens contextuel : « Dans ces représentations, il n'est pas tenu compte du contexte discursif, seul le point de vue paradigmatique est donc en cause ». L'approche des concepteurs du *DES* a été de chercher à voir s'il est possible de capturer ce que Venant & Victorri (2012) appellent le noyau[8] de sens, une sorte d'invariant à partir duquel on expliquerait les effets de sens représentés par les glissements métasémiques, eux-mêmes représentés par les synonymes clés. Venant & Victorri (2012) ont observé la compatibilité de cette approche avec une vision cognitiviste du sens. Il ne s'agit pas d'une formalisation d'un sens schématique. Ce rapprochement entre perception cognitiviste du sens et représentation géométrique s'explique dans la mesure où la construction du sens reflète la construction conceptuelle.

Ce type de perspective paradigmatique se distingue nettement des travaux actuels visant à calculer le sens à travers une analyse de sémantique distributionnelle à partir de corpus (voir Glynn 2010[9]). La sémantique distributionnelle vise à analyser le sens d'une forme à travers ses relations avec les mots employés dans le contexte, appelés collexèmes, ou « *collocates* » en anglais. Un collexème correspond donc à un mot du contexte qui sert à analyser le comportement sémantique d'un lexème[10]. C'est une approche *usage-based* qui permet de déterminer non pas des noyaux de sens mais des profils comportementaux. Aujourd'hui, tout un pan disciplinaire travaille sur la notion et le calcul de la proximité sémantique à travers les collexèmes. Toutefois, cette

8. « Le noyau de sens est, en effet, ce qui assure la cohérence sémantique d'une unité lexicale, ce qui fait que l'on peut toujours passer, par des glissements sémantiques successifs, d'un sens à un autre. La synonymie est l'outil idéal pour capturer ce continuum car n'étant, presque toujours, que partielle, elle permet de représenter à la fois en quoi les sens de mots synonymes se recouvrent et en quoi ils diffèrent » (Venant & Victorri 2012 : 74).
9. « *The current state of the art divides into the study of lexical near-synonyms* […] *and syntactic alternations* […] » (Glynn 2010 : 90).
10. Voir Firth (1957), considéré comme le père de la sémantique distributionnelle.

proximité sémantique n'est pas la synonymie. Des synonymes ne sont pas des formes sémantiquement proches, mais des formes substituables en contexte. On s'accorde à accepter que la synonymie absolue (et donc l'équivalence parfaite) n'existe pas (Cruse 1986 ; Spärck Jones 1986 ; Kleiber 2009 ; Murphy 2016) car des distinctions existent toujours dans le choix d'une forme plutôt qu'une autre (distinctions purement sémantiques, ou discursives, pragmatiques, selon les connotations, les registres, etc.). La similarité distributionnelle n'est pas un bon indicateur de synonymie selon Budanitsky & Hirst (2006 : 42) car les relations sont tributaires du corpus, de sa taille, de sa représentativité, mais aussi parce qu'il n'y a aucune symétrie en sémantique distributionnelle (un mot peut être synonyme d'un autre en contexte mais ce n'est pas toujours le cas dans le sens inverse). De ce fait, Budanitsky & Hirst 2006 rejettent cette méthodologie comme modèle de rapprochement sémantique.

Au contraire, certains chercheurs soutiennent que la similarité distributionnelle est un indicateur de similarité lexicale (Kotlerman *et al.* 2010 ; Desagulier 2014), alors que d'autres vont jusqu'à affirmer que le sens hors contexte n'existe pas, même parmi les lexicographes, tels que Patrick Hanks, qui a participé à l'élaboration de l'un des premiers dictionnaires fondés sur corpus numérisé, le *Collins Cobuild*. Hanks (2013) développe la théorie des normes et exploitations, à partir de l'usage en contexte de lexèmes : un lexème aurait un profil comportemental en évolution constante, les normes correspondent aux emplois fréquents, alors que les exploitations s'appuient sur des potentiels sémantiques, idée que l'on retrouve dans le concept de *lexical priming*, ou « amorçage » de Hoey (2005) ; voir aussi Pace-Sigge (2013).

Nous tenterons dans une quatrième partie (section 5) de faire un test de sémantique distributionnelle pour explorer la possibilité de combiner les approches lexicographique et de corpus (qui sont devenues inextricables puisque les dictionnaires se fondent sur les données) dans le futur MultiDES.

3.2. Le calcul de la proximité sémantique : modèle des cliques des réseaux synonymiques

Il s'agit de représenter géométriquement l'organisation du sens des lexèmes, en appliquant la théorie des graphes aux structures lexicales d'une langue (Ploux & Victorri 1998). Comme le soulignent Venant & Victorri (2012 : 74) :

> Nous utilisons la synonymie pour représenter le sens d'une unité lexicale. Nous cherchons aussi bien à capturer son potentiel de sens

> hors contexte, qu'à caractériser le sens plus ou moins précis qu'elle prend en contexte.

Il s'agit de confronter formalisation et usage en étudiant les espaces sémantiques à différentes échelles (niveau local, niveau global, niveau intermédiaire). Le niveau intermédiaire pourrait :

> […] caractériser des relations sémantiques entre quelques unités sémantiques, comme, par exemple, le fait que *sec* et *froid* entretiennent une relation de synonymie très partielle, valable uniquement dans un contexte psychologique. *Sec* et *froid* peuvent être considérés comme étant synonymes dans *un sourire sec* et *un sourire froid*, mais pas *un temps sec* et *un temps froid*. (Venant & Victorri 2012 : 76)

Ainsi, si une simple liste de synonymes n'est pas distinctive, l'organisation des synonymes par cliques l'est davantage. Une clique se définit comme un sous-ensemble de synonymes de la vedette qui sont synonymes entre eux. À partir de ces cliques, la représentation géométrique permet de distinguer des sens différents et donc des usages lexicalisés différents (la polysémie de la vedette). La méthodologie d'analyse du graphe consiste à commencer par explorer les sens de « cœur », au centre du graphe (le sens primaire), et ensuite aller vers les cliques périphériques (souvent les sens spécialisés) pour expliquer les rapprochements entre les sens. Cette méthode s'est avérée fonctionnelle pour les adjectifs, mais l'exploration des sens verbaux est légèrement différente de celle des adjectifs (comme le sens formel de ces catégories en est distinct) selon Venant & Victorri (2012 : 85).

La validation des résultats des graphes de synonymie et de leur interprétation est possible par des méthodes expérimentales, notamment deux méthodes de validation mises en place communément par protocole : d'une part, des tests d'ordre psycholinguistique chez des locuteurs natifs pour déterminer si les synonymes sont validés et, d'autre part, une validation par analyse de corpus. Ces deux méthodes de validation des résultats de la visualisation géométrique de la synonymie lexicale sont toutefois coûteuses en temps, et n'ont donc pas pu être menées à grande échelle.

3.3. Synonymie conceptuelle et approche translangue

Une approche translangue ou multilangue ne peut se justifier que dans une approche conceptuelle, qui sera effectivement la nôtre. Si considérer la synonymie entre deux systèmes de langue semble improbable dans une perspective grammaticale formelle, où deux systèmes de langue ont leurs propriétés distinctives et ne se recoupent pas, en revanche,

dans une approche conceptuelle, comparer les termes exprimant des concepts similaires est parfaitement adapté, et en particulier concernant un concept universel comme celui des couleurs. Hirst (1995)[11] définit la synonymie de manière indirecte, en se concentrant sur la relation entre le concept et son expression linguistique langagière dans une langue donnée. Sur le plan sociolinguistique, la notion de synonymie permet de comparer et distinguer le réseau de formes linguistiques qui expriment un aspect du même concept.

