



**HAL**  
open science

## La lexicométrie : un outil efficient pour l'analyse du discours

Hadria Khelifi

► **To cite this version:**

Hadria Khelifi. La lexicométrie : un outil efficient pour l'analyse du discours. Colloque international Médias numériques, langues, discours, pratiques et interculturalité, Université Paris 8 Vincennes-Saint-Denis (France); Université Grenoble Alpes (France); Université Ibn Zohr (Maroc), Oct 2016, Agadir, Maroc. pp.151-161. hal-01730406

**HAL Id: hal-01730406**

**<https://hal.univ-lorraine.fr/hal-01730406>**

Submitted on 13 Mar 2018

**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.

# **La lexicométrie : un outil efficient pour l'analyse du discours**

**Hadria KHELIFI**

**Doctorante à université de Lorraine**

[hadriakhelifi@gmail.com](mailto:hadriakhelifi@gmail.com)

## **Résumé**

Cet exposé s'inscrit dans la perspective de l'analyse du discours ayant un format numérique. Il est destiné, prioritairement, aux jeunes chercheurs qui étudient leur corpus de façon traditionnelle dans différents domaines notamment en discours et langue car la recherche statistique et/ou automatique en sciences sociales est une ancienne tendance.

Aujourd'hui, il existe des logiciels informatiques qui prolongent avantageusement certaines lacunes des méthodologies classiques de la recherche scientifique. Leur usage semble toutefois restreint. S'agit-il d'un déficit de performance ou de difficulté d'interprétation des résultats, à titre d'exemple ? Les chercheurs estiment-ils que les méthodes de recherche et d'exégèse classique donnent amplement de satisfaction ?

## **Mots clés**

Analyse du discours, interprétation, lexicométrie, logiciel, *Hyperbase*.

## **Introduction**

Depuis le passage à l'écrit numérique, le traitement des textes via des logiciels impacte de plus en plus la méthodologie de la recherche dans différentes disciplines des sciences humaines. La raison de l'usage de ces outils vient de la nécessité d'analyser des corpus volumineux, ce qui requiert une quantité considérable de travail. Ces mêmes outils sont également valables pour explorer et analyser de petites de documents (articles de revues, des livres, etc.) et avoir divers résultats, ce qui rend leur utilisation très alléchante.

Le logiciel assure un résultat hautement performant, satisfaisant et beaucoup plus ordonné qu'un travail manuel étant exécuté via une entité informatique intelligente. En effet, les logiciels sont connus par leur efficacité à permettre au chercheur de retourner facilement au

texte qui constitue le corpus par un simple clic et à partir de n'importe quelle fonction. Il est ainsi possible de faire appel à des phrases, des paragraphes pour vérifier l'emplacement des formes, lire aisément et comprendre leur contexte. Mais cette approche par logiciel n'est pas exempte de difficultés telles que la collecte des données qui doivent être à la fois larges et diversifiées, ou les problèmes techniques (compétence à utiliser le logiciel) que peut rencontrer l'utilisateur.

Malgré son apparition ancienne, l'exploitation de cette démarche automatique du corpus reste limitée, la raison pour laquelle de nombreuses journées<sup>1</sup> scientifiques ont été organisées autour du sujet, et des articles ont été publiés en faveur de l'analyse automatique dans divers domaines de recherche.

Grâce aux moyens de communication et d'usage collectif fournis par les nouvelles technologies, la statistique n'est plus réservée au champ scientifique ou économique, elle est aujourd'hui, en usage général dans tous les domaines comme celui du discours universitaire que nous choisissons de traiter à travers l'outil *Hyperbase* dans cette proposition de communication.

## **1. Traitement automatique de textes**

Via l'analyse numérique, la vision du texte et la façon de le lire changent ; nous ne lisons plus ce dernier d'une manière purement linéaire mais plutôt en produisant des liaisons transversales entre les mots fonctionnant en réseau. Selon D. Mayaffre, « le numérique complète-augmente la linéarité par des réticularités »<sup>2</sup>. À travers le moyen informatique, qui peut porter des millions de documents, le texte peut être présenté en tant que mobile, dynamique animant le parcours de la lecture du chercheur en le permettant de se déplacer temporairement et simultanément que le déplacement des données du texte en cas d'une lecture hypertextuelle. Il est par ailleurs judicieux de souligner que l'analyse via les logiciels n'enferme pas l'analyste dans une technique imposée, et ne prétend pas à une substitution à la pensée humaine, mais une orientation de cette dernière vers les différentes interprétations verticales et horizontales du texte.

---

<sup>1</sup> Nous vous conseillons de consulter les JADT (journée internationale d'Analyse statistique des Données Textuelles) qui réunissent des chercheurs travaillant sur l'analyse automatique et statistique du texte tous les deux ans depuis 1990. Ces recherches portent sur des textes de domaines différents et sur la façon d'utiliser le logiciel (préparation du texte, installation du logiciel, possibilités de la lecture des résultats, comparaison entre des logiciels existant...)

<sup>2</sup> Mayaffre D., 2015.

## 2. Linguistique, statistique et informatique

Selon E. Brunet, « la statistique linguistique - et la statistique tout court- ne peut atteindre son plein développement qu'en association avec l'informatique. »<sup>3</sup> À première pensée, nous négligeons toute relation entre la linguistique et la statistique mais si nous réfléchissons aux premiers grands corpus comme celui du T.L.F<sup>4</sup> qui traite une masse de trente millions de personnes, nous nous interrogeons certainement comment son traitement a pu être réalisé en absence de la statistique, avant de le présenter en forme de dictionnaire écrit. Ainsi, comment cela a été informatisé, par la suite ? Tous les savoirs sont donc indissociables depuis toujours.

## 3. Lexicométrie

### 3.1. Historique

Selon M. Tournier, la lexicométrie aussi appelée : logométrie, analyse automatique, « **statistique linguistique** (Guiraud, 1959, 1960), **statistique lexicale** ou **linguistique quantitative** (Muller, 1964, 1967, 1973, 1979), **statistique textuelle** (Salem 1987, 1994), voire **analyse des données en linguistique** (Benzecri 1981), **la lexicométrie** (Tournier 1975, Lafon 1984) n'est pas une théorie mais une méthodologie d'étude du discours, qui se veut exhaustive, systématique est automatisée. »<sup>5</sup>

Cette pratique est un inventaire de mots ou un répertoire linguistique, une alliance de l'informatique, de la statistique et des sciences humaines et sociales, vu que les différentes analyses proposées par cette étude se basent sur le calcul entier de tous les mots du corpus. L'année 1969 a connu par la suite l'apparition de l'Analyse automatique du discours grâce à M. Pêcheux<sup>6</sup>. La lexicométrie politique construite à la fin des années 1960, a connue en France une certaine popularité mais qui s'estompe au fil du temps. Elle est nommée « textométrie »<sup>7</sup> qu'au milieu des années 1990. De nos jours, un déclin alors même dans notre essor caractérisé par des approches technologiques avancées et des ressources matérielles en abondance (ordinateurs, textes numérisés sous divers formats). En plus, de nombreuses versions de différents logiciels sont téléchargées gratuitement sur Internet, des logiciels aux capacités de traitement énormes pour les macros-corpus.

---

<sup>3</sup> Brunet E., 2011, p.40-41.

<sup>4</sup> Trésor de la langue Française, un dictionnaire de la langue française des XIXème et XXème siècle, publié entre 1971 et 1994.

<sup>5</sup> Tournier M., 2002, p. 342-343.

<sup>6</sup> Leimendorfer F., Salem A., 1995, Usage de la lexicométrie en analyse de discours, selon l'adresse : [http://horizon.documentation.ird.fr/exl-doc/pleins\\_textes/pleins\\_textes\\_4/sci\\_hum/41740.pdf](http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_4/sci_hum/41740.pdf), [consulté le 25/01/2016].

<sup>7</sup> Rastier F., Que cachent les « données textuelles » ?, 2008, selon l'adresse : <http://lexicometrica.univ-paris3.fr/jadt/jadt2008/pdf/rastier.pdf>, [consulté le 14/06/16].

Selon L. Lebart et A. Salem, les premiers travaux quantitatifs sur le texte portent sur l'ensemble des unités linguistiques que l'on peut répertorier dans un même texte (phonème, lexème). Quant à la statistique lexicale, en reprenant la nomination de P. Guiraud puis de C. Muller, elle entreprend d'étudier avant tout le vocabulaire des anciens grands auteurs français du 17<sup>ème</sup> siècle<sup>8</sup>. Ils rajoutent que G. Herdan (1964) considère la linguistique statistique comme « une branche de la linguistique structurale, avec pour principale fonction la description statistique du fonctionnement (dans des corpus de textes) des unités définies par le linguiste aux différents niveaux de l'analyse linguistique (phonologique, lexical, phrastique) »<sup>9</sup>. Quant aux recherches récentes effectuées via l'analyse automatique, nous citons à titre d'exemple, le projet Scientext conduit principalement par les équipes pluridisciplinaires suivantes : IDILEM<sup>10</sup>, LiCoRN<sup>11</sup>, LLS<sup>12</sup> compare des écrits scientifiques (thèses, écrit d'étudiants, articles, actes de colloque) dont l'objectif est une étude linguistique du positionnement de l'auteur à travers les marques syntaxiques et énonciatives liées à la causalité.

### **3.2. Fonctionnement lexicométrique**

Afin de réaliser sa tâche, le logiciel chargé de la pratique lexicométrique exige un texte fermé et stabilisé. Nous soulignons qu'il est très important que la donnée de départ (corpus) ne subisse aucune modification ou transformation par rapport au discours tel qu'il est écrit ou dit par son auteur notamment quand il s'agit d'un corpus constitué d'une transcription d'entretiens. À travers sa nomination, « statistique lexicale ou linguistique quantitative » que représente la lexicométrie, C. Muller vise à mesurer la quantité des mots d'un texte, plus précisément les lexèmes qui couvrent toutes les classes (substantifs, temps, modes d'un verbe, etc.<sup>13</sup>). Avant de commencer le traitement, le logiciel découpe le corpus en parties puis s'attèle à décompter le nombre de chaque unité linguistique dans le corpus, à étudier sa fréquence et son occurrence, à définir son évolution, etc.

Quant à l'interprétation des résultats, c'est à l'utilisateur de gérer son analyse de corpus selon l'approche qu'il a choisi. Les résultats fournis par le logiciel orientent l'analyste, notamment s'il est un non-spécialiste. D. Mayaffre, un historien qui a étudié de nombreux

---

<sup>8</sup> Lebart L., Salem A., 1994, p.16.

<sup>9</sup> Ibid., p.17.

<sup>10</sup> Laboratoire de Linguistique et de Didactique des Langues Etrangères et Maternelles de Grenoble

<sup>11</sup> Linguistique de corpus de Lorient.

<sup>12</sup> Langage, Littérature, Société de Chambéry.

<sup>13</sup> Muller C., 1992, p.10.

discours politiques, reconnaît que lui-même ne tient ni à la méthode binomiale<sup>14</sup> ni le modèle hypergéométrique en statistique lexicale, et que ses connaissances linguistiques ne suffisent pas pour nourrir le débat épistémologique sur le lien entre Histoire et Linguistique dans l'analyse du discours<sup>15</sup> et pourtant, la plupart de ses analyses se basent sur des recherches statistiques obtenu à l'aide des logiciels d'analyse linguistique. En effet, c'est le logiciel qui, en offrant diverses pistes de recherche, suite à ses différentes fonctionnalités et selon les composantes du discours (verbe, lemme, nom propre, adjectif, adverbe, ponctuation, mots outils, etc.), orientent le chercheur vers les différentes manières d'interpréter les résultats. C'est la méthode lexicométrique qui organise le travail du chercheur en proposant des analyses distinctes. En d'autres termes, le logiciel apprête le texte au chercheur en le triant, calculant les différents profils qu'il suppose être cherchés, organisant toutes les données sous forme de tableaux, graphiques ou histogramme mais en n'imposant rien à l'analyste en ce qui concerne sa lecture des résultats, sa créativité ou son esprit critique.

Il existe divers logiciels pour effectuer une analyse de discours très longs notamment : *Alceste* (de M. Reinert), *Lexico 1 et 2* (de A. Salem), *Tropes* (de P. Molette et A. Landré), *Decision Explorer* (de S. Heiden), *NVivo* (de QSR<sup>16</sup>), *Hyperbase* (par E. Brunet).

## **4. Exemple d'analyse via le logiciel *Hyperbase***

### **4.1. *Hyperbase***

*Hyperbase*<sup>17</sup> est un outil d'analyse statistique et graphique à partir duquel il est possible de définir la quantité de l'écrit, calculer la haute fréquence des éléments utilisés dans tout le corpus, la distance lexicale entre les textes du corpus, la richesse du vocabulaire et l'état de son évolution, la typologie du discours, les thèmes transversaux du corpus, la prise de position par les scripteurs-étudiants, la cohésion et la cohérence, etc. Ainsi, il permet de comparer un texte à d'autres textes ou une comparaison interne selon l'objectif du chercheur.

Grace aux fonctions proposées par ce logiciel, l'angle de vision de l'analyste par rapport au texte change. Ce dernier n'est plus qu'une suite linéaire de phrases étudiées telles quelles selon leur position dans le texte, ou une suite linéaire de mots analysés selon leur ordre dans

---

<sup>14</sup> Une loi qui spécifie le nombre de succès obtenus lors de la répétition de plusieurs unités identiques. Consulter : Muller C., 1992, *Initiation aux méthodes de la statistique linguistique*, p.42.

<sup>15</sup> Mayaffre D., 2000, p.749.

<sup>16</sup> *Quality Systems Registrars*.

<sup>17</sup> Figure parmi les logiciels disponibles pour la plateforme Windows et sous Mac OS comme *Alceste* et *Lexico*.

la phrase (sujet – verbe – complément), ce même ordre qui leur gratifie du sens, et fait du mot un « objet réticulaire »<sup>18</sup>, c'est-à-dire une construction textuelle qui fonctionne en réseau. Cette structure se construit à partir du vocabulaire situé d'une manière désorganisée dans le texte mais qui peut être porteur du sens, créateur de logique et de cohérence. *Hyperbase* a été utilisé pour l'étude de nombreux corpus de la littérature française (le Clézio, Hugo, Sand, Rimbaud, Aragon, du Bellay...), la base Balzac, Rabelais, preuve de son efficacité et de sa pertinence. Afin d'illustrer au mieux l'usage de ce logiciel, nous vous proposons ci-dessous deux exemples dont l'un concerne le positionnement du scripteur-étudiant à travers son texte et l'autre établit une recherche thématique au sein du discours universitaire.

## 5. Traces du scripteur dans le discours universitaire

Dans le cadre d'une recherche sur le discours universitaire écrit, nous avons choisi d'étudier la prise en charge de l'énonciation par le scripteur-étudiant à travers l'usage des pronoms personnels *je* et *nous*. Le rédacteur exprime ses différents points de vue en prenant parfois en charges ses propos à travers les pronoms *je* ou *nous*. Le tableau ci-dessous montre le chiffre de ces deux pronoms.

Discipline	Je	Nous	Discipline	Je	Nous
Anthropologie	11	92	Gestion	10	100
Architecture		148	Mathématiques		51
Bibliothéconomie		53	Médecine vétérinaire		59
Didactique	39	134	Psychologie	55	215
Éducation sportive		102	Science alimentaire		112
Informatique	1	139	Texte littéraire	20	117

**Tableau 1 :** Fréquence des pronoms *je*, *nous*,

Selon le tableau ci-dessus, le pronom sujet *je* se montre 136 fois, un chiffre beaucoup moins que celui de *nous* qui s'élève à 1322 usages. *Je* est complètement absent dans certaines disciplines qui constituent le corpus comme : l'architecture, la bibliothéconomie, l'éducation sportive, les mathématiques, la médecine vétérinaire, la science alimentaire. Les auteurs de ces discours se sont déresponsabilisés de toutes les informations apportées dans leurs écrits

---

<sup>18</sup> Mayaffre D., 2015.

car le principe du *je* est d'« assumer la responsabilité. »<sup>19</sup>. L'absence totale de ce pronom confirme qu'il est sûrement masqué derrière les pronoms *nous* et *on* (que nous ne traitons pas dans cet exemple). La présence et/ ou l'absence de *je* et *nous* peut avoir des raisons diverses :

### **5.1.1. Contenu des différentes parties du mémoire de fin d'étude**

Il est possible que cette réduction dans l'emploi de *je* soit liée au contenu des différentes parties composant le mémoire de fin d'étude. Par exemple, dans le développement (la plus longue partie) ou la conclusion, l'implication du rédacteur est moins imposante que dans l'introduction.

### **5.1.2. Typologie du texte**

Nous supposons que la typologie du texte est concernée par l'usage de *je* ou de *nous*. Par exemple, dans un texte expressif ou narratif (quand le narrateur fait partie de l'histoire), l'emploi du *je* ou *nous* est prépondérant.

### **5.1.3. Spécificité des disciplines à l'université**

Le *je* apparaît en progression dans la plupart des discours en sciences humaines. Dans les discours nommés durs et économiques, le pronom *je* est absent. Là aussi, nous pouvons dire qu'il est courant que les spécialités connues comme scientifiques posent immédiatement la problématique, discutent directement le sujet et proposent aussitôt des solutions. L'éloquence et la manifestation du soi sont rares chez l'étudiant des sciences dures contrairement à un rédacteur en sciences humaines qui rédige son texte en le traçant une finalité esthétique, s'impliquant parfois involontairement.

À vrai dire, nous ne pouvons pas définir les divers rôles de deux pronoms à partir du graphique seul. Nous ne savons pas, à titre d'exemple, quel *je* (« *je* acteur, *je* chercheur, *je* réflexif, *je* intime »<sup>20</sup>) a été utilisé par le scripteur, mais à l'aide du logiciel *Hyperbase*, nous avons pu rapidement chercher les différents emplois du *je* comme dans le contexte ci-dessous :

---

<sup>19</sup> Reutner U., 2010, *De nobis ipsis silemus ?* Les marques de personne dans l'article Scientifique, U. Reutner, In *Énonciation et rhétorique dans l'écrit scientifique*, selon l'adresse : <http://lidil.revues.org/3013>

<sup>20</sup> Deronne C., 2001, p.149-150.



L'anthropologie du développement ( l' anthropologie dynamique ) s' intéresse aux représentations qui dominent chez les différents acteurs concernés tel que la vision de la population locale sur les agents de développement . Les représentations sociales des uns et des autres sont en effet une donnée de base pour comprendre les stratégies de chacun des acteurs sociaux et c' est pourquoi , **je** tente dans ma recherche d' explorer les présentations des agents de développement rural par la population locale .

Le représentant le plus important de ce courant rénovateur est l' anthropologue français George Balandier , qui a travaillé en Afrique , comme en témoignent ses ouvrages : « sociologie actuelle de l' Afrique Noire ( 1955 ) , Afrique Ambigüe ( 1957 ) , Sens et puissance ( 1974 ) » qui sont à l' origine de la notion « situation coloniale » et qui a tourné son regard vers les changements qui secouent ces sociétés . Il s' est intéressé à la dynamique du dedans et du

**Figure 1** : contexte de *je* dans le mémoire *anthropologie*

#### **5.1.4. Usage volontaire et/ ou involontaire des pronoms *je* et *nous***

Certains auteurs veillent à utiliser un seul pronom, le *je*, à titre d'exemple, parmi les trois pronoms personnels connus pour marquer leur présence, mais ils peuvent glisser aux deux autres pronoms (*nous*, *on*) soit par inadvertance ou pour un usage métadiscursif, soit volontairement. Dans ce dernier cas, le locuteur peut choisir implicitement ou complètement de se décharger de son énoncé en ayant un statut neutre, ou bien partiellement en cas de manque d'information sur le sujet en question.

### **6. Étude thématique**

En se basant sur *Hyperbase* et à partir des mots les plus fréquents, nous avons choisi un mot-pivot (*travail*) pour devenir un thème. Ce dernier sera défini en lien avec tous les mots qui l'entourent. Nous choisissons comme mot-pôle les lemmes et non pas les verbes, les adjectives ou les adverbes parce que les lemmes sont les plus performants<sup>21</sup>.

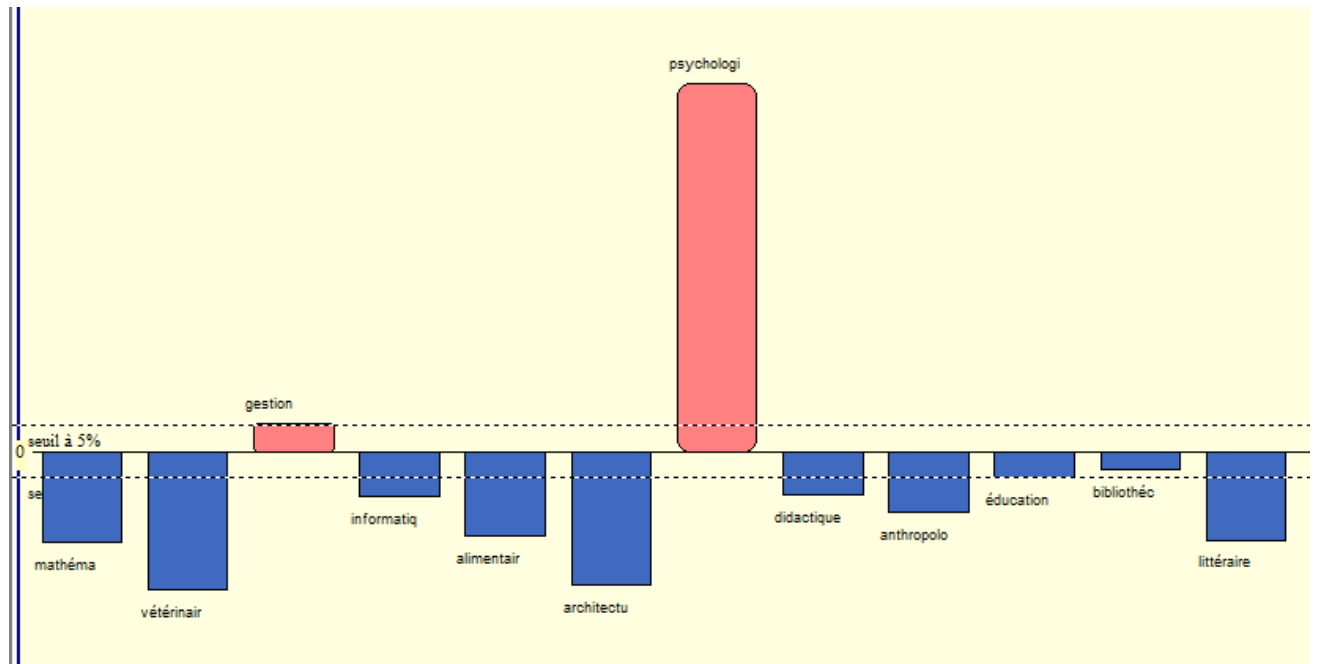
#### **6.1. Le thème *travail***

Nous avons privilégié *travail* comme mot-pôle étant la prochaine occupation des étudiants