



Quel sens prennent les mots anglais les plus fréquents du discours de la santé ? Une étude lexicographique du corpus ScienText

Fabrice Gilles

► To cite this version:

Fabrice Gilles. Quel sens prennent les mots anglais les plus fréquents du discours de la santé ? Une étude lexicographique du corpus ScienText. LTT 2018, Sep 2018, Grenoble, France. ⟨hal-02013951⟩

HAL Id: hal-02013951

<https://hal.science/hal-02013951v1>

Submitted on 25 Jul 2019

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.




HAL Authorization

Quel sens prennent les mots anglais les plus fréquents du discours de la santé ? Une étude lexicographique du corpus ScienText

Conférence LTT 2018, Grenoble

Fabrice GILLES, LIDILEM Université Grenoble Alpes

gilles.fabrice@wanadoo.fr



LIDILEM
UNIVERSITÉ
Grenoble
Alpes

INTRODUCTION

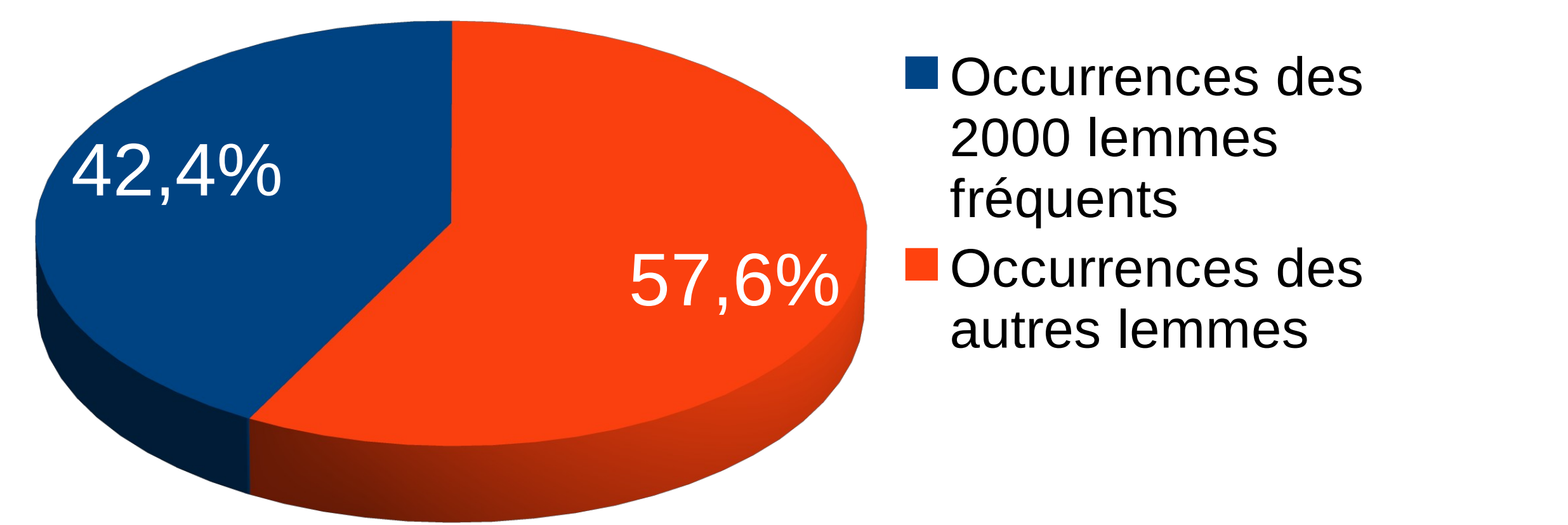
Nous avons trié parmi les mots anglais fréquents du discours scientifique de la santé ceux du Lexique Scientifique Transdisciplinaire (LST) et ceux du Lexique de la santé.

Notre objet d'étude est le corpus ScienText : 35 Mln de mots, 7564 textes tirés de 136 revues publiées par BioMedCentral.

La finalité est didactique : ces mots anglais ont été traduits en espagnol, français, italien et portugais afin de dresser un interlexique, autrement dit un lexique des mots analogues dans les cinq langues (similaires dans leur forme et leur sens)

1 – Description linguistique fondée sur l’usage

Extraction avec Treetagger des 2000 plus fréquents
Fréquence minimale : 1556
Rectification des erreurs d’étiquetage :catégorie grammaticale (*fat* étiqueté comme adjectif), etc.
Homogénéisation : graphie américaine color, program, etc.



Résultat : 476 adjectifs, 1150 noms, 374 verbes
Limite : cout en temps de la désambiguïisation manuelle

2 – Détermination des acceptions fréquentes des 2000 mots anglais fréquents

Utilisation de Lexicoscope (Kraif & Diwersy, 2012) pour établir le profil combinatoire (Blumental, 2007) et discerner les acceptions fréquentes
<https://frama.link/qCzuF29W>
Seuil de fréquence des cooccurents : 50

Lexicogramme									
Graphiques									
Show	25	entries							
I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likeliho	
dependence_N	on_O	~NOMPREP ~NN PREP ~DET	347	7492	273594	47778862	1	855,4856	
dependence_N	alcohol_N	NN	92	7492	3728	47778862	1	751,3441	
dependence_N	of_O	~NOMPREP NOMPREP ~NN ~SUJ PREP NNSPE OBJ ~DET PART	797	7492	2653809	47778862	1	295,0641	
dependence_N	density_N	~CC NN	60	7492	15309	47778862	1	271,7004	
dependence_N	the_D	~NOMPREP ~PREP DET PREP NN ~DET	346	7492	1740174	47778862	1	18,7977	

Le profil combinatoire du mot *dependence* suggère qu’il admet les deux acceptions définies par le Wiktionary

Cite this page	Noun [edit]
Visibility	dependence (countable and uncountable, plural dependences)
Show translations	1. The state of being dependent , of relying upon another. He had a deep dependence on her for guidance.
In other languages	2. An irresistible physical or psychological need , especially for a chemical substance. Her dependence on cocaine led to her ruin.
Català	
Čeština	
Ελληνικά	

Résultat : 2208 acceptions pour les 2000 mots fréquents
90 % des mots sont monosémiques
2/3 des polysémiques sont des noms
Limites : couteux en temps
Piste : abaisser le seuil

3 – Catégorisation sémantique

Dans le Lexique Scientifique Transdisciplinaire (LST) (Tutin, 2007) :

- orientation de la recherche, démarche scientifique, mesure, interprétation
- connaissance, dynamiques sociales, description des caractéristiques

Dans le lexique de spécialité de la santé (LS), nous rangeons ;

- anatomie générale, antomie comparée
- anatomopathologie générale, biologie médicale
- thérapeutique

	ADJECTIFS	NOMS	VERBES
LST	different high low other specific	analysis case control datum difference factor figure group level method model number percent rate result study table time type value	be base compare find have report show suggest use
Lexique de spécialité de la santé	clinical human	activity analysis cancer cell disease effect expression gene health mouse patient protein receptor response sample sequence treatment tumor	

Les 53 mots du premier quartile d’occurrences

Résultats sur les 2208 acceptions

LST : 53,9 % des acceptions, **59,3 % des occurrences**
LS : 46,1 % des acceptions, 40,68 % des occurrences

LS : 67 % des acceptions et 77 % des occurrences sont des noms
1/4 des acceptions sont des verbes mais 1/14 des occurrences seulement
Adjectifs peu nombreux dans le premier quartile mais importance croissante des adjectifs dans le lexique de spécialité

CONCLUSION
Ces résultats montrent l’importance du lexique scientifique transdisciplinaire dans le discours scientifique de la santé et confirme la place première des noms.
Application envisagée : un dispositif didactique qui valorise les analogies entre l’anglais et les langues romanes.