Cette approche pourra précisément permettre de mieux comprendre les distinctions d'emplois entre termes sémantiquement proches d'une langue à l'autre. Dans une approche linguistique formelle, deux items lexicaux distincts, avec des étymologies distinctes, ont toutes les chances de présenter des différences importantes dans deux systèmes de langue distincts. Toutefois, si l'on considère qu'un synonyme est un recoupement de sens partiel, alors deux unités lexicales appartenant à des langues distinctes peuvent être synonymes sur le plan pragmatique. D'un point de vue cognitiviste, le projet de recherche de synonymes interlangues peut répondre à des questions ontologiques essentielles, en rapport avec le sens primitif. Cette entreprise a un intérêt indiscutable comme outil heuristique dans une perspective traductologique, comme le décrivent Edmonds & Hirst (2002 : 111) dans leur comparaison des distinctions de sens entre les noms *woods*, *forest* et *wald*. On peut donc imaginer qu'une analyse des synonymes translangues d'une paire de vedettes anglais-français permettra de mettre en évidence les différences d'extension polysémique des lexèmes *green* et *vert*, donnant ainsi accès à une visualisation de la différence de leurs usages, ou à leurs profils conceptuels.

4. Étude de synonymie translangue : les adjectifs VERT et GREEN

4.1. Génération des synonymes et organisation de la polysémie

Commençons tout d'abord par évoquer la question de la source de la génération des synonymes : commencer par un lexème exige qu'on

11. « *I thus sidestep the long-debated question of exactly what degree of inter-substitutability should count as synonymy or near-synonymy (Egan 1942 ; Spärck Jones 1986 ; Cruse 1986). For this work, synonymy and near-synonymy arise at the point in the conceptual network at which a (language-independent) concept diverges into the set of (language-dependent) lexical entries for the words that, in one way or another, denote that concept. That is, the groups of words among which we need to discriminate are exactly the groups of words (in each language of interest) that correspond to each single concept in the taxonomic hierarchy of the knowledge base of the system* » (Hirst 1995 : 53).

fasse un choix de langue de départ, ce qui dans notre cas est le français, puisque la base existante du *DES* est en français. Toutefois ce choix peut avoir des conséquences sur les sens représentés ou non dans le graphe et en particulier le risque d'exclusion de sens anglais.

Pour le mot vedette VERT (sans distinction de classe grammaticale), le *DES* donne 54 synonymes[12] qui se divisent en 47 cliques, c'est-à-dire des sous-ensembles de mots synonymes entre eux, signifiant une grande polysémie du lexème *vert*. Le dictionnaire en ligne *Synonymo.fr* donne 18 synonymes alors que *Le synonymeur* en donne 46[13]. On peut aussi s'étonner que le *Larousse langue française* en ligne donne 8 sens différents alors que le *DES* donne 47 cliques. Le *TLFi*[14] propose des sens plus complets en usage, hiérarchisés en fonction des glissements de sens et effets, connotations et registres. On peut remarquer que le sens « – *Pop., fam. Être vert*. Être pâle de déception, subir une grave déconvenue, tromperie » est bien donné, mais que les adjectifs *déçu* et *dépité* ne figurent pas parmi les synonymes dans le *DES*, ni même dans les autres ressources synonymiques du français telles que le *Larousse* et le *Robert* en ligne. Dans ce sens, il est clair que l'activation de l'acception <dépité> vient par métonymisation de cause à effet ; la couleur blafarde ou blême matérialise en quelque sorte l'émotion de déception. Tout comme en anglais *green* est associé à l'émotion de jalousie (*green with envy*) mais non dans le sens de déception. Voici, dans le tableau 1 ci-contre, les synonymes de VERT organisés en groupes de sens.

Il s'agit des synonymes de VERT (adjectif et nom) dans le *DES*, avec en parallèle les synonymes de l'anglais GREEN issus d'une compilation manuelle (37 synonymes de type adjectival). On note que le *DES* n'intègre pas de distinction de classe grammaticale dans sa base de données, ce qui explique que les synonymes de la vedette VERT mêlent à la fois des sens adjectivaux et des sens nominaux tels que <herbe, green>. Comme nous avons pour objectif de traiter du sens de l'adjectif *vert*, les sens nominaux de la vedette VERT en catégorie 6 du tableau 1

12. 54 synonymes selon une recherche menée en juillet 2019, 60 synonymes en février 2020.
13. Voir http://www.synonymo.fr/synonyme/vert ; https://www.synonymeur.com/synonyme/vert.
14. En ce qui concerne le *TLFi* : « La rédaction est terminée depuis 1994 et la plupart des contributrices et contributeurs ont quitté le laboratoire. Il n'a pas vocation à être mis à jour. Cette ressource, qui ne fait pas l'objet d'une veille lexicographique, est donc close "en l'état". Il est donc tout à fait naturel que les définitions qui s'y trouvent ne rendent pas compte des évolutions de la société » (http://www.atilf.fr/spip.php?article192).

Tableau 1. Classification des synonymes
de la vedette VERT avec les synonymes anglais en parallèle

Sens conceptuels	Synonymes français de VERT	Synonymes anglais de GREEN
1	acerbe, aigre, aigrelet, râpeux / raide, acidulé, rude / brutal, violent, vif	*unripe*, *raw*
2	écologique / éco-	*ecological*, *eco-friendly*, *ozone-friendly*, *non-polluting*, *recyclable*, *sustainable*, *conservationist*, *green collar*, *durable*
3	nouveau, jeune / juvénile, immature, novice / inexpert, bleu	*immature*, *youthful*, *young*, *wet behind the ears*, *uninitiated*, *unfledged*
4	hardi, mordant, herbeux, valide, vaillant, vigoureux, vif	*vigorous*, *verdurous*, *verdant / leafy*, *grassy*
5	gaillard, cru, gaulois, graveleux, osé, polisson, licencieux, égrillard / leste, libre	
6	herbe, prairie, green	*turf*, *lawn*
7	blême, pâle, glauque, bilieux	*pasty*, *pallid*, *wan*, *sallow*, *ashen*, *glaucous*, *pinched*
8	vert bouteille, pers, émeraude, pomme	*pea green*, *olive green*, *greenish*, *green*, *bottle green*, *jade*

seront exclus pour l'étude des synonymes adjectivaux translangues à suivre. Il s'agit du sens <herbe, gazon, prairie, green>. Il faudrait également s'interroger sur le statut adjectival ou non des lexèmes des sens 2 et 3 <écologiste, nouveau, inexpert et bleu> qui semblent correspondre à des relations synonymiques du nom VERT uniquement. Il est en vérité assez complexe de démêler emplois nominaux et emplois adjectivaux du fait de la substantivation aisée des adjectifs français. Il est manifeste que *vert* et *bleu* ne sont synonymes que dans un contexte particulier, dans la désignation d'une personne inexperte (*un bleu* et

un vert seraient ainsi des formes de désignations synonymes, mais les adjectifs *vert* et *bleu* ne seraient pas synonymes dans un emploi adjectival : « il est bleu / il est vert »). Dissocier l'emploi de la forme donne lieu à des résultats confus.

La précision (ou granularité) des regroupements est volontairement grossière pour permettre une clarté des regroupements métasémiques : on prend en compte les glissements de sens pour les regrouper de façon à améliorer la lisibilité à la manière des entrées lexicographiques. On obtient une vision de la polysémie de *vert* en français avec les synonymes, dans laquelle on peut introduire les synonymes de *green* en anglais qui correspondent aux mêmes effets de sens conceptuels.

Les synonymes de couleur se trouvent dans la dernière catégorie 8, exclue de l'étude, étant donné qu'il s'agit de décliner des types de couleurs, donc en quelque sorte des hyponymes de la couleur verte et non de véritables synonymes.

La génération des synonymes en anglais (voir *supra*) est issue de la consultation de plusieurs ressources numérisées qui ont été croisées pour retenir les synonymes avérés de GREEN dans *Merriam-Webster*, *Collins*, *Oxford Living* (*Lexico*). Toutes ces ressources ont la caractéristique d'être consultables en ligne et sont mises à jour régulièrement ; elles sont également toutes étoffées de recherche de synonymes et de thesaurus. Il y a donc 37 synonymes anglais adjectivaux uniquement correspondant aux sens de vert.

Le tableau 2 ci-contre présente la classification des synonymes du lexème adjectival GREEN de l'anglais uniquement, pour lesquels on obtient 6 catégories (si l'on exclut les adjectifs de couleur hyponymes en 8 et les sens nominaux en 7), contre 6 dans le cas de l'adjectif *vert* (une fois que l'on a exclu le sens nominal <herbe, gazon, prairie, green> et les sous-types de couleur et si on maintient le sens 3). On constate que, malgré un nombre similaire d'acceptions, le traitement lexicographique des adjectifs *green* et *vert* montre des divergences attendues, notamment l'absence en anglais du sens <brutal / vif> qui part du sens 4 <hardi, vigoureux>, et aussi de la métasémie <gaillard, grivois>. Il semblerait donc que la métasémie <vigueur> a donné plus de relations en français qu'en anglais.

Toutefois, si l'on peut organiser les sens de l'adjectif *green* en 6 catégories conceptuelles, on note une variation dans le traitement des synonymes au stade lexicographique. *Merriam-Webster Thesaurus* propose un regroupement en trois catégories des sens de GREEN : « *1) covered with a thick, healthy natural growth* […] *2) lacking in adult experience or maturity* […] *3) lacking in worldly wisdom or informed judgment* ».

Tableau 2. Classification des synonymes de GREEN en anglais

Sens conceptuels	Synonymes de GREEN
1	*inexperienced*, *unpolished*, *unsophisticated*, *untrained*, *novice*, *unversed* / *inexpert*
2	*ingenuous*, *innocent* / *ignorant*, *gullible*, *credulous*, *jejune*
3	*non-polluting*, *durable*, *sustainable*, *ecological*, *eco-friendly*
4	*unhealthy*, *sick*, *nauseous*, *under the weather*
5	*common*, *simple*, *naïve* / *naif*
6	*verdurous*, *grassy*, *vigorous*, *leafy*
7	*callow*, *foliage*, *greenery*, *leafage*, *vegetable*
8	*pea green*, *olive green*, *grassy*, *greenish*, *green*, *bottle green*, *jade*

D'autre part, dans un système de langue donné, on s'attend bien entendu à ce que les usages et répartitions des spécialisations sémantiques se fassent naturellement de manière distincte. Il est possible, au-delà des distinctions dictionnairiques, que la métasémie des lexèmes VERT et GREEN présente bien des divergences naturelles, et que leurs usages soient différenciés.

Cependant, l'objectif est ici de permettre de mieux comprendre les affinités sémantiques de VERT et GREEN, et de mieux visualiser la manière dont les extensions de sens peuvent diverger d'une langue à l'autre. Dans cette optique, les synonymes sont des outils permettant de mettre en évidence les distinctions, donc des outils de discrimination sémantique.

4.2. Polysémie comparée de VERT et GREEN

On observe une asymétrie des résultats et des regroupements sémantiques entre les deux langues.

Notamment le sens 5 en français n'est pas observé en anglais : cela signifie que cette métasémie est non existante pour GREEN. Le sens 2 français reste proche du sens propre alors que le sens 3 anglais fait apparaître des sens connexes par métonymie plus éloignés tels <*durable*, *sustainable*>. Le sens 4 anglais est également distinct du sens 7 français <blême / pâle> : le sens français renvoie à un état émotionnel alors que le

sens anglais semble renvoyer plutôt à la maladie, à un état physiologique et non émotif.

La matrice d'adjacence combinée de VERT et GREEN (voir la figure 1 ci-contre) confirme visuellement les impressions formulées à partir des listes et des regroupements de synonymes. Ce graphique représente les réseaux de synonymes translangues des adjectifs *vert* et *green*. Les synonymes sont représentés sous forme de bulles, dont la taille dépend du nombre de relations synonymiques dans lesquelles entre le mot. Le nombre de relations apparaît également chiffré à la suite du mot. On voit apparaître sur le graphe multilingue des regroupements de sens plus denses, avec parfois une absence de termes anglais ou français. Ce cloisonnement montre la spécificité d'un sens dans une langue par rapport à l'autre.

Nous excluons de ce graphe les sens nominaux de GREEN et VERT (<gazon, *turf*>) et les sous-types de couleur dans les deux langues (<*pea green*, *emerald*, émeraude, céladon>) puisqu'il s'agit de déclinaisons de types de couleur, donc de potentiels hyponymes et non de véritables synonymes. Nous avons laissé les sens nominaux / adjectivaux difficilement démêlables <écologiste, bleu, immature, nouveau>. Ils seront exclus par la suite car il s'agit de données problématiques.

On constate donc pour les synonymes de GREEN que le sens <inexpert> est majoritairement peuplé par des adjectifs descriptifs anglais, de même pour le sens <malade / nauséeux>. En français, cependant, les sens <inexpert / bleu> sont majoritairement des sens nominaux (désignant des humains) et non des sens adjectivaux.

En parallèle sur le réseau de synonymie de VERT, on voit apparaître une zone peuplée de termes anglais essentiellement avec le sens <écologique>, alors que le sens de <vigoureux, osé, gaillard / violent / acide> n'a pas d'équivalent anglais. On distingue bien les divergences sémantiques entre les réseaux de synonymie.

4.3. L'espace sémantique combiné : visualisation des relations synonymiques dans un espace sémantique

Notons tout d'abord que les sens nominaux <gazon, prairie, green, *ecologist* n, bleu, inexpert> ne sont pas exclus ici car le *DES* ne fait pas de distinction de classe grammaticale dans sa base de données.

La figure 2 ci-après permet une visualisation de l'espace sémantique combiné des vedettes GREEN et VERT. L'espace sémantique utilise le regroupement en cliques (groupes de mots entrant en relation synonymique les uns avec les autres) pour représenter géométriquement les différentes zones de sens. Autour des cliques on voit apparaître les enveloppes qui

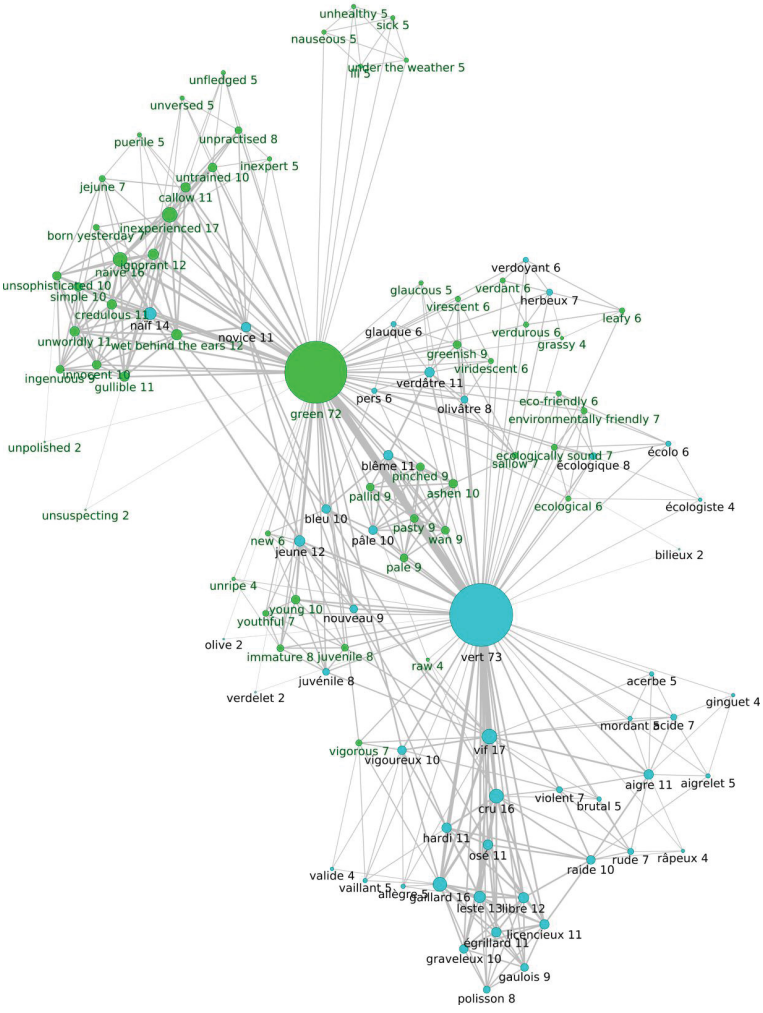

Figure 1. Graphe d'adjacence combiné de *green* et *vert*

représentent les cloisonnements des sens distincts. Il s'agit donc d'une visualisation très différente de la figure 1 (données brutes sans calcul).

On voit apparaître, à partir de cet espace sémantique combiné, trois grands espaces distincts qui représentent les trois principaux réseaux de sens de l'adjectif VERT. En haut à gauche, le plus éloigné de la vedette correspond au sens <écologique> : on constate que le sens anglais <*eco-friendly*, *sustainable*> se trouve relativement isolé, étant le plus éloigné des sens anglais, alors que, pour le français, les sens

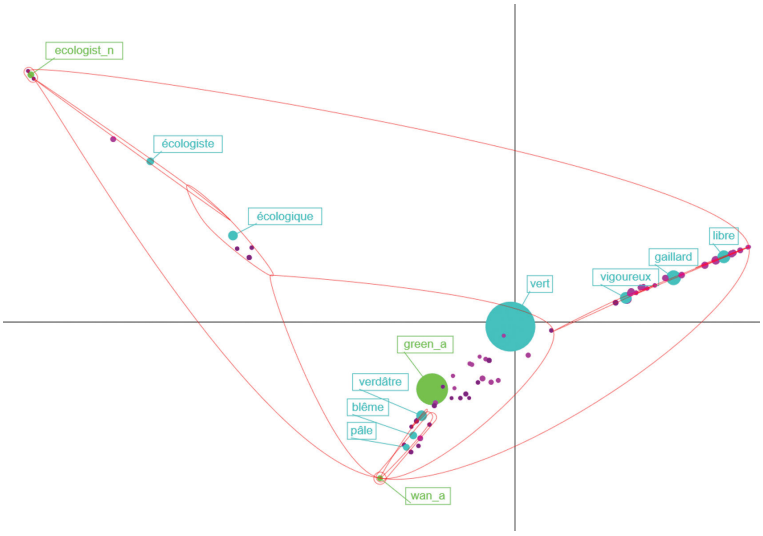

Figure 2. Espace sémantique combiné de GREEN et VERT

<écolo-, écologiste> (à la fois adjectival et nominal) et le sens adjectival <écologique> se trouvent être les plus éloignés de tous (ici <écologique> regroupé dans une clique avec d'autres synonymes).

Le deuxième sens en bas correspond au sens couleur de la vedette VERT, avec des références à la couleur verte, mais aussi à la pâleur, au teint de peau verdâtre qui trahit des émotions.

En ce qui concerne le troisième réseau de sens isolé dans l'espace sémantique, on voit le sens <vigoureux>, le plus proche de la vedette (vert, qui pousse, donc dynamique), une extension métaphorique du monde végétal, puis, un peu plus loin, le sens de <gaillard> (extension métaphorique pour désigner un tempérament, quelqu'un qui a un comportement osé), et enfin les sens de <libre, osé> (extension du sens de <gaillard> avec des connotations sexuelles), donc <grivois>, etc.

Le problème qui se pose ici est avant tout celui de la classe grammaticale à inclure : les sens nominaux et adjectivaux sont manifestement distincts, et la capacité de changement de classe est un facteur supplémentaire à inclure dans la base de données.

La version 3D (voir la figure 3 ci-contre) de l'espace sémantique permet de mieux distinguer les sens et leur position. Toutefois, le nombre de sens à représenter étant très important, le schéma 3D est difficile à lire sur une page en version 2D. Pour la réalisation du graphe et la clarté des relations, il est ainsi nécessaire de réduire le nombre de

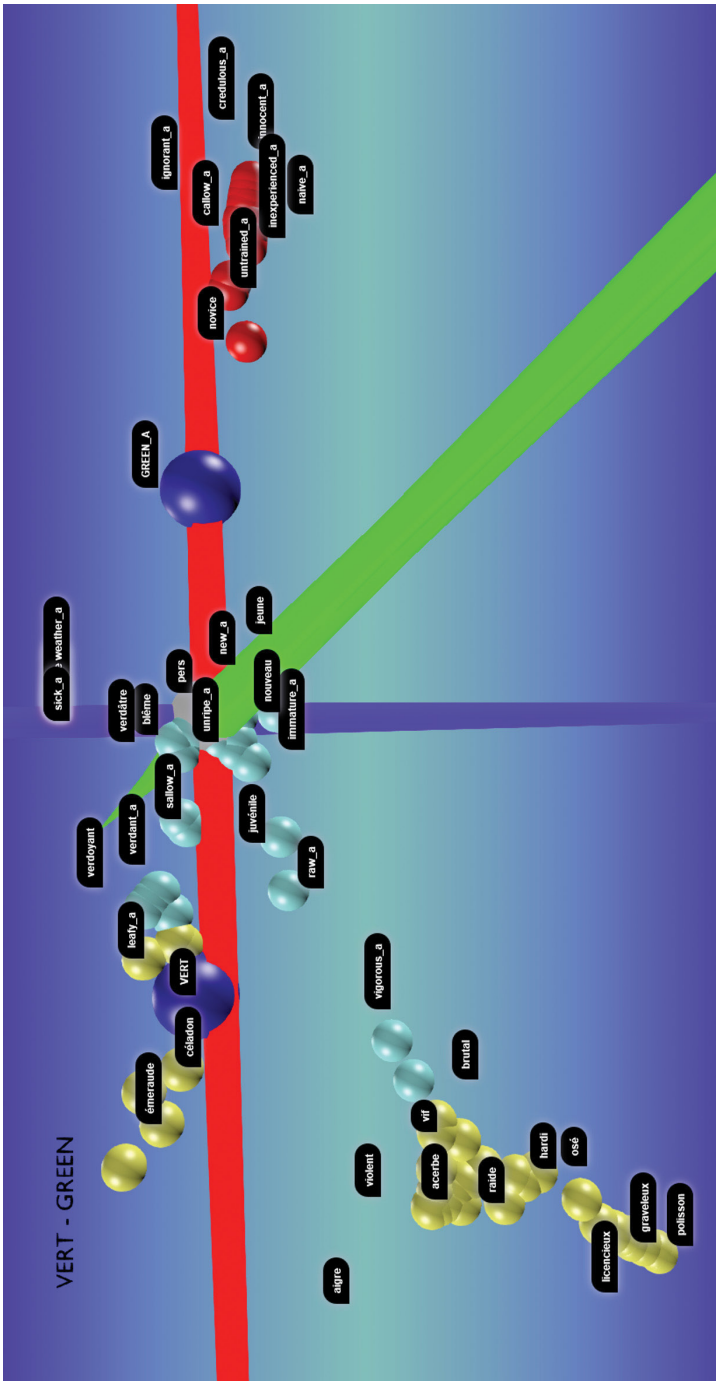

Figure 3. Espace sémantique 3D des deux vedettes GREEN et VERT

relations synonymiques, afin de mieux cerner où se situent les zones de recoupement entre les synonymes des deux vedettes GREEN et VERT qui sont représentées par les boules bleues. Les synonymes de GREEN sont représentés en rouge, les synonymes de VERT en jaune, et les synonymes de la zone commune sont représentés en bleu clair[15].

La sélection des sens à supprimer s'est faite à partir de deux critères : le faible nombre de relations synonymiques du mot et une fréquence moindre par rapport aux autres. Cela supprime ainsi, dans la zone de sens <naïf>, les relations avec *inexpert*, *unversed*, *jejune*, *unfledged*, *born yesterday*, *unpractised*, qui ont moins de dix relations synonymiques.

La question se pose toujours de manière renforcée de la relation entre usage en corpus et traitement lexicographique dans le *DES* initial. Il s'agit aussi de considérer la meilleure manière d'intégrer les synonymes de l'anglais afin de préserver la nature même du *DES*, mais également de rendre l'outil plus performant et plus en phase avec l'usage, potentiellement de façon diachronique si les relations synonymiques anciennes, archaïques, désuètes sont maintenues.

La section 5 vise à tester la relation synonymique *vert-vigoureux* et *green-vigorous* dans plusieurs corpus français et anglais. Comme les résultats préliminaires dans le MultiDES tendent à suggérer que cette relation existe dans les deux langues, mais sans indication de fréquence d'usage, interroger un corpus permettra de confirmer si l'usage est effectivement comparable ou non. On vise ainsi à confronter les résultats obtenus dans cette approche manuelle de traitement lexicographique des lexèmes GREEN et VERT à des données contextuelles de corpus.

5. Analyse comparative des collexèmes de *vert* et *green*

5.1. Objectifs du test

L'objectif ici est de confronter les résultats obtenus à travers les espaces sémantiques translangues anglais français avec le profil comportemental de GREEN et VERT dans des corpus. Nous allons donc faire un premier test pour valider la relation *vert-vigoureux* et la relation *green-vigorous* dans des corpus Web contemporains de grande taille :

15. La version 3D peut être visualisée *via* le lien suivant : https://crisco2.unicaen.fr/des3d/vert-green.html. Il est possible de faire tourner la vue manuellement en cliquant au centre du graphique et en faisant glisser la souris sur l'axe rouge ou sur l'axe bleu selon que l'on souhaite faire une rotation par rapport à l'axe des ordonnées ou des abscisses. Il est également possible d'activer une rotation automatique *via* les paramètres disponibles en bas à droite du graphique, en cliquant sur la roue crantée puis sur la flèche circulaire.

enTenTen15 (collection de textes Internet anglais de 15 milliards de mots) et frTenTen17 (collection de textes Internet français de 5,7 milliards de mots), tous deux accessibles *via* Sketch Engine[16]. On utilisera également, *via* Sketch Engine, OEC (Oxford English Corpus, corpus contemporain de genres multiples et aires géographiques variées, de 2,1 milliards de mots) pour tester l'effet du choix du corpus sur les résultats en anglais, et Frantext (textes littéraires du XVIII^{e}-XX^{e} siècle, de 15 millions de mots) pour interroger les résultats obtenus en français dans un corpus Web contemporain.

Il s'agit de lancer une analyse collocationnelle des lexèmes recherchés, en opérant une comparaison entre les collexèmes de chacun des mots cibles. Les résultats montreront la zone de recoupement des deux adjectifs dans chacun des corpus. On pourra alors constater, dans un corpus Web contemporain, si les relations sémantiques sont équivalentes ou non.

5.2. Méthodologie : sémantique distributionnelle comparée

On sélectionne l'un des synonymes clés de *vert* dans la matrice d'adjacence : prenons le sens <inexpérimenté>, que l'on retrouve pour chacune des vedettes GREEN et VERT. Il s'agit de mener une analyse de sémantique distributionnelle comparée entre les relations en utilisant la fonction Sketch (recherche de collexèmes triés par score de cooccurrence) dans Sketch Engine.

Le tableau 3 ci-après montre les résultats de la comparaison entre les collexèmes de *vert* et *vigoureux* dans le corpus Web frTenTen17. Les collexèmes dans la position *et / ou* sont proposés pour le lexème *vert* (en gris clair) et pour le lexème *vigoureux* (en gris foncé). En blanc apparaît la zone de recoupement, c'est-à-dire les emplois communs, l'espace sémantique partagé des lexèmes *vert* et *vigoureux*.

Les chiffres dans la première colonne indiquent la fréquence brute de la collocation avec *vert*, alors que la colonne 2 indique la fréquence brute de la collocation avec *vigoureux*. La colonne 3 donne l'indice de proximité de la collocation avec *vert* et la colonne 4 l'indice de proximité avec *vigoureux*. L'indice de proximité correspond au calcul pondéré du lien collocationnel entre les termes : le calcul permet de prendre en compte les fréquences d'emploi de chaque terme pour calculer si la proximité des termes se situe au-dessus du seuil de signification

16. Sketch Engine est une plateforme mettant à disposition de nombreux corpus multilingues ainsi qu'une interface et un outil performants d'analyse sémantique distributionnelle.

Tableau 3. Collexèmes communs de *vert* et *vigoureux* dans frTenTen17

Collexèmes	**Nombre d'occurrences en position *et / ou* avec *vert***	**Nombre d'occurrences en position *et / ou* avec *vigoureux***	**Score de collocation avec *vert***	**Score de collocation avec *vigoureux***
bleu	19 269	0	11.2	
rouge	17 058	0	10.8	
jaune	10 590	0	10.6	
rose	2 935	0	9.1	
violet	2 170	0	8.8	
blanc	7 725	13	9.2	0.5
sain	135	319	4.6	7.4
fertile	74	111	4.1	7.9
précoce	6	66	0.5	7.1
productif	14	315	1.6	8.8
grimpant	0	47		7.2
florifère	0	130		8.7

ou non. Il s'agit de la méthode de calcul LogDice, testée et avérée par Rychlý 2008 comme ayant des résultats significatifs en comparaison avec d'autres méthodes de calcul (T-score et MI score).

On voit que les adjectifs *sain* et *fertile* se trouvent dans la zone de recoupement rapproché, alors que les adjectifs *blanc*, *précoce* et *productif* se trouvent dans une zone où les scores diffèrent de manière plus prononcée. Le concordancier montre que la collocation de *vert* + *sain* correspond à des emplois végétaux (il s'agit de pelouse, feuillage ou encore de vacances vertes, saines). La collocation *sain* + *vigoureux* correspond plus à des emplois végétaux (arbustes, plantes, racines) mais aussi humains et métaphoriques / métonymiques (amour, génération, âme, nourriture, etc.).

Les résultats issus de Frantext sont reproduits dans le tableau 4 ci-contre, et montrent que l'adjectif *frais* correspond à la zone de recoupement cette fois-ci. Le concordancier montre 15 emplois de *frais* + *vert* qui correspondent à des descripteurs physiques de lieux (vallons, île, lierre, jardin, paradis). Dans le cas de *vigoureux* + *frais*, il

Tableau 4. Collexème commun de *vert* et *vigoureux* dans Frantext

Collexèmes	**Nombre d'occurrences en position *et / ou* avec *vert***	**Nombre d'occurrences en position *et / ou* avec *vigoureux***	**Score de collocation avec *vert***	**Score de collocation avec *vigoureux***
jaune	21	0	9.8	
bleu	24	0	9.6	
rouge	24	0	9.6	
gris	8	0	8.4	
gros	11	0	8.4	
frais	15	3	9.0	7.0
vivace	0	2		8.0
rusé	0	2		8.1
agile	0	3		8.4
adroit	0	4		8.6
mâle	0	4		8.6
sain	0	20		10.3

s'agit de 3 emplois descripteurs respectivement d'un enfant, d'organes et d'arbres.

Le même test est reproduit dans les tableaux 5 et 6 ci-après pour les lexèmes *green* et *vigorous* en anglais à partir du corpus Web enTenTen15 puis dans le corpus contemporain OEC. Aucune zone de recoupement n'est visible dans enTenTen15 pour *green* et *vigorous*, les scores étant très différenciés pour *green* et pour *vigorous* comme on l'observe dans le tableau 5.

En revanche, dans le corpus OEC, une zone de recoupement apparaît avec l'adjectif *healthy* comme on le voit dans le tableau 6. Dans le concordancier, on constate que la collocation *green* + *healthy* visible dans le tableau 7 ci-après s'avère être associée à la nourriture ou au feuillage après vérification des contextes d'occurrence. Pour ce qui est de *vigorous* + *healthy*, la différence est beaucoup plus marquée, avec peu de recoupements ; quasiment uniquement des sens métaphoriques pour *healthy* (*healthy debate*) et des emplois végétaux pour *vigorous*, avec l'exception du nom *debate* (*vigorous debate*).

Tableau 5. Collexèmes communs de *green* et *vigorous* dans enTenTen15

Collexèmes	**Nombre d'occurrences en position *and / or* avec *green***	**Nombre d'occurrences en position *and / or* avec *vigorous***	**Score de collocation avec *green***	**Score de collocation avec *vigorous***
blue	27 293	0	10.2	
red	26 135	0	10.0	
yellow	13 775	0	9.5	
dark	11 547	10	9.0	0.0
clean	11 409	20	8.9	1.0
lush	14 515	12	9.9	2.8
upright	23	86	0.7	6.6
moderate	25	296	0.7	7.0
hardy	20	96	0.5	6.9
free-flowering	0	39		6.1
youthful	0	88		6.5
aerobic	0	85		6.7

Tableau 6. Collexèmes communs à *green* et *vigorous* dans OEC

Collexèmes	**Nombre d'occurrences en position *and / or* avec *green***	**Nombre d'occurrences en position *and / or* avec *vigorous***	**Score de collocation avec *green***	**Score de collocation avec *vigorous***
red	3 257	0	10.0	
blue	2 372	0	9.8	
yellow	1 775	0	9.7	
lush	1 096	0	9.5	
dark	1 659	0	9.2	
fluorescent	737	0	8.9	

healthy	125	129	5.9	7.2
upright	5	13	1.8	6.1
moderate	5	46	1.6	6.7
aerobic	0	12		6.1
youthful	0	19		6.5
robust	0	35		6.5

Tableau 7. Collexèmes comparés de *healthy* + *vigorous* dans OEC

X modif N					
healthy			***vigourous***		
Collexèmes	**Nombre d'occurrences**	**Score**	**Collexèmes**	**Nombre d'occurrences**	**Score**
chest	1	1.25	Okinawans	1	8.34
democracy	2	0.57	termite	1	5.52
debate	4	0.19	alfalfa	1	5.18
cat	1	0.16	perennial	1	4.52
			strawberry	1	3.7
			seedling	3	3.46
			bud	1	2.06
			jew	1	1.68
			shoot	1	1.2
			vegetable	1	0.92
			grass	1	0.61
			debate	5	0.51

On voit donc le potentiel de cette démarche pour tester le rapport entre sémantique distributionnelle et proximité sémantique. On constate cependant la difficulté à établir des relations sémantiques de façon générale, car toute relation sémantique est tributaire de contextes. Les

questionnements sont similaires à ceux qui se posent dans d'autres travaux sur la synonymie et le calcul de la proximité sémantique, avec deux approches distinctes dont l'enjeu serait de réussir à les combiner. L'approche sur corpus est tributaire d'un type de corpus, de sa datation, des données disponibles, alors que l'approche lexicographique (telle qu'elle existe actuellement dans le *DES*) est dénuée des indicateurs contextuels et fréquentiels sur les usages (ces indicateurs sont désormais présents dans les ressources lexicographiques telles que l'*Oxford English Dictionary*, mais la fréquence n'est valable que d'un point de vue synchrone). Il y aura donc forcément une dichotomie entre les deux visions, l'une exhaustive du point de vue diachronique et l'autre tributaire des données d'un corpus.

C'est cette grande palette de synonymes proposée par le *DES* qui est à la fois son point fort et son point faible : la prise en compte de la désambiguïsation contextuelle est nécessaire pour la sélection du sens.

6. Conclusion de cette étude préliminaire : enjeux et pistes

L'objectif de cette étude est double. Tout d'abord, il s'agit d'expliquer la méthodologie (et, à terme, de développer le projet d'extraction automatique des synonymes) et les problématiques liées au projet d'extension du *DES* à l'anglais. L'un des grands obstacles à ce projet consiste à faire évoluer le *DES*, un dictionnaire « cumulatif et épuré » (Doualan 2013a), mais aussi un dictionnaire de synonymes diachronique provenant de sources hétéroclites et sans indication de contextes et de fréquence d'usage. De fait, le *DES* actuel présente des listes brutes de synonymes sans différenciation, hormis un classement et une liste de cliques parfois très longue, qui représente tous les effets de sens existants pour tous les synonymes, sans pondération de leur véritable fréquence d'usage, ni de la restriction de ce même usage. On peut citer, par exemple, dans la liste des synonymes de *femme*, les termes plus génériques et non sexués *adulte*, *personne*, *individu* ou encore *moitié*, sans compter les termes de registre populaire, argotique ou injurieux (*grognasse*, *poule*), ou encore, parmi la liste des synonymes de *chien*, le mot *lésineur* comme premier synonyme (mot peu fréquent). Cette problématique est évoquée par Hirst (1995 : 51) qui résume les caractéristiques fondamentales distinguant les parasynonymes, ou « *plesionyms* » de Cruse 1986 (voir le tableau 8 ci-contre).

La méthodologie de génération de synonymes dans le MultiDES va donc s'adosser à un outil existant multiforme. L'enjeu du projet d'extension des synonymes est de réfléchir à des pistes de travail de génération automatique de synonymes, d'intégration de contextes d'usage,

Tableau 8. Les facteurs distinctifs entre synonymes selon Hirst 1995

Distinctions	Exemples
Dénotation *coarse-grained* (distinction moyenne)	*yawl*, *ketch*
Dénotation *fine-grained* (distinction fine)	*lie*, *fib*
Dénotation *fuzzy* (distinction floue)	*forest*, *woods*
Emphase	*foe*, *enemy*
Implicature	*mislay*, *lose*
Formalité	*drunk*, *pissed*
Attitude	*skinny*, *slim*

de datation de contextes d'usage pour des relations synonymiques potentiellement désuètes, rares. Comme le souligne Doualan 2013a :

> Un dictionnaire électronique doit continuellement être alimenté pour suivre la langue dans ses évolutions, dans le cas contraire il risque de se figer. Dès lors, il rencontrerait les mêmes problèmes qu'un dictionnaire papier, les apports de la dictionnairique informatique se trouvant réduits à néant.

Le deuxième objectif est de tester un espace translangue GREEN et VERT. En ce qui concerne le test translangue, on voit le potentiel du modèle à visualiser les espaces de recoupement et de distinction entre des mots vedettes non problématiques (des adjectifs de couleur pour lesquels il existe souvent un sens « universel »). Ensuite, la pertinence d'une relation synonymique translangue est mise à l'épreuve par une approche de sémantique distributionnelle. Le test sur *vert-vigoureux* et *green-vigorous* montre que le potentiel existe, mais que l'entreprise est vaste, et surtout tributaire du type de corpus sélectionné, et dépendra ainsi de la mise en place de corpus parallèles français et anglais.

Cette étude préliminaire pose en conséquence la question de l'impact, de la pertinence d'une approche translangue. Quel emploi scientifique peut-on faire du *DES* et du MultiDES ? Qu'en serait-il pour un choix de vedettes plus problématiques, comme les étiquettes nominales, ou encore les faux amis ? Selon Victorri, la géométrie des graphes représenterait en quelque sorte d'éventuels sens primitifs dans la cognition humaine.

Peut-on ainsi améliorer sa lisibilité et donner accès à ces représentations mentales de manière plus intuitive ? L'autre question importante est celle du rapport entre ces représentations et les présuppositions que l'on entretient sur la structuration du lexique :

- La visualisation graphique coïncide-t-elle avec le lexique mental des usagers ?
- La visualisation graphique coïncide-t-elle avec l'organisation lexicographique ontologique ? A-t-elle une réalité psycholinguistique ou est-elle purement instrumentale ?

Selon Hirst (1995 : 51), la hiérarchisation du monde trouve un écho dans la hiérarchisation lexicale (ou *vice versa*) :

> *If two or more words denote the same concept, all will be included ; if a word is ambiguous, its different senses appear at their respective nodes. And conversely, it is assumed that, for each word sense in the lexicon, there is a node (or ensemble of nodes) somewhere in the hierarchy for it to be attached to ; otherwise, the system simply doesn't "understand" that word sense.*

Références bibliographiques

Études

Budanitsky A., Hirst G. (2006), « Evaluating WordNet-Based Measures of Lexical Semantic Relatedness », *Computational Linguistics*, vol. 32, nº 1, p. 13-47, en ligne : https://www.aclweb.org/anthology/J06-1003.pdf.

Cigada S. (2009), « Le concept de synonymie entre "langue" et "parole" », *Cahiers de l'Association internationale des études françaises*, nº 61, p. 13-29, en ligne : https://www.persee.fr/doc/caief_0571-5865_2009_num_61_1_2566.

Cruse D. A. (1986), *Lexical Semantics*, Cambridge – New York, Cambridge University Press.

Desagulier G. (2014), « Visualizing Distances in a Set of Near-Synonyms : *Rather*, *Quite*, *Fairly* and *Pretty* », in *Corpus Methods for Semantics : Quantitative Studies in Polysemy and Synonymy*, D. Glynn, J. A. Robinson (dir.), Amsterdam – Philadelphie, J. Benjamins, p. 145-178.

Desalle Y., Gaume B., Duvignau K., Cheung H., Hsieh S.-K., Magistry P., Nespoulous J.-L. (2014), « Skillex, an Action Labelling Efficiency Score : The Case for French and Mandarin », *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 36 (juillet 2014, Québec), p. 409-414.

Divjak D., Gries S. (2008), « Clusters in the Mind ? Converging Evidence from Near Synonymy in Russian », *The Mental Lexicon*, vol. 3, nº 2, p. 188-213.

Doualan G. (2011), *Cahier du CRISCO*, nº 32, *Introduction à une approche instrumentée de la synonymie*.

Doualan G. (2013a), « Analyse métalexicographique d'un dictionnaire des synonymes informatisé : l'exemple du Dictionnaire électronique des synonymes du CRISCO », *L'information grammaticale*, nº 136, p. 20-24.

Doualan G. (2013b), « La synonymie, relation d'équivalence, un artefact de la pensée ? », *Équivalences*, vol. 40, nº 1-2, *Lexicologie Terminologie Traduction : nouvelles recherches au cœur d'un système*, p. 15-42.

Edmonds P. (1998), « Translating Near-Synonyms : Possibilities and Preferences in the Interlingua », in *Proceedings of the AMTA / SIG-IL Second Workshop on Interlinguas* (Langhorne, octobre 1998), Technical Report MCCS-98-316, Computing Research Laboratory, New Mexico State University, p. 23-30.

Edmonds P., Hirst G. (2002), « Near-Synonymy and Lexical Choice », *Computational Linguistics*, vol. 28, nº 2, p. 105-144, en ligne : https://www.cs.toronto.edu/pub/gh/Edmonds+Hirst-2002.pdf.

Egan R. F. ([1942] 1968), « Survey of the History of English Synonymy » et « Synonym : Analysis and Definition », reproduits dans *Webster's New Dictionary of Synonyms*, P. B. Gove (dir.), Springfield, Merriam-Webster, p. 5a-31a.

Firth J. R. (1957), « A Synopsis of Linguistic Theory, 1930-1955 », in *Studies in Linguistic Analysis*, J. R. Firth (dir.), Oxford, B. Blackwell (numéro spécial de *Philological Society*), p. 1-32.

François J., Victorri B., Manguin J.-L. (2005), « Polysémie adjectivale et synonymie : l'éventail des sens de *curieux* », in *La polysémie*, O. Soutet (dir.), Paris, Presses de l'université Paris-Sorbonne, p. 157-170.

Gaillard B., Navarro E., Gaume B. (2010a), « From Binary Synonymy to Near Synonymy by Optimal Proxemy of Lexical Resources », in *Proceedings of Workshop on Computational Approaches to Synonymy* (Actes du colloque « Re-Thinking Synonymy », Helsinki, octobre 2010).

Gaillard B., Chudy Y., Magistry P., Hsieh S.-K., Navarro E. (2010b), « Graph Representation of Synonymy and Translation Resources for Cross-Linguistic Modelisation of Meaning », in *PACLIC 24 Proceedings* (Tohoku University, Sendai, novembre 2010), p. 819-830, en ligne : https://www.aclweb.org/anthology/Y10-1094.pdf.

Gaume B., Duvignau K., Prévot L., Desalle Y. (2008), « Toward a Cognitive Organization for Electronic Dictionaries, the Case for Semantic Proxemy », in *Coling 2008 : Proceedings of the Workshop on Cognitive Aspects of the Lexicon (Cogalex 2008)* (Manchester, août 2008), p. 86-93, en ligne : https://www.aclweb.org/anthology/W08-1912.pdf.

Glynn D. (2010), « Synonymy, Lexical Fields, and Grammatical Constructions. A Study in Usage-Based Cognitive Semantics », in *Cognitive Foundations of Linguistic Usage Patterns*, H.-J. Schmid, S. Handl (dir.), Berlin, De Gruyter Mouton, p. 89-118.

Hanks P. (2013), *Lexical Analysis : Norms and Exploitations*, Cambridge, MIT Press.

Hirst G. (1995), « Near-Synonymy and the Structure of Lexical Knowledge », in *AAAI Symposium on Representation and Acquisition of Lexical Knowledge : Polysemy, Ambiguity, and Generativity* (Stanford, mars 1995), p. 51-56, en ligne : https://www.aaai.org/Papers/Symposia/Spring/1995/SS-95-01/SS95-01-011.pdf.

Hoey M. (2005), *Lexical Priming : A New Theory of Words and Language*, Londres – New York, Routledge.

Hüllen W. (2004), *A History of Roget's Thesaurus*, Oxford, Oxford University Press.

Inkpen D., Hirst G. (2006), « Building and Using a Lexical Knowledge Base of Near-Synonym Differences », *Computational Linguistics*, vol. 32, nº 2, en ligne : https://www.aclweb.org/anthology/J06-2003.pdf.

Kilgarriff A. (1997), « "I Don't Believe in Word Senses" », *Computers and the Humanities*, nº 31, p. 91-113.

Kleiber G. (2009), « La synonymie–"identité de sens" n'est pas un mythe », *Pratiques*, nº 141/142, *La synonymie*, p. 9-25, en ligne : https://journals.openedition.org/pratiques/1262.

Kotlerman L., Dagan I., Szpektor I., Zhitomirsky-Geffet M. (2010), « Directional Distributional Similarity for Lexical Inference », *Natural Language Engineering*, vol. 16, nº 4, *Distributional Lexical Semantics*, p. 359-389.

Masseron C. (2009), « Présentation. Les paradoxes de la synonymie », *Pratiques*, nº 141/142, *La synonymie*, p. 3-8, en ligne : https://journals.openedition.org/pratiques/1260.

Murphy L. (2016), « Meaning Relations in Dictionaries : Hyponymy, Meronymy, Synonymy, Antonymy, and Contrast », in *The Oxford Handbook of Lexicography*, P. Durkin (dir.), Oxford, Oxford University Press, p. 439-456.

Pace-Sigge M. (2013), « The Concept of Lexical Priming in the Context of Language Use », *ICAME Journal*, nº 37, p. 149-174.

Ploux S., Boussidan A., Ji H. (2010), *The Semantic Atlas : An Interactive Model of Lexical Representation*, communication à la 7e Conference of International Language Resources and Evaluation, Valletta, Malte, mai 2010, en ligne : https://hal.archives-ouvertes.fr/hal-00933294/document.

Ploux S., Victorri B. (1998), « Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes », *TAL*, vol. 39, nº 1, p. 161-182.

Rychlý P. (2008), « A Lexicographer-Friendly Association Score », in *Proceedings of Recent Advances in Slavonic Natural Language Processing*, P. Sojka, A. Horák (dir.), Brno, Masaryk University, p. 6-9, en ligne : https://www.sketchengine.eu/wp-content/uploads/2015/03/Lexicographer-Friendly_2008.pdf.

Spärck Jones K. (1986), *Synonymy and Semantic Classification*, Édimbourg, Edinburgh University Press.

Stefanowitsch A., Gries S. T. (2003), « Collostructions : Investigating the Interaction of Words and Constructions », *International Journal of Corpus Linguistics*, vol. 8, nº 2, p. 209-243.

Stefanowitsch A., Gries S. T. (2004), « Covarying Collexemes in the *Into*-Causative », in *Language, Culture and Mind*, M. Achard, S. Kemmer (dir.), Stanford, CSLI Publications.

Venant F. (2004), « Polysémie et calcul du sens », in *Le poids des mots* (Actes des 7es journées internationales d'analyse statistique des données textuelles, Louvain-la-Neuve, 10-12 mars 2004), Louvain-la-Neuve, Presses universitaires de Louvain, p. 1145-1156.

Venant F., Victorri B. (2012), « La synonymie comme accès à la structure sémantique du lexique adjectif et verbal du français », in *La synonymie*, F. Berlan, G. Berthomieu (dir.), Paris, Presses de l'université Paris-Sorbonne, p. 73-90.

Vetulani Z., Vetulani G. (2015), « Synonymie et granularité dans les bases lexicales du type WordNet », *Studia Romanica Posnaniensia*, vol. 42, nº 1, p. 113-127.

Références lexicographiques

Collins English Thesaurus online, https://www.collinsdictionary.com/dictionary/english-thesaurus/green.

DES : *Dictionnaire électronique des synonymes*, Centre de recherches inter-langues sur la signification en contexte (CRISCO), université de Caen Normandie, https://crisco2.unicaen.fr/des/synonymes/vert.

Glosbe, https://glosbe.com/fr/en/vert.

Larousse langue française en ligne, https://larousse.fr/dictionnaires/francais/vert/81664.

Lexico, https://www.lexico.com/definition/green.

Merriam-Webster Thesaurus online, https://www.merriam-webster.com/thesaurus/green.

Le Robert en ligne, https://dictionnaire.lerobert.com/definition/vert.

Sensagent, http://dictionary.sensagent.com/vert/fr-en.

Le synonymeur, https://www.synonymeur.com/synonyme/vert/.

Synonymo.fr., http://www.synonymo.fr/synonyme/vert.

TLFi : *Trésor de la langue française informatisé*, laboratoire Analyse et traitement informatique de la langue française (ATILF), CNRS et université de Lorraine, http://atilf.atilf.fr.

Dictionnaires sources du DES

Bailly R. (1947), *Dictionnaire des synonymes de la langue française*, Paris, Larousse.

Bénac H. (1956), *Dictionnaire des synonymes conforme au dictionnaire de l'Académie française*, Paris, Hachette.

Bertaud du Chazaud H. (1979), *Dictionnaires des synonymes*, Paris, Robert.

Guilbert L., Lagane R., Niobey G. (dir.) (1971-1989), *Grand Larousse de la langue française*, Paris, Larousse, 7 vol.

Guizot F. (1809), *Dictionnaire universel des synonymes de la langue française*, Paris, Didier.

Lafaye B. (1858), *Dictionnaire des synonymes de la langue française*, Paris, Hachette.

Rey A. (dir.) (1985), *Le Grand Robert de la langue française. Dictionnaire alphabétique et analogique de la langue française*, Paris, Robert.

Corpus

BNC : British National Corpus, https://www.english-corpora.org/bnc.

COCA : Corpus of Contemporary American English, https://www.english-corpora.org/coca.

OEC : Oxford English Corpus, https://www.sketchengine.eu/oxford-english-corpus.