

Annexes

Table of Contents

Annexes	1
<i>Détail des résultats présentés dans le rapport.....</i>	<i>3</i>
TF-IDF + LDA.....	3
Doc2Vec + K-means	6
InstructorEmbeddings + K-means	9
Classe “sleep”	12
<i>Classification en 5 clusters.....</i>	<i>13</i>
TF-IDF + LDA.....	13
Doc2Vec + K-means	14
InstructorEmbeddings + K-means	15
<i>Autres méthodes d’encodage des mots</i>	<i>16</i>
Word2Vec.....	16
Small Bert 128.....	18
Glove.....	20
Extraction de mots-clés (keywords) puis LDA	22

TABLE 1: DETAILS DES 10 CLASSES OBTENUES AVEC L'ENCODAGE TF-IDF ET LA LDA.....	4
TABLE 2: DETAILS DES 10 CLASSES OBTENUES AVEC L'ENCODAGE Doc2Vec ET LE PARTITIONNEMENT AVEC LES K-MEANS.....	7
TABLE 3 : DETAILS DES 10 CLASSES OBTENUES AVEC L'ENCODAGE INSTRUCTOR EMBEDDINGS ET LE PARTITIONNEMENT AVEC LES K-MEANS.	10
TABLE 4 : TERMES REPRÉSENTATIFS DES 5 CLASSES IDENTIFIÉES AVEC LA LDA.	13
TABLE 5 : TERMES MeSH REPRÉSENTATIFS DES 5 CLASSES IDENTIFIÉES APRÈS PARTITIONNEMENT AVEC K-MEANS.....	14
TABLE 6 : INSTRUCTOR EMBEDDINGS. TERMES MeSH REPRÉSENTATIFS DES 5 CLASSES IDENTIFIÉES APRÈS PARTITIONNEMENT AVEC K-MEANS.....	15
TABLE 7 : WORD2VEC. TERMES MeSH REPRÉSENTATIFS DES 10 CLASSES IDENTIFIÉES APRÈS PARTITIONNEMENT AVEC K-MEANS.	16
TABLE 8 : SMALLBERT. TERMES MeSH REPRÉSENTATIFS DES 10 CLASSES IDENTIFIÉES APRÈS PARTITIONNEMENT AVEC K-MEANS.	19
TABLE 9 : GLOVE. TERMES MeSH REPRÉSENTATIFS DES 9 CLASSES IDENTIFIÉES APRÈS PARTITIONNEMENT AVEC K-MEANS.....	21
TABLE 10: TERMES REPRÉSENTATIFS DES 10 CLASSES IDENTIFIÉES AVEC LA LDA.	22

FIGURE 2 : ANALYSE SÉMANTIQUE LATENTE APRÈS ENCODAGE Doc2Vec. VISUALISATION DES RÉSUMÉS D'ARTICLES DANS LES 3 PREMIÈRES DIMENSIONS EN FONCTION DE LA LONGUEUR DES RÉSUMÉS.	6
FIGURE 3 : ANALYSE SÉMANTIQUE LATENTE APRÈS ENCODAGE INSTRUCTOR EMBEDDINGS. VISUALISATION DES RÉSUMÉS D'ARTICLES DANS LES 3 PREMIÈRES DIMENSIONS EN FONCTION DE LA LONGUEUR DES RÉSUMÉS.	9
FIGURE 4: NUAGE DES NOMS DE MOLÉCULES ACTIVES ("DRUGS") ASSOCIÉS AUX ARTICLES COLLECTÉS DANS LE GROUPE "SOMMEIL" POUR CHACUNE DES 3 MÉTHODES PRÉSENTÉES.	12
FIGURE 5: NOMBRE D'ARTICLES PAR AN DANS LA CLASSE "SOMMEIL" POUR CHAQUE MÉTHODE PRÉSENTÉE.	12
FIGURE 6 : ANALYSE SÉMANTIQUE LATENTE. PROJECTION SUR LES COMPOSANTES 1 ET 3 DE L'ACP EFFECTUÉE SUR LES RÉSUMÉS ENCODÉS EN VECTEURS AVEC LA MÉTHODE TF-IDF. LES COULEURS CORRESPONDENT AUX CLUSTERS DÉTERMINÉS PAR LDA.....	13
FIGURE 7 : ANALYSE SÉMANTIQUE LATENTE. PROJECTION SUR LES COMPOSANTES 1 ET 3 DE L'ACP EFFECTUÉE SUR LES RÉSUMÉS ENCODÉS EN VECTEURS AVEC LA MÉTHODE Doc2Vec. LES COULEURS CORRESPONDENT AUX 5 CLUSTERS DÉTERMINÉS PAR L'ALGORITHME DES K-MOYENNES.....	14
FIGURE 8 : COEFFICIENT DE SILHOUETTE OBTENU APRÈS LA PARTITION EN N CLUSTERS AVEC L'ALGORITHME DES K-MOYENNES.....	14
FIGURE 9 : INSTRUCTOR EMBEDDINGS. COEFFICIENT DE SILHOUETTE OBTENU APRÈS LA PARTITION EN N CLUSTERS AVEC L'ALGORITHME DES K-MOYENNES.....	15
FIGURE 10 : ANALYSE SÉMANTIQUE LATENTE. PROJECTION SUR LES COMPOSANTES 1 ET 3 DE L'ACP EFFECTUÉE SUR LES RÉSUMÉS ENCODÉS EN VECTEURS AVEC LA MÉTHODE INSTRUCTOR EMBEDDINGS. LES COULEURS CORRESPONDENT AUX 5 CLUSTERS DÉTERMINÉS PAR L'ALGORITHME DES K-MOYENNES.....	15
FIGURE 11 : WORD2VEC, COEFFICIENT DE SILHOUETTE OBTENU APRÈS LA PARTITION EN N CLUSTERS AVEC L'ALGORITHME DES K-MOYENNES.....	16
FIGURE 12 : SMALLBERT. COEFFICIENT DE SILHOUETTE OBTENU APRÈS LA PARTITION EN N CLUSTERS AVEC L'ALGORITHME DES K-MOYENNES.....	18
FIGURE 13 : GLOVE. COEFFICIENT DE SILHOUETTE OBTENU APRÈS LA PARTITION EN N CLUSTERS AVEC L'ALGORITHME DES K-MOYENNES.....	20

FIGURE 1 : ANALYSE SÉMANTIQUE LATENTE APRÈS ENCODAGE TF-IDF. VISUALISATION DES RÉSUMÉS D'ARTICLES DANS LES 3 PREMIÈRES DIMENSIONS EN FONCTION DE LA LONGUEUR DES RÉSUMÉS.....

3

Détail des résultats présentés dans le rapport

TF-IDF + LDA

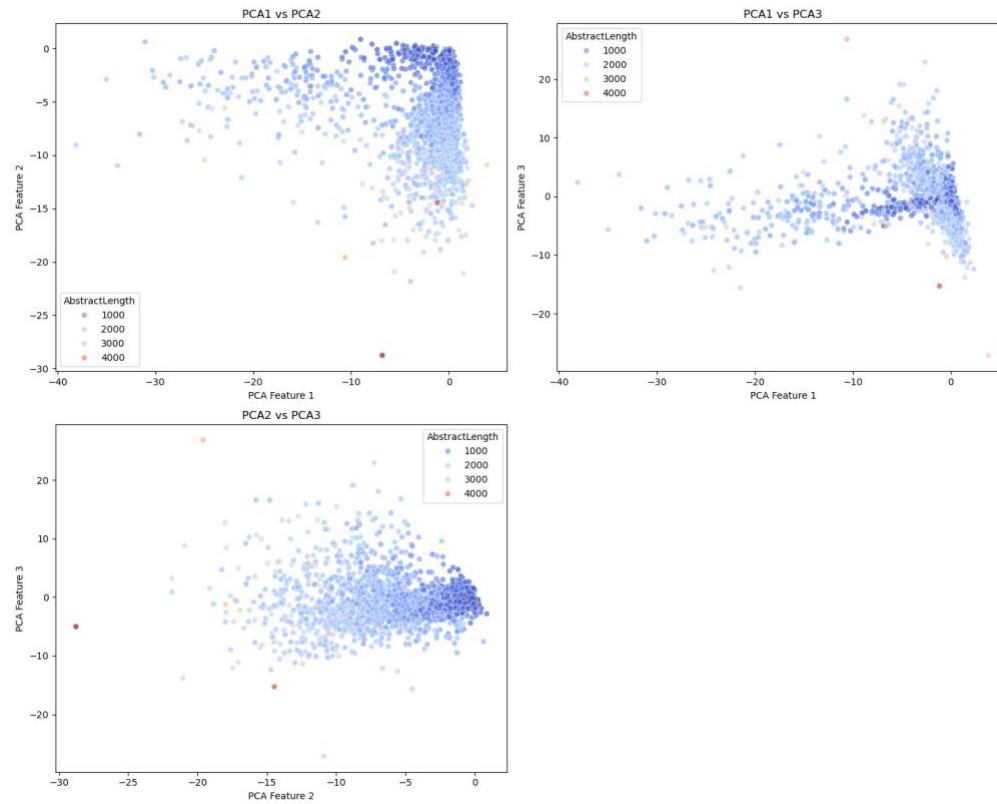


Figure 1 : Analyse sémantique latente après encodage TF-IDF. Visualisation des résumés d'articles dans les 3 premières dimensions en fonction de la longueur des résumés.

Table 1: Details des 10 classes obtenues avec l'encodage TF-IDF et la LDA.

Classe	Total nb. articles	Top 20 des termes MeSH	Nom attribué	Test de validation 20 articles
0	142	'Genotype', 'Gene Frequency', 'Polymorphism, Single Nucleotide', 'Catechol O-Methyltransferasegenetics', 'Alleles', 'Polymorphism, Genetic', 'Evoked Potentials, Auditoryphysiology', 'Loudness Perceptionphysiology', 'Genetic Predisposition to Disease', 'Reaction Timegenetics', 'Polymorphism, Single Nucleotidegenetics', 'Serotonin Plasma Membrane Transport Proteinsgenetics', 'Polymorphism, Geneticgenetics', 'Evoked Potentialsgenetics', 'Serotoninmetabolism', 'Dopamine Plasma Membrane Transport Proteinsgenetics', 'Genetic Variation', 'Methioninegenetics', 'Valinegenetics', 'Serotoninphysiology'	Monoamines and genetic variation	12/20
1	237	'Models, Neurological', 'Antiparkinson Agentstherapeutic use', 'Levodopatherapeutic use', 'Parkinson Diseasedrug therapyphysiopathology', 'Magnetoencephalographymethods', 'Evoked Potentialsphysiology', 'Beta Rhythmphysiology', 'Magnetoencephalography', 'Psychomotor Performancedrug effectsphysiology', 'Brain', 'Pattern Recognition, Visualphysiology', 'Motor Cortexdrug effectsphysiopathology', 'Wavelet Analysis', 'Evoked Potentials, Visualphysiology', 'Motor Cortexdrug effectsphysiology', 'Attentionphysiology', 'Neural Pathwaysdrug effectsphysiology', 'Neural Pathwaysphysiopathology', 'Neuronsdrug effectsphysiology', 'Evoked Potentialsdrug effectsphysiology'	Parkinson and motor control	8/20
2	275	'Sleep, REMdrug effects', 'Sleep Stagesdrug effects', 'Sleepdrug effectsphysiology', 'Sleep, REMphysiology', 'Polysomnographdrug effects', 'Sleep Stagesphysiology', 'Sleep Stagesdrug effectsphysiology', 'Sleep, REMdrug effectsphysiology', 'Wakefulnessdrug effects', 'Zolpidem', 'Hypnotics and Sedativespharmacology', 'Pyridinespharmacology', 'Azabicyclo Compounds', 'Sleep Stages', 'Polysomnographymethods', 'Arousal', 'Flunitrazepampharmacology', 'Paroxetinepharmacology', 'Plants, Medicinal', 'Sleep'	Sleep	15/20
3	265	'Fluorodeoxyglucose F18', 'Aspartic Acidanalogs & derivativesmetabolism', 'Creatinemetabolism', 'Magnetic Resonance Spectroscopy', 'Cholinemetabolism', 'Retrospective Studies', 'Tomography, Emission-Computed', 'Radiopharmaceuticals', 'Magnetic Resonance Imagingmethods', 'Anticonvulsantstherapeutic use', 'Positron-Emission Tomographymethods', 'Magnetic Resonance Spectroscopymethods', 'Hippocampuspathology', 'Protons', 'Glucosemetabolism', 'Aspartic Acidanalogs & derivativesanalysis', 'Seizures', 'Biomarkersblood', 'Cholineanalysis', 'Creatineanalysis'	Other neural imaging methods: PET, MRI in diagnosis	5/20
4	260	'Hydrocortisoneblood', 'Area Under Curve', 'Adrenocorticotrophic Hormoneblood', 'Half-Life', 'Affectdrug effects', 'Drug Administration Schedule', 'Drug Interactions', 'Hydrocortisonemetabolism', 'Prolactinblood', 'Circadian Rhythmdrug effects', 'Hormonesblood', 'Injections, Intravenous', 'Zolpidem', 'Saccadesdrug effects', 'Growth Hormoneblood', 'Central Nervous Systemdrug effects', 'Circadian Rhythmdrug effectsphysiology', 'Headachechemically induced', 'Human Growth Hormonemetabolism', 'Serotonin Antagonistspharmacology'	Pharmacokinetics and effect of drug administration on hormones levels	18/20
5	113	'Caffeinepharmacology', 'Breath Tests', 'Caffeine', 'Automobile Driving', 'Insulinblood', 'Alanine Transaminaseblood', 'Body Temperaturephysiology', 'Body Weight', 'Ghrelinblood', 'Lighting', 'Prone Position', 'Psychotic Disorders', 'Supine Position', 'Central Nervous System Stimulantspharmacology', 'Galvanic Skin Responsedrug effects', 'Memoryphysiology', 'Biomarkerscerebrospinal fluid', 'Blood Glucosedrug effectsmetabolism', 'Caffeinepharmacologytherapeutic use', 'Energy Intake'	Stimulants and attention	13/20

6	357	'Reward', 'Transcranial Magnetic Stimulationmethods', ' Evoked Potentialsphysiology ', ' Painphysiopathology ', 'Electric Stimulation', ' Evoked Potentials, Somatosensoryphysiology ', 'Cues', 'Motivation', ' Pain Thresholdphysiology ', ' Painchemically inducedphysiopathology ', ' Prefrontal Cortexphysiology ', ' Pain Measurement ', 'Emotionsphysiology', 'Conflict, Psychological', 'Methylphenidatepharmacology', 'Inhibition, Psychological', ' Evoked Potentials, Somatosensorydrug effects ', 'Methamphetamineadverse effects', ' Pain Thresholddrug effects ', ' Capsaicinpharmacology '	Sensory processing and pain	9/20
7	153	'Acoustic Stimulation', ' Schizophreniadrug therapyphysiopathology ', 'Auditory Perceptiondrug effectsphysiology', 'Evoked Potentials, Auditorydrug effectsphysiology', 'Acoustic Stimulationmethods', ' Antipsychotic Agentstherapeutic use ', ' Schizophrenic Psychology ', 'Contingent Negative Variationdrug effects', 'Excitatory Amino Acid Antagonistspharmacology', 'Ketaminepharmacology', 'Sensory Gatingdrug effects', 'Sensory Gatingdrug effectsphysiology', ' Antipsychotic Agentspharmacology ', 'Evoked Potentials, Auditoryphysiology', ' Schizophreniaphysiopathology ', 'Cerebral Cortexdrug effectsphysiopathology', ' Schizophreniadrug therapy ', 'Homovanillic Acidblood', 'Nicotinic Agonistsadministration & dosage', 'Prepulse Inhibitiondrug effectsphysiology'	Schizophrenia	10/20
8	177	' Propofolpharmacology ', ' Anesthesia, General ', ' Sevoflurane ', ' Propofoladministration & dosage ', ' Monitoring, Intraoperativemethods ', ' Anesthetics, Intravenousadministration & dosage ', ' Anesthesia Recovery Period ', ' Conscious Sedation ', ' Hypnotics and Sedativesadministration & dosage ', ' Remifentanil ', ' Anesthetics, Intravenouspharmacology ', ' Anesthetics, Intravenous ', 'Entropy', ' Propofol ', ' Anesthetics, Inhalation ', ' Monitoring, Intraoperative ', ' Consciousness Monitors ', ' Nitrous Oxide ', ' Midazolam ', ' Conscious Sedationmethods '	Anesthesia and sedation	19/20
9	115	'Models, Biological', 'Remifentanil', 'Memory Consolidationdrug effectsphysiology', 'Receptors, Muscarinicmetabolism', 'Scopolaminepharmacology', 'Central Nervous Systemdrug effects', 'Cycloserinepharmacology', 'Analgesiamethods', 'Benzodiazepinesadministration & dosageblood', 'Curcuminpharmacology', 'Cytochrome P-450 CYP3A Inhibitors', 'Furansadverse effectspharmacokineticspharmacology', 'Glycine Plasma Membrane Transport Proteinsantagonists & inhibitors', 'Nicotinic Antagonistspharmacology', 'Painchemically induceddrug therapy', 'Psychomotor Disorderschemically induced', 'Receptors, Glycinemetabolism', 'Signal Transductiondrug effects', 'Affectdrug effects', 'Cholinesterase Inhibitorspharmacology'	-	-

Doc2Vec + K-means

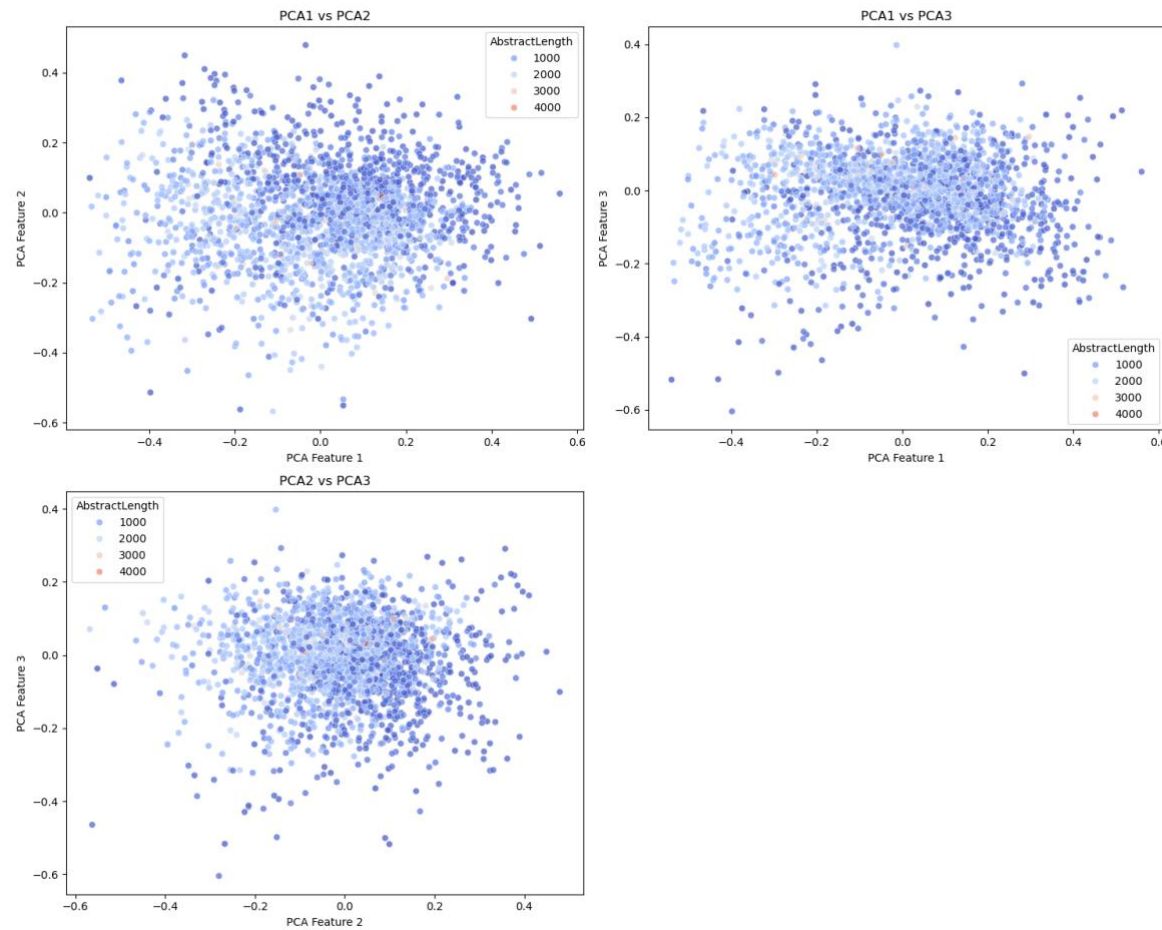


Figure 2 : Analyse sémantique latente après encodage Doc2Vec. Visualisation des résumés d'articles dans les 3 premières dimensions en fonction de la longueur des résumés.

Table 2: Details des 10 classes obtenues avec l'encodage Doc2Vec et le partitionnement avec les K-means.

Classe	Total nb. articles	Top 20 des termes MeSH	Nom attribué	Test de validation 20 articles
0	175	'Braindiagnostic imagingdrug effectsmetabolism', 'Receptors, Serotoninindrug effects', 'Autonomic Nervous Systemdrug effectsphysiology', 'Japan', 'Braindrug effectsmetabolism', 'Analgesiamethods', 'Cold Temperatureadverse effects', 'Curcuminpharmacology', 'Cytochrome P-450 CYP3A Inhibitors', 'Furansadverse effectspharmacokineticspharmacology', 'Glycine Plasma Membrane Transport Proteinsantagonists & inhibitors', 'N,N-Dimethyltryptaminepharmacology', 'Nicotinic Antagonistspharmacology', 'Orientationdrug effects', 'Paclitaxeladverse effectstherapeutic use', 'Psychomotor Disorderschemically induced', 'Receptors, Glycinemetabolism', 'Selective Serotonin Reuptake Inhibitorsadverse effects', 'Serotonin Receptor Agonistsadverse effectspharmacology', 'Solventstoxicity'	-	
1	209	'Genotype', 'Polymorphism, Single Nucleotide', 'Evoked Potentialsphysiology', 'Evoked Potentials, Auditoryphysiology', 'Acoustic Stimulationmethods', 'Loudness Perceptionphysiology', 'Dopaminemetabolism', 'Polymorphism, Geneticgenetics', 'Psychomotor Performancephysiology', 'Reward', 'Alleles', 'Reaction Timegenetics', 'Evoked Potentials', 'Prefrontal Cortexphysiology', 'Case-Control Studies', 'Serotonin Plasma Membrane Transport Proteinsgenetics', 'Evoked Potentialsgenetics', 'Evoked Potentials, Auditorydrug effectsphysiology', 'Executive Functionphysiology', 'Inhibition, Psychological'	-	
2	195	'Sevoflurane', 'Remifentanyl', 'Consciousness Monitors', 'Anesthetics, Intravenousadministration & dosage', 'Propofol', 'Monitoring, Intraoperativemethods', 'Monitoring, Intraoperative', 'Blood Pressuredrug effects', 'Entropy', 'Infusions, Intravenous', 'Anesthetics, Intravenous', 'Anesthesia, General', 'Cerebrovascular Circulationdrug effects', 'Hemodynamicsdrug effects', 'Propofolpharmacology', 'Anesthesia Recovery Period', 'Anesthetics, Inhalationpharmacology', 'Propofoladministration & dosage', 'Respiratory Mechanicsdrug effects', 'Carbon Dioxideblood'	Anaesthetics and sedatives	13/20
3	103	'Sleep, REMdrug effects', 'Sleep Stagesdrug effects', 'Sleepdrug effects', 'Polysomnographydrug effects', 'Wakefulnessdrug effects', 'Sleepdrug effectsphysiology', 'Azabicyclo Compounds', 'Hypnotics and Sedativesadministration & dosagepharmacology', 'Sleep Initiation and Maintenance Disorderschemically induced', 'Sodium Oxybatepharmacology', 'Zolpidem', 'Piperazinespharmacology', 'Darkness', 'Depressive Disorderdiagnosisdiagnostic imagingmetabolism', 'Hydroxybutyratespharmacology', 'Sleep Deprivationdrug therapyphysiopathologypsychology', 'Sleep Wake Disordersblood', 'Sleep Wake Disorderschemically induced', 'Temazepamadministration & dosage', 'Norepinephrineblood'	Sleep 1	14/20
4	160	'Half-Life', 'Saccadesdrug effects', 'Administration, Oral', 'Infusions, Intravenous', 'Metabolic Clearance Rate', 'Central Nervous Systemdrug effects', 'Injections, Intravenous', 'Arousaldrug effects', 'Drug Interactions', 'Models, Biological', 'Biological Availability', 'Histamine H1 Antagonistspharmacology', 'Fluoroquinolones', 'Hypnotics and Sedativesadverse effectspharmacokineticspharmacology', 'Piperazinesadministration & dosageadverse effectspharmacokinetics', 'Quinolonesadministration & dosageadverse effectspharmacokinetics', 'Tablets', 'Blood Pressuredrug effects', 'Zolpidem', 'Drug Administration Schedule'	Drug administration dynamics and kinetics	11/20

5	159	'Psychometrics', 'Caffeinepharmacology', 'Bromazepampharmacology', 'Ethanolpharmacology', 'Antipsychotic Agentsadverse effectspharmacokineticspharmacology', 'Atenololpharmacology', 'Benzilates', 'Bridged Bicyclo Compounds, Heterocyclic', 'Mandelic Acidsadverse effectspharmacology', 'Muscarinic Antagonistsadverse effectspharmacology', 'Nortropanesadverse effectspharmacology', 'Pemolinepharmacology', 'Problem Solvingdrug effects', 'Riluzole', 'Blood Gas Analysis', 'Anticonvulsantspharmacology', 'Beta Rhythmdrug effects', 'Oxazepampharmacology', 'Anti-Anxiety Agentsadministration & dosage pharmacology', 'Electrocardiographydrug effects'	-	
6	296	'Magnetic Resonance Imaging', 'Fluorodeoxyglucose F18', 'Case-Control Studies', 'Retrospective Studies', 'Aspartic Acidanalogs & derivativesmetabolism', 'Creatinemetabolism', 'Magnetic Resonance Spectroscopy', 'Cholinemetabolism', 'Anticonvulsants therapeutic use', 'Tomography, Emission-Computed', 'Infant', 'Aspartic Acidanalogs & derivativesanalysis', 'Seizures', 'Age of Onset', 'Radiopharmaceuticals', 'Magnetic Resonance Spectroscopymethods', 'Positron-Emission Tomography', 'Disease Progression', 'Brainpathology', 'Cholineanalysis'	Other neural imaging methods: PET, MRI in diagnosis	10/20
7	435	'Acoustic Stimulation', 'Evoked Potentialsphysiology', 'Reaction Timephysiology', 'Evoked Potentials, Somatosensoryphysiology', 'Evoked Potentialsdrug effects', 'Evoked Potentialsdrug effectsphysiology', 'Brainphysiology', 'Reaction Time', 'Electric Stimulation', 'Reaction Timedrug effectsphysiology', 'Attentionphysiology', 'Photic Stimulationmethods', 'Cognitionphysiology', 'Transcranial Magnetic Stimulationmethods', 'Evoked Potentials', 'Event-Related Potentials, P300drug effects', 'Brain Mappingmethods', 'Evoked Potentials, Auditorydrug effectsphysiology', 'Pain Measurement', 'Cognition'	Evoked potentials	13/20
8	208	'Alpha Rhythm', 'Rest', 'Antiparkinson Agents therapeutic use', 'Beta Rhythm', 'Brain', 'Alpha Rhythmdrug effectsmethods', 'Models, Neurological', 'Magnetoencephalography', 'Delta Rhythm', 'Frontal Lobephysiology', 'Theta Rhythmdrug effects', 'Hand Strengthphysiology', 'Parkinson Disease', 'Levodopatherapeutic use', 'Motor Cortexdrug effectsphysiology', 'Restphysiology', 'Propofol pharmacology', 'Cerebral Cortexphysiology', 'Magnetic Resonance Imaging', 'Beta Rhythmdrug effects'	Brain waves at rest	8/20
9	154	'Sleep Stagesphysiology', 'Sleepdrug effects', 'Sleep, REMphysiology', 'Sleepdrug effectsphysiology', 'Wakefulnessphysiology', 'Sleep, REMdrug effectsphysiology', 'Sleep Stagesdrug effectsphysiology', 'Sleep', 'Sleep, REMdrug effects', 'Polysomnographymethods', 'Sleep Stages', 'Sleep Stagesdrug effects', 'Polysomnographydrug effects', 'Hydrocortisonemetabolism', 'Adrenocorticotrophic Hormonemetabolism', 'Depressive Disorderphysiopathology', 'Dexamethasonepharmacology', 'Memory Consolidationdrug effectsphysiology', 'Succinimides', 'Periodicity'	Sleep 2	13/20

InstructorEmbeddings + K-means

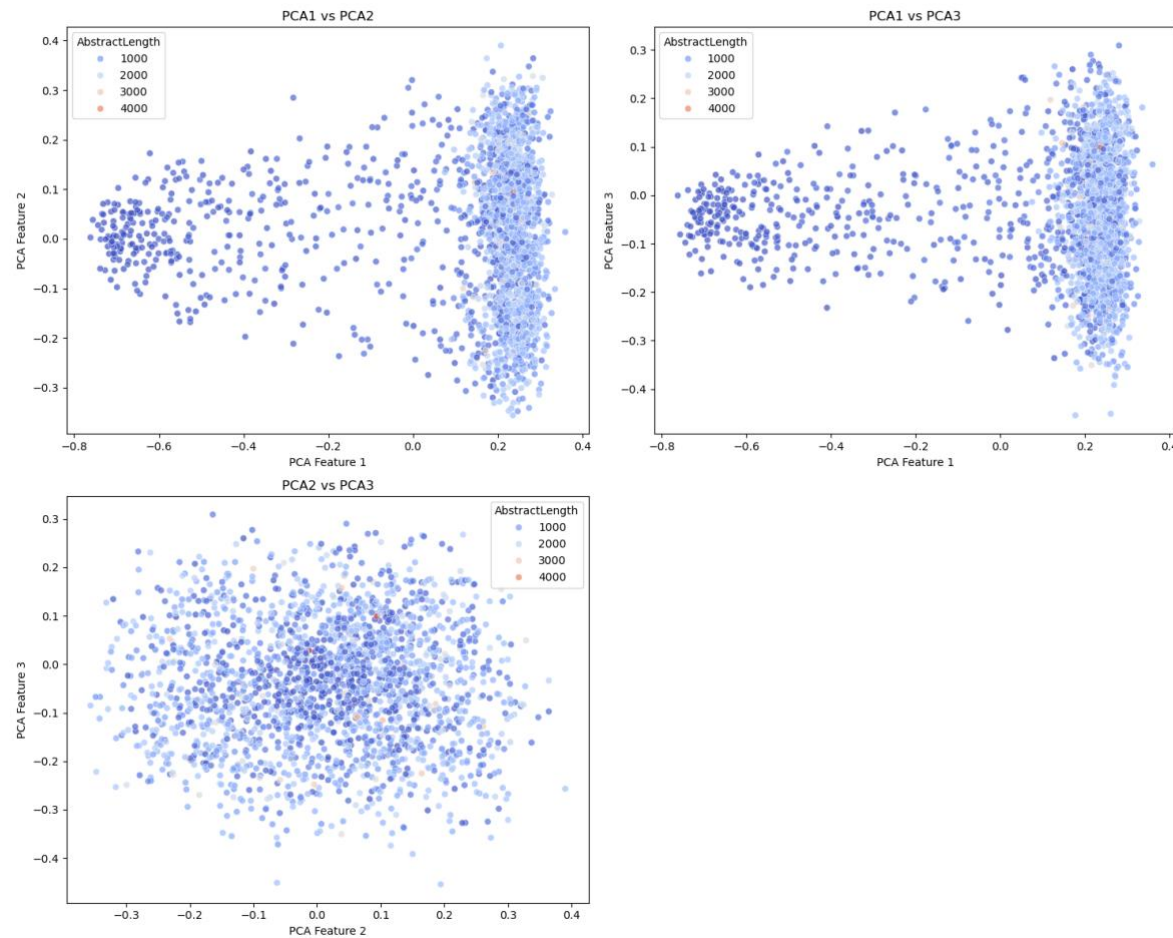


Figure 3 : Analyse sémantique latente après encodage *InstructorEmbeddings*. Visualisation des résumés d'articles dans les 3 premières dimensions en fonction de la longueur des résumés.

Table 3 : Details des 10 classes obtenues avec l'encodage *InstructorEmbeddings* et le partitionnement avec les *K-means*.

Classe	Total nb. articles	Top 20 des termes MeSH	Nom attribué	Test de validation 20 articles
0	237	'Alpha Rhythm', 'Algorithms', ' Anticonvulsantspharmacology ', ' Ethanolpharmacology ', 'Isoflurane', 'Electroencephalographydrug effectsmethods', ' Propofolpharmacology ', ' Diazepampharmacology ', ' Levetiracetampharmacology ', ' Migraine Disordersdrug therapyphysiopathology ', 'Seizureschemically induced', 'Predictive Value of Tests', 'Rest', 'Unconsciousnesschemically inducedphysiopathology', 'Motor Cortexdrug effectsphysiology', ' Bromazepampharmacology ', 'Random Allocation', 'Topiramate', 'Electroencephalographymethodsstandards', ' Ethanolpharmacokineticspharmacology '	Drug pharmacology	8/20
1	275	'Electroencephalographydrug effectsmethods', 'Fluorodeoxyglucose F18', 'Glucosemetabolism', 'Benzodiazepines', ' Anesthetics, Inhalationpharmacology ', ' Hypnotics and Sedativesadministration & dosagepharmacokinetics ', 'Midazolampharmacokineticspharmacology', 'Motor Skillsphysiology', ' Propofolpharmacokineticspharmacology ', 'Sensitivity and Specificity', 'Infant', ' Analgesics, Opioidadverse effectspharmacology ', 'Autonomic Nervous Systemphysiopathology', 'Benzodiazepinesadministration & dosageblood', 'Child Behavior', 'Child Developmentphysiology', 'Cognition Disorderschemically inducedphysiopathology', 'Environment', 'Epilepsydiagnosisphysiopathology', 'Glasgow Coma Scale'	-	
2	202	' Capsaicinpharmacology ', ' Evoked Potentials, Somatosensoryphysiology ', ' Painchemically inducedphysiopathology ', ' Hot Temperature ', 'Polymorphism, Single Nucleotidegenetics', 'Receptors, Serotonin, 5-HT1', ' Vagus Nerve Stimulation ', ' Lasers ', ' Hydrochloric Acidadministration & dosage ', 'Ketanserinpharmacology', 'Nerve Fibers, Myelinatedphysiology', 'Seizuresdrug therapy', 'Ultraviolet Raysadverse effects', 'Evoked Potentials, Somatosensorydrug effectsphysiology ', ' Pain Measurementmethods ', ' Evoked Potentials, Somatosensorydrug effects ', 'Phenotype', 'Drug Interactions', ' Evoked Potentials, Somatosensory ', 'Skininnervation'	Pain	11/20
3	225	[' Sleep, REMdrug effects ', ' Sleepdrug effects ', ' Sleep Stagesdrug effects ', ' Sleepdrug effectsphysiology ', ' Sleepphysiology ', ' Polysomnographydrug effects ', ' Sleep, REMphysiology ', ' Sleep, REM ', ' Sleep ', ' Wakefulnessphysiology ', ' Polysomnographymethods ', ' Wakefulnessdrug effects ', ' Sleep Stages ', 'Zolpidem', ' Sleep Stagesdrug effectsphysiology ', ' Sleep Stagesphysiology ', ' Circadian Rhythmdrug effects ', 'Arousal', ' Sleep, REMdrug effectsphysiology ', ' Sleep Quality ']	Sleep	20/20
4	313	['Aspartic Acidanalogs & derivativesmetabolism', 'Magnetic Resonance Spectroscopy', 'Creatinemetabolism', ' Antiparkinson Agentstherapeutic use ', ' Levodopatherapeutic use ', 'Cholinemetabolism', 'Fluorodeoxyglucose F18', 'Aged, 80 and over', 'Image Processing, Computer-Assistedmethods', ' Parkinson Diseasedrug therapyphysiopathology ', 'Positron-Emission Tomographymethods', 'Infant', 'Disease Progression', 'Protons', 'Magnetic Resonance Imagingmethods', 'Phosphocreatinemetabolism', 'Transcranial Magnetic Stimulationmethods', 'Follow-Up Studies', 'Tomography, Emission-Computed', 'Radiopharmaceuticals']	-	
5	300	['Psychomotor Performancedrug effects', 'Saccadesdrug effects', ' Half-Life ', ' Drug Administration Schedule ', 'Remifentanyl', 'Affectdrug effects', 'Central Nervous Systemdrug effects', ' Area Under Curve ', ' Hemodynamicsdrug effects ', 'Memorydrug effects', ' Oxazepampharmacology ', 'Conscious Sedationmethods', ' Metabolic Clearance Rate ', 'Sevoflurane', 'Blood Gas Analysis', 'Electrocardiographydrug effects', 'Monitoring, Intraoperative', ' Pursuit, Smoothdrug effects ', ' Postural Balancedrug effects ', ' Hypnotics and Sedativespharmacology ']	PK-PD studies	9/20

6	124	'Schizophrenic Psychology', 'Schizophreniadrug therapyphysiopathology', 'Antipsychotic Agentstherapeutic use', 'Schizophreniaphysiopathology', 'Sensory Gatingdrug effectsphysiology', 'Sensory Gatingdrug effects', 'Acoustic Stimulationmethods', 'Schizophreniadiagnosisdrug therapyphysiopathology', 'Antipsychotic Agentspharmacologytherapeutic use', 'Clozapinepharmacologytherapeutic use', 'Schizophrenia', 'Auditory Cortexphysiopathology', 'Prefrontal Cortexmetabolism', 'Evoked Potentials, Auditoryphysiology', 'Olanzapine', 'Evoked Potentials, Auditorydrug effectsphysiology', 'Amisulpride', 'Benzodiazepinetherapeutic use', 'Clozapinetherapeutic use', 'Genome-Wide Association Study'	Schizophrenia	11/20
7	37	'Contingent Negative Variationdrug effects', 'Excitatory Amino Acid Antagonistspharmacology', 'Ketaminepharmacology', 'Nicotinic Agonistsadministration & dosage', 'Psychoses, Substance-Inducedphysiopathology', 'Acetylcysteinepharmacologytherapeutic use', 'Orientation drug effects', 'Nicotineadministration & dosage', 'Receptor, Serotonin, 5-HT _{2A} ', 'Receptors, N-Methyl-D-Aspartateantagonists & inhibitors', 'Acetylcysteinepharmacology', 'Alcoholismepidemiologyphysiopathology', 'Alzheimer Disease therapy', 'Amphetamine-Related Disordersphysiopathologydiagnosis', 'Anti-N-Methyl-D-Aspartate Receptor Encephalitisdiagnostic imagingmetabolismphysiopathology', 'Aphasiaphysiopathology', 'Basal Nucleus of Meynert', 'Brainphysiologyphysiopathologydrug effects', 'Bromocriptineadministration & dosage pharmacology', 'Cannabinolanalogs & derivativespharmacology'	-	
8	67	'Hydrocortisoneblood', 'Adrenocorticotrophic Hormoneblood', 'Estradiolblood', 'Sleepphysiology', 'Human Growth Hormonemetabolism', 'Pituitary-Adrenal Systemphysiology', 'Adrenocorticotrophic Hormonebloodmetabolism', 'Hydrocortisonebloodmetabolism', 'Progesteroneblood', 'Aldosteroneblood', 'Corticotropin-Releasing Hormonepharmacology', 'Dexamethasonepharmacology', 'Growth Hormone-Releasing Hormoneadministration & dosage pharmacology', 'Hypothalamo-Hypophyseal Systemphysiology', 'Mifepristonepharmacology', 'Testosteronemetabolism', 'Growth Hormoneblood', 'Hydrocortisonemetabolism', 'Imipraminetherapeutic use', 'Reninblood'	Hormones studies	17/20
9	314	['Reward', 'Evoked Potentialsphysiology', 'Cues', 'Inhibition, Psychological', 'Attentionphysiology', 'Evoked Potentials, Auditoryphysiology', 'Brainphysiology', 'Reaction Time', 'Genotype', 'Emotionsphysiology', 'Reaction Timephysiology', 'Psychomotor Performancephysiology', 'Conflict, Psychological', 'Cocaine-Related Disordersphysiopathology', 'Cognitionphysiology', 'Polymorphism, Single Nucleotide', 'Methylphenidatepharmacology', 'Serotoninmetabolism', 'Choice Behaviorphysiology', 'Methylphenidatetherapeutic use']	-	

Classe "sleep"



Figure 4: Nuage des noms de molécules actives ("drugs") associés aux articles collectés dans le groupe "sommeil" pour chacune des 3 méthodes présentées.

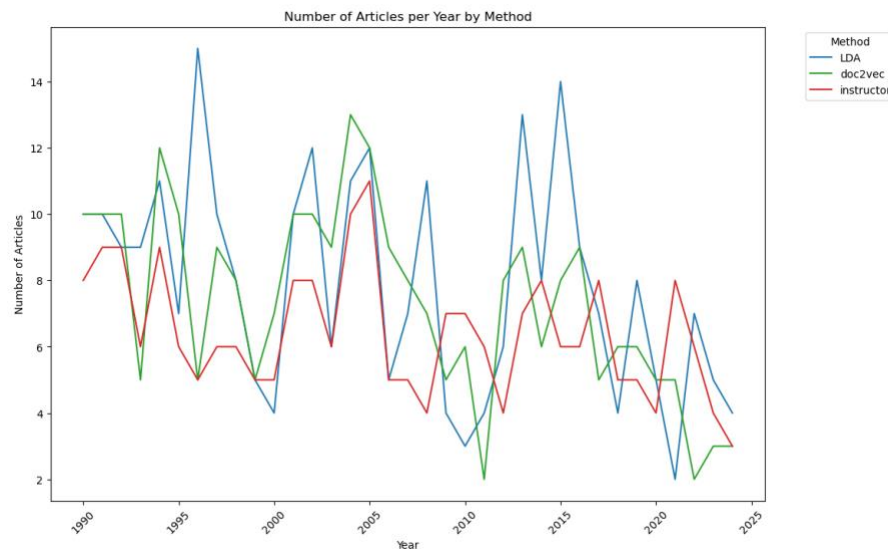


Figure 5: Nombre d'articles par an dans la classe "sommeil" pour chaque méthode présentée.

Classification en 5 clusters

TF-IDF + LDA

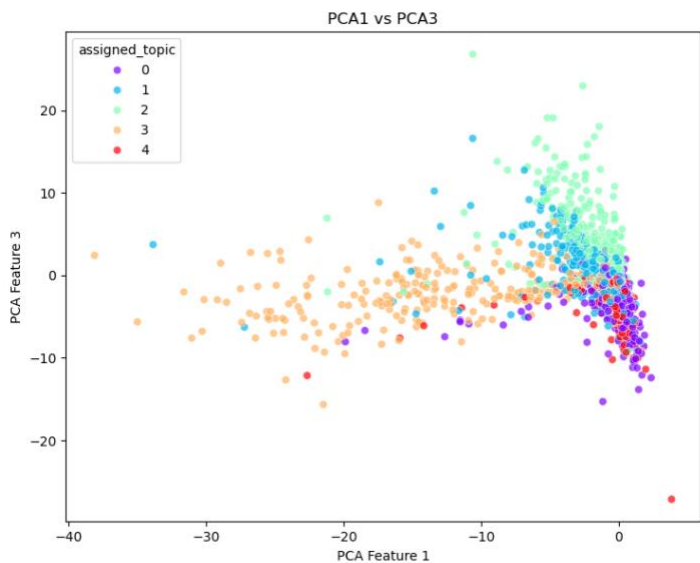


Figure 6 : Analyse sémantique latente. Projection sur les composantes 1 et 3 de l'ACP effectuée sur les résumés encodés en vecteurs avec la méthode TF-IDF. Les couleurs correspondent aux clusters déterminés par LDA.

Classe	10 termes les plus représentatifs
0	['component', 'motor', 'response', 'cortex', 'cortical', 'network', 'baseline', 'plasma', 'beta', 'process']
1	['power', 'alpha', 'task', 'rest', 'cognitive', 'analysis', 'theta', 'erp', 'stimulus', 'placebo']
2	['include', 'dose', 'concentration', 'mg', 'reflect', 'sedation', 'perform', 'infusion', 'min', 'due']
3	['sleep', 'night', 'rem', 'cortisol', 'marker', 'hormone', 'general', 'interval', 'ratio', 'receptor']
4	['patient', 'epilepsy', 'seizure', 'child', 'group', 'single', 'control', 'ad', 'recent', 'feature']

Table 4 : Termes représentatifs des 5 classes identifiées avec la LDA.

Doc2Vec + K-means

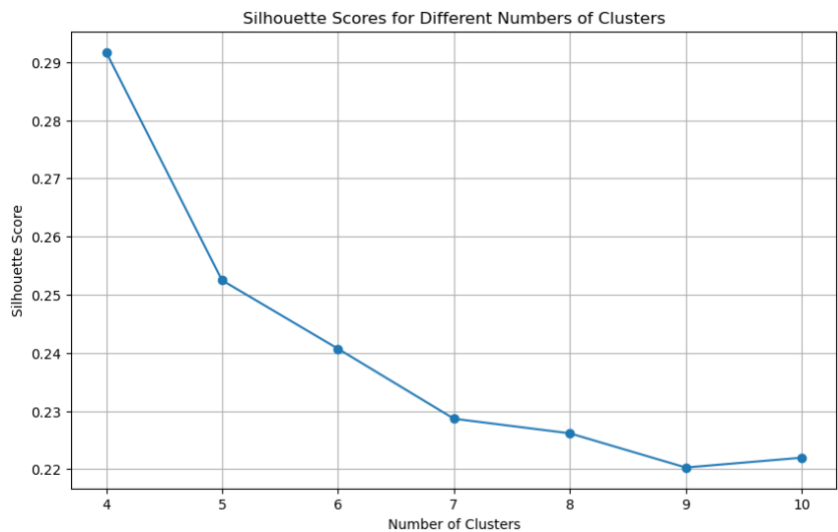


Figure 8 : Coefficient de silhouette obtenu après la partition en n clusters avec l'algorithme des K-moyennes.

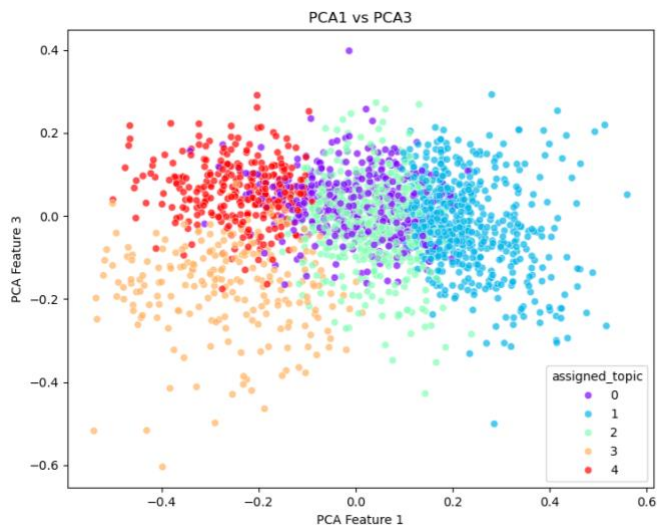


Figure 7 : Analyse sémantique latente. Projection sur les composantes 1 et 3 de l'ACP effectuée sur les résumés encodés en vecteurs avec la méthode Doc2Vec. Les couleurs correspondent aux 5 clusters déterminés par l'algorithme des K-moyennes

Classe	Total nb. articles	Top 10 des termes MeSH
0	408	['Alpha Rhythm', 'Brainphysiology', 'Magnetoencephalography', 'Rest', 'Brain', 'Brain Mappingmethods', 'Reproducibility of Results', 'Delta Rhythm', 'Cognition', 'Motor Cortexdrug effectsphysiology']
1	572	['Case-Control Studies', 'Magnetic Resonance Imaging', 'Evoked Potentialsphysiology', 'Evoked Potentials, Auditoryphysiology', 'Evoked Potentials', 'Reward', 'Acoustic Stimulationmethods', 'Brainphysiopathology', 'Fluorodeoxyglucose F18', 'Psychomotor Performancephysiology']
2	592	['Infant', 'Hypnotics and Sedativespharmacology', 'Evoked Potentials, Auditorydrug effects', 'Event-Related Potentials, P300drug effects', 'Follow-Up Studies', 'Remifentanyl', 'Evoked Potentialsdrug effectsphysiology', 'Sevoflurane', 'Anesthesia, General', 'Models, Biological']
3	226	['Sleep, REMdrug effects', 'Sleep Stagesdrug effects', 'Sleepdrug effectsphysiology', 'Polysomnographdrug effects', 'Sleep Stagesdrug effectsphysiology', 'Wakefulnessdrug effects', 'Sleep Stagesphysiology', 'Sleep, REMphysiology', 'Wakefulnessphysiology', 'Sleep Stages']
4	296	['Infusions, Intravenous', 'Half-Life', 'Drug Administration Schedule', 'Saccadesdrug effects', 'Injections, Intravenous', 'Zolpidem', 'Central Nervous Systemdrug effects', 'Electrocardiographdrug effects', 'Metabolic Clearance Rate', 'Pain Measurementdrug effects']

Table 5 : Termes MeSH représentatifs des 5 classes identifiées après partitionnement avec K-means.

InstructorEmbeddings + K-means

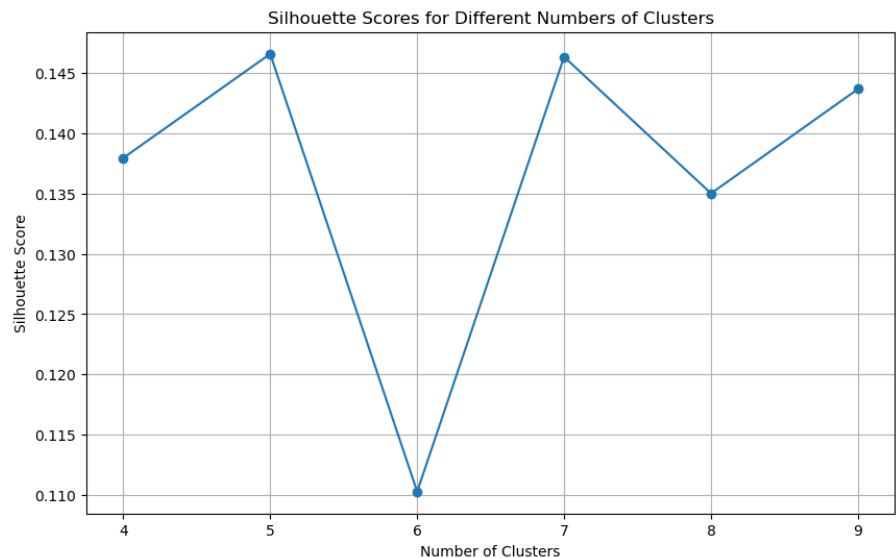


Figure 9 : InstructorEmbeddings. Coefficient de silhouette obtenu après la partition en n clusters avec l'algorithme des K-moyennes.

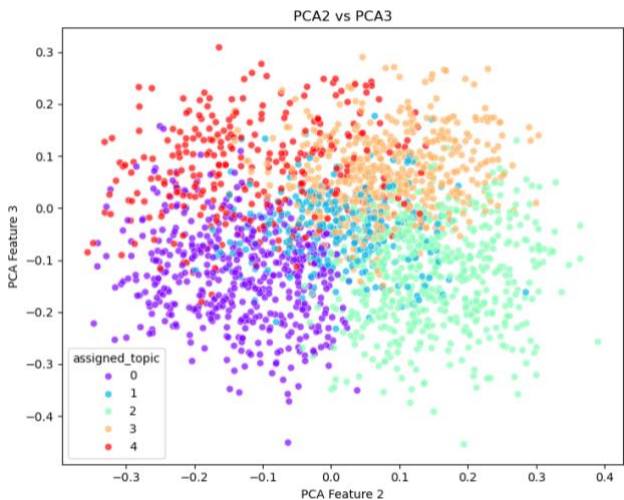


Figure 10 : Analyse sémantique latente. Projection sur les composantes 1 et 3 de l'ACP effectuée sur les résumés encodés en vecteurs avec la méthode InstructorEmbeddings. Les couleurs correspondent aux 5 clusters déterminés par l'algorithme des K-moyennes

Classe	Total nb. articles	Top 20 des termes MeSH
0	473	['Administration, Oral', 'Half-Life', 'Remifentanyl', 'Drug Interactions', 'Central Nervous Systemdrug effects', 'Hemodynamicsdrug effects', 'Sevoflurane', 'Saccadesdrug effects', 'Chromatography, High Pressure Liquid', 'Electric Stimulation']
1	277	['Glucosemetabolism', 'Administration, Intranasal', 'Pregnancy', 'Benzodiazepines', 'Consciousnessdrug effectsphysiology', 'Machine Learning', 'Memorydrug effectsphysiology', 'Affectdrug effectsphysiology', 'Anti-Anxiety Agentspharmacokineticspharmacology', 'Breath Tests']
2	545	['Magnetic Resonance Imagingmethods', 'Fluorodeoxyglucose F18', 'Magnetic Resonance Spectroscopy', 'Reproducibility of Results', 'Magnetoencephalography', 'Follow-Up Studies', 'Aspartic Acidanalogs & derivativesmetabolism', 'Transcranial Magnetic Stimulation', 'Brain', 'Creatinemetabolism']
3	511	['Acoustic Stimulationmethods', 'Evoked Potentials, Auditoryphysiology', 'Evoked Potentialsphysiology', 'Reward', 'Evoked Potentials', 'Evoked Potentials, Auditorydrug effectsphysiology', 'Cognitionphysiology', 'Reaction Time', 'Photic Stimulation', 'Schizophrenic Psychology']
4	288	['Sleep, REMdrug effects', 'Sleepdrug effectsphysiology', 'Sleep Stagesdrug effects', 'Polysomnographydrug effects', 'Sleep, REMphysiology', 'Sleep Stagesdrug effectsphysiology', 'Wakefulnessdrug effects', 'Wakefulnessphysiology', 'Zolpidem', 'Sleep Stagesphysiology']

Table 6 : InstructorEmbeddings. Termes MeSH représentatifs des 5 classes identifiées après partitionnement avec K-means.

Autres méthodes d'encodage des mots

Word2Vec

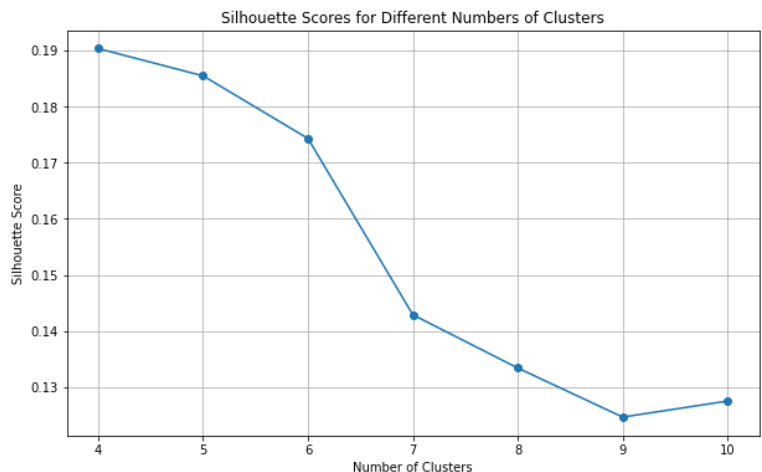


Figure 11 : Word2Vec, coefficient de silhouette obtenu après la partition en n clusters avec l'algorithme des K-moyennes.

Table 7 : Termes MeSH représentatifs des 10 classes identifiées après partitionnement avec K-means.

Classe	Total nb. articles	Top 20 des termes MeSH	Nom attribué	Test de validation 20 articles
0	210	['Rest', 'Restphysiology', 'Alpha Rhythmdrug effectsmethods', 'Alpha Rhythm', 'Alpha Rhythmdrug effectsphysiology', 'Levodopatherapeutic use', 'Delta Rhythm', 'Methioninegenetics', 'Valinegenetics', 'Brain Wavesphysiology', 'Hand Strengthphysiology', 'Brain', 'Cerebral Cortexphysiology', 'Magnetic Resonance Imaging', 'Alcohol Drinking', 'Action Potentialsdrug effectsphysiology', 'Magnetoencephalography', 'Antipsychotic Agentsadministration & dosagetherapeutic use', 'Brain Wavesgenetics', 'Brainphysiologyphysiopathology']	Drugs EEG effects	30%
1	221	['Magnetic Resonance Imaging', 'Evoked Potentials, Auditoryphysiology', 'Auditory Perceptionphysiology', 'Fluorodeoxyglucose F18', 'Sertralinetherapeutic use', 'Antidepressive Agentstherapeutic use', 'Schizophreniadrug therapyphysiopathology', 'Retrospective Studies', 'Radiopharmaceuticals', 'Cognitive Behavioral Therapy', 'Prefrontal Cortexphysiopathology', 'Bipolar Disorderphysiopathology', 'Epilepsy, Temporal Lobediagnostic imagingsurgery', 'Antipsychotic Agentstherapeutic use', 'Cerebral Cortexphysiopathology', 'Aspartic Acidanalogs & derivativesanalysis', 'Acoustic Stimulationmethods', 'Loudness Perceptionphysiology', 'Schizophreniaphysiopathology', 'Antiparkinson Agentstherapeutic use']	Psychosis and schizophrenia	30%

2	371	['Models, Neurological', 'Polymorphism, Single Nucleotide', 'Brainphysiology', 'Attentionphysiology', 'Alleles', 'Psychomotor Performancephysiology', 'Gene Frequency', 'Evoked Potentialsgenetics', 'Ethanoladministration & dosage', 'Catechol O-Methyltransferasegenetics', 'Genetic Variation', 'Dopaminemetabolism', 'Inhibition, Psychological', 'Evoked Potentials', 'Photic Stimulationmethods', 'Cognitionphysiology', 'Reaction Time', 'Citaloprampharmacology', 'Cues', 'Psychomotor Performancedrug effectsphysiology']	Monoamines effect on ERP	35%
3	151	['Hot Temperatureadverse effects', 'Cerebrovascular Circulation', 'Deoxyglucoseanalogs & derivatives', 'Magnetics', 'Phenylethyl Alcoholadministration & dosage', 'Carbamazepinetherapeutic use', 'Xenon Radioisotopes', 'Anticonvulsantstherapeutic use', 'Epilepsy, Temporal Lobemetabolism', 'Human Growth Hormonebloodmetabolism', 'Inositolmetabolism', 'Magnetic Resonance Spectroscopy', 'Neural Conductionphysiology', 'Phenytointherapeutic use', 'Tidal Volume', 'tau Proteinsmetabolism', 'Aspartic Acidanalogs & derivativesmetabolism', 'Electroencephalographydrug effectsstatistics & numerical data', 'Linear Models', 'Magnetic Resonance Imaging']	-	
4	159	['Polysomnographydrug effects', 'Sleepdrug effectsphysiology', 'Hydrocortisoneblood', 'Sleep Stagesdrug effectsphysiology', 'Wakefulnessdrug effects', 'Sleep Stages', 'Sleep, REMdrug effectsphysiology', 'Circadian Rhythmphysiology', 'Adrenocorticotrophic Hormoneblood', 'Azabicyclo Compounds', 'Circadian Rhythmdrug effects', 'Zolpidem', 'Sleep', 'Wakefulnessphysiology', 'Growth Hormoneblood', 'Human Growth Hormonemetabolism', 'Sodium Oxybatepharmacology', 'Piperazinespharmacology', 'Sleep, REM', 'Accidents, Traffic']	Sleep	95%
5	136	['Evoked Potentials, Auditory, Brain Stem', 'Gabapentin', 'Acetatespharmacologytherapeutic use', 'Amines', 'Chinaepidemiology', 'Consciousness Monitorsstatistics & numerical data', 'Cyclohexanecarboxylic Acids', 'DNA Primersgenetics', 'Epilepsy, Temporal Lobediagnosis', 'Temporal Lobediagnostic imagingmetabolism', 'Thiazolespharmacology', 'Neuropsychological Testsstatistics & numerical data', 'Retrospective Studies', 'Acoustic Stimulationmethodpsychology', 'Association Learningdrug effects', 'Cerebral Cortexdiagnostic imagingmetabolism', 'Copperblood', 'Electroencephalographystatistics & numerical data', 'Fructoseanalogs & derivativespharmacology', 'Mental Status Schedule']	-	
6	160	['Remifentanyl', 'Anesthesia, Intravenousmethods', 'Anesthetics, Intravenousadministration & dosage', 'Piperidinespharmacokineticspharmacology', 'Propofoladministration & dosage', 'Propofolpharmacology', 'Anti-Anxiety Agentspharmacokineticspharmacology', 'Models, Biological', 'Drug Interactionsphysiology', 'Temazepamadministration & dosagepharmacology', 'ATP Binding Cassette Transporter, Subfamily B', 'Analgesiamethods', 'Analgesics, Opioidadverse effectspharmacology', 'Benzodiazepinesadministration & dosageblood', 'Cold Temperatureadverse effects', 'Cytochrome P-450 Enzyme Inhibitors', 'Epilepsydiagnosisphysiopathology', 'Lighting', 'Paclitaxeladverse effectstherapeutic use', 'Painchemically induceddrug therapy']	Aneshtesia and sedatives	30%
7	158	['Consciousnessdrug effectsphysiology', 'Evoked Potentials, Auditoryphysiology', 'Reward', 'Reaction Time', 'Evoked Potentials, Motordrug effects', 'Acoustic Stimulationmethods', 'Nicotinepharmacology', 'Evoked Potentials', 'Nicotinic Agonistspharmacology', 'Adaptation, Physiologicaldrug effectsphysiology', 'Antipsychotic Agentsbloodpharmacology', 'Dopamine Agentspharmacologytherapeutic use', 'Esophagusphysiopathology', 'Gene-Environment Interaction', 'Nicotine', 'Nicotinic Antagonistspharmacology', 'Noise', 'Phenylalanineblooddeficiency', 'Phenylalaninedeficiency', 'Schizophrenia, Paranoiddiagnosisphysiopathology']	-	

8	324	['Pain Measurement', 'Fentanyl', 'Task Performance and Analysis', 'Nitrous Oxide', 'Pursuit, Smoothdrug effects', 'Propofol', 'Capsaicin', 'Monitoring, Physiologic', 'Piperidinesadverse effectspharmacology', 'Sensationdrug effects', 'Pain Thresholdphysiology', 'Hydrocortisoneblood', 'Administration, Intranasal', 'Evoked Potentials, Somatosensorydrug effects', 'Stimulation, Chemical', 'Bromazepampharmacology', 'Prolactinblood', 'Cycloserinepharmacology', 'Diagnosis, Computer-Assistedmethods', 'Schizophreniadrug therapy']	Pain	20%
9	204	['Half-Life', 'Area Under Curve', 'Drug Interactions', 'Triazolampharmacology', 'Chromatography, High Pressure Liquid', 'Metabolic Clearance Rate', 'Central Nervous Systemdrug effects', 'Midazolamadministration & dosagepharmacokineticspharmacology', 'Postural Balancedrug effects', 'Beta Rhythmdrug effects', 'Wakefulnessdrug effects', 'Electrocardiographdrug effects', 'Fluoroquinolones', 'Piperazinesadministration & dosageadverse effectspharmacokinetics', 'Quinolonesadministration & dosageadverse effectspharmacokinetics', 'Tablets', 'Triazolespharmacology', 'Drug Administration Schedule', 'Hemodynamicsdrug effects', 'Saccadesdrug effects']	Drugs pharmacodynamic	60%

Small Bert 128

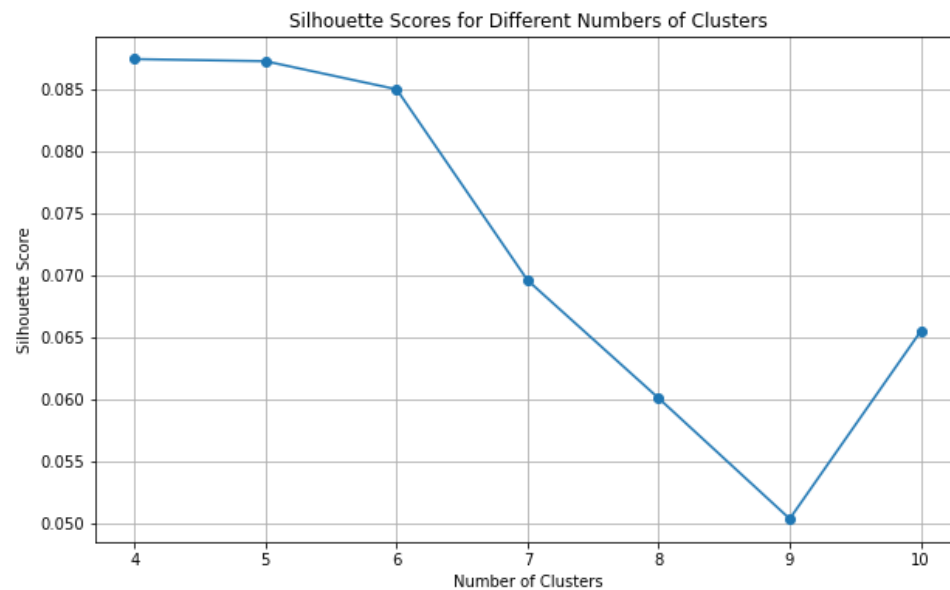


Figure 12 : SmallBert. Coefficient de silhouette obtenu après la partition en n clusters avec l'algorithme des K -moyennes.

Table 8 : *SmallBert. Termes MeSH représentatifs des 10 classes identifiées après partitionnement avec K-means.*

Classe	Total nb. articles	Top 20 des termes MeSH
0	92	[Epilepsy, Temporal Lobediagnosisphysiopathology', 'Epilepsy, Temporal Lobediagnosisdiagnostic imaging', 'Fluorodeoxyglucose F18', 'Dominance, Cerebralphysiology', 'Magnetic Resonance Spectroscopy', 'Aspartic Acidanalog & derivativesmetabolism', 'Creatinemetabolism', 'Cyclohexanes', 'Energy Metabolismphysiology', 'Epilepsy, Temporal Lobediagnosis', 'Frontal Lobediagnosisdiagnostic imaging', 'Glasgow Coma Scale', 'Neural Pathwaysdiagnostic imaging', 'Neurosurgical Procedures', 'Prevalence', 'Receptor, Serotonin, 5-HT1Ametabolism', 'Magnetic Resonance Spectroscopy', 'Cholinemetabolism', 'Retrospective Studies', 'Epilepsy, Temporal Lobemetabolism']
1	329	[Psychomotor Performancephysiology', 'Prefrontal Cortexphysiology', 'Models, Neurological', 'Methylphenidatepharmacology', 'Memory, Short-Termphysiology', 'Photic Stimulation', 'Transcranial Magnetic Stimulation', 'Conflict, Psychological', 'Neural Inhibitionphysiology', 'Magnetic Resonance Imaging', 'Executive Functionphysiology', 'Transcranial Magnetic Stimulation', 'Motor Cortexdrug effectsphysiology', 'Polymorphism, Single Nucleotide', 'Evoked Potentials', 'Motor Cortexphysiology', 'Norepinephrinephysiology', 'Magnetoencephalography', 'Gamma Rhythmphysiology', 'Beta Rhythmdrug effectsphysiology']
2	311	[Sleep, REMdrug effects', 'Azabicyclo Compounds', 'Sleep Stagesdrug effects', 'Drug Administration Schedule', 'Half-Life', 'Triazolampharmacology', 'Central Nervous Systemdrug effects', 'Beta Rhythmdrug effects', 'Ethanolpharmacology', 'Saccadesdrug effects', 'Chromatography, High Pressure Liquid', 'Anti-Anxiety Agentspharmacology', 'Delta Rhythmdrug effects', 'Mental Recalldrug effects', 'Zolpidem', 'Metabolic Clearance Rate', 'Flunitrazepampharmacology', 'Migraine Disordersphysiopathology', 'Postural Balancedrug effects', 'Capsaicin']
3	213	[Anesthesia, General', 'Nitrous Oxideadministration & dosage', 'Artificial Intelligence', 'China', 'Caffeinepharmacology', 'Artifacts', 'Analgesics, Opioidadministration & dosage', 'Anesthetics, Inhalationadministration & dosage', 'Anestheticspharmacology', 'Conscious Sedation', 'Diagnosis, Computer-Assisted', 'Electroencephalographyinstrumentation', 'Infusion Pumps', 'Propofoladministration & dosage', 'Reward', 'Anesthetics, Intravenouspharmacology', 'Electrodes', 'Machine Learning', 'Accelerometryinstrumentation', 'Endpoint Determination']
4	172	[Sleep Stagesdrug effects', 'Sleep, REMdrug effects', 'Polysomnographydrug effects', 'Circadian Rhythmdrug effects', 'Wakefulnessphysiology', 'Polysomnographymethods', 'Wakefulnessdrug effects', 'Sleep, REM', 'Sleep Stages', 'Sleep Stagesdrug effectsphysiology', 'Wakefulnessdrug effectsphysiology', 'Arousal', 'Circadian Rhythmphysiology', 'Adrenocorticotrophic Hormoneblood', 'Circadian Rhythm', 'Sleep Deprivationphysiology', 'Health', 'Sleep, REMphysiology', 'Automobile Driving', 'Growth Hormoneblood']
5	213	[Ligands', 'Macaca mulatta', 'Mifepristonepharmacology', 'Pregnanoloneblood', 'Receptors, GABA-Adrug effectsmetabolism', 'Receptors, Muscarinicmetabolism', 'Area Under Curve', 'Rats', 'Flumazenilpharmacology', 'Piperazinespharmacology', 'Chromatography, Liquid', 'Cytochrome P-450 Enzyme Inhibitors', 'Epilepsyblood', 'Fatty Acids, Nonesterifiedblood', 'Ghrelin', 'Glucose Clamp Technique', 'Manganeseblood', 'Megestrol Acetatepharmacology', 'Melatoninmetabolism', 'Nicotine']
6	215	[Schizophreniadrug therapyphysiopathology', 'Bipolar Disorderphysiopathology', 'Electromyographymethods', 'Quality of Life', 'Recognition, Psychologydrug effects', 'Loudness Perceptionphysiology', 'Nicotinepharmacology', 'Heroin', 'Quetiapine Fumarate', 'Nicotinic Agonistspharmacology', 'Antipsychotic Agentsbloodpharmacology', 'Cognitive Dysfunctionetiology', 'Dibenzothiazepinespharmacologytherapeutic use', 'Esophagusphysiopathology', 'Estradiolmetabolism', 'Fingers', 'Fluoxetine therapeutic use', 'Gene-Environment Interaction', 'Receptors, Serotoninmetabolism', 'Walkingphysiology']
7	128	[Midazolampharmacokineticspharmacology', 'ATP Binding Cassette Transporter, Subfamily B', 'GABA Agonists therapeutic use', 'Isoxazole therapeutic use', 'Paclitaxel adverse effectstherapeutic use', 'Models, Biological', 'Anti-Anxiety Agentspharmacokineticspharmacology', 'Metabolic Clearance Rate', 'Anesthetics, Intravenousadministration & dosage pharmacokinetics', 'Antineoplastic Combined Chemotherapy Protocolsadverse effectstherapeutic use', 'Cognition Disorderschemically induced', 'Paindrug therapy', 'Phytotherapy', 'Plant Extractsadministration & dosage pharmacology', 'Benzodiazepines', 'ATP Binding Cassette Transporter, Subfamily B, Member 1 antagonists & inhibitorsmetabolism', 'Abnormalities, Drug-Inducedpsychology', 'Absorption', 'Acyclic Monoterpenes', 'Adenocarcinoma, Clear Celldrug therapypsychology']
8	174	[Genetic Predisposition to Disease', 'Gene Frequency', 'Alleles', 'Chi-Square Distribution', 'Phenotype', 'Polymorphism, Single Nucleotidegenetics', 'Serotonin Plasma Membrane Transport Proteinsgenetics', 'Event-Related Potentials, P300genetics', 'Genome-Wide Association Study', 'Anesthetics, Intravenouspharmacokineticspharmacology', 'Base Sequence', 'Bipolar Disordergenetics', 'Brain Wavesgenetics', 'Brainmetabolismphysiology', 'Central Nervous System Stimulantsadverse effects', 'Epilepsy, Complex Partialmetabolism', 'Extraversion, Psychological', 'Fathers', 'Haloperidoladverse effectstherapeutic use', 'Quinolonespharmacology']
9	247	[Risk Factors', 'Selective Serotonin Reuptake Inhibitortherapeutic use', 'Cognitive Behavioral Therapy', 'Delta Rhythm', 'Emotionsphysiology', 'Succinimides', 'Frontal Lobephysiopathology', 'Retrospective Studies', 'Attention Deficit Disorder with Hyperactivitydrug therapy', 'Occipital Lobephysiopathology', 'Parasympathomimetics', 'Biomarkersblood', 'Longitudinal Studies', 'Antipsychotic Agentsadverse effectstherapeutic use', 'Brain Stemphysiopathology', 'Cannabidiolpharmacology', 'Cognition Disorderschemically induceddiagnosis', 'Cognition Disordersphysiopathologypsychology', 'Cognitive Behavioral Therapymethods', 'Epilepsy, Temporal Lobediagnosis imagingmetabolismpathology']

Glove

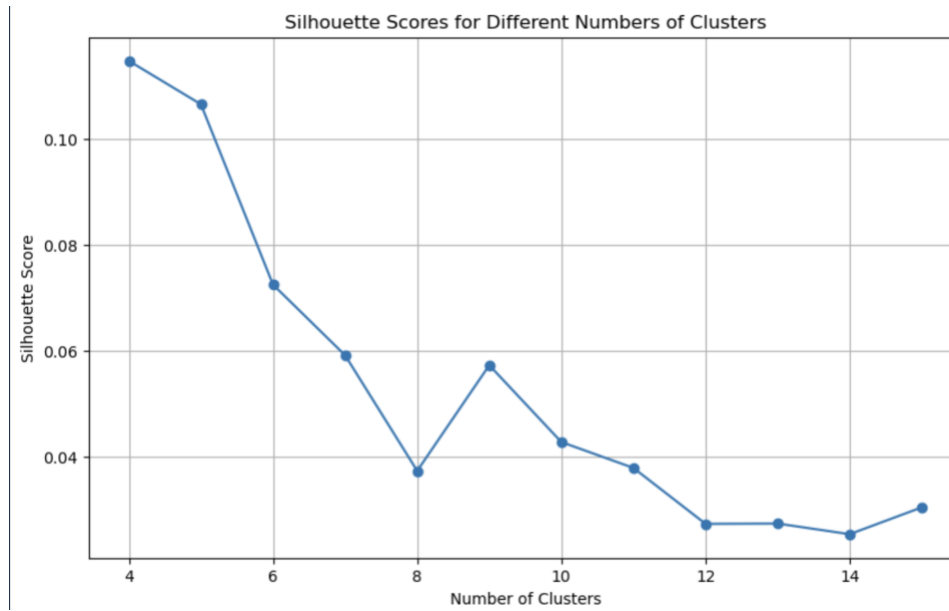


Figure 13 : Glove. Coefficient de silhouette obtenu après la partition en n clusters avec l'algorithme des K -moyennes.

Table 9 : Glove. Termes MeSH représentatifs des 9 classes identifiées après partitionnement avec K-means.

Classe	Total nb. articles	Top 10 des termes MeSH
0	352	['Blood Pressuredrug effects', 'Infusions, Intravenous', 'Placebos', 'Administration, Oral', 'Braindrug effects', 'Hydrocortisoneblood', 'Half-Life', 'Wakefulnessdrug effects', 'Drug Administration Schedule', 'Adrenocorticotrophic Hormoneblood']
1	450	['Brainphysiology', 'Evoked Potentialsphysiology', 'Psychomotor Performancephysiology', 'Evoked Potentialsdrug effects', 'Evoked Potentials', 'Magnetoencephalography', 'Attentiondrug effects', 'Reward', 'Reaction Timephysiology', 'Attentionphysiology']
2	233	['Schizophreniadrug therapyphysiopathology', 'Schizophrenic Psychology', 'Antipsychotic Agentstherapeutic use', 'Consciousnessdrug effectsphysiology', 'Propofoladministration & dosage', 'Sex Factors', 'Age Factors', 'Anesthetics, Intravenouspharmacology', 'Sleep', 'Anesthesia']
3	322	['Tomography, Emission-Computed', 'Positron-Emission Tomography', 'Functional Laterality', 'Transcranial Magnetic Stimulationmethods', 'Evoked Potentials, Somatosensoryphysiology', 'Brainmetabolism', 'Glucosemetabolism', 'Serotoninmetabolism', 'gamma-Aminobutyric Acidmetabolism', 'Antiparkinson Agentstherapeutic use']
4	168	['Fluorodeoxyglucose F18', 'Transcranial Magnetic Stimulation', 'Aspartic Acidanalogs & derivativesmetabolism', 'Anticonvulsantspharmacology', 'Creatinemetabolism', 'Glutaminemetabolism', 'Magnetic Resonance Spectroscopymethods', 'Electroencephalographystatistics & numerical data', 'Positron-Emission Tomographymethods', 'Theta Rhythmdrug effectsphysiology']
5	1	['Anthropometry', 'Diethylcarbamazinetherapeutic use', 'Epilepsy, Frontal Lobedrug therapyphysiopathology', 'Lipoxygenase Inhibitorstherapeutic use']
6	217	['Psychiatric Status Rating Scales', 'Retrospective Studies', 'Risk Factors', 'Selective Serotonin Reuptake Inhibitorstherapeutic use', 'Antidepressive Agentstherapeutic use', 'Brain Wavesphysiology', 'Pregnancy', 'Cognitive Behavioral Therapy', 'Incidence', 'Orexins']
7	68	['Benzodiazepinespharmacology', 'Epilepsy, Temporal Lobediagnostic imaging', 'Epilepsymetabolism', 'Pedigree', 'Positron Emission Tomography Computed Tomography', 'Receptors, N-Methyl-D-Aspartateantagonists & inhibitors', 'Receptors, Nicotinicgenetics', 'Seizures', 'Spasms, Infantilediagnosisdrug therapyphysiopathology', 'Thalamusdiagnostic imaging']
8	283	['Sleepdrug effects', 'Sleep, REMdrug effects', 'Sleepdrug effectsphysiology', 'Arousaldrug effects', 'Affectdrug effects', 'Polysomnographdrug effects', 'Sleep, REMphysiology', 'Central Nervous System Stimulantspharmacology', 'Polysomnographymethods', 'Alpha Rhythm']

Extraction de mots-clés (keywords) puis LDA

Étape d'extraction des mots clés incluse dans la pipeline de traitement Spark des résumés.

Step 4: Keyword Extraction

```
keywords = YakeKeywordExtraction() \
    .setInputCols("token") \
    .setOutputCol("keywords") \
    .setMinNGrams(1) \
    .setMaxNGrams(2) \
    .setNKeywords(10) \
    .setStopWords(StopWordsCleaner().getStopWords())
```

Table 10: Termes représentatifs des 10 classes identifiées avec la LDA.

Classe	10 termes les plus représentatifs
0	['sleep', 'group', 'dopamine', 'rem sleep', 'gating', 'system', 'sleep eeg', 'functional', 'individuals', 'eeg power']
1	['mg', 'propofol', 'auditory', 'cortical', 'receptor', 'visual', 'concentrations', 'evoked', 'analysis', 'disorder']
2	['brain', 'treatment', 'pain', 'pd', 'attention', 'doses', 'blood', 'dose', 'healthy controls', 'related']
3	['eeg', 'effects', 'activity', 'cognitive', 'changes', 'alpha', 'effect', 'alcohol', 'cortex', 'drugs']
4	['subjects', 'beta', 'depression', 'children', 'drug', 'stimulation', 'acute', 'eeg', 'whether', 'eeg changes']
5	['performance', 'task', 'h', 'motor', 'temporal', 'clinical', 'neural', 'stimuli', 'function', 'studies']
6	['caffeine', 'memory', 'controls', 'sedation', 'ad', 'sensory', 'processing', 'central', 'however', 'heart rate']
7	['power', 'response', 'frequency', 'levels', 'serotonergic', 'disorders', 'anesthesia', 'eeg power', 'theta', 'p']
8	['p', 'study', 'placebo', 'rem', 'increased', 'amplitude', 'control', 'administration', 'midazolam', 'mmn']
9	['patients', 'schizophrenia', 'theta', 'healthy', 'epilepsy', 'oscillations', 'volunteers', 'used', 'responses', 'healthy subjects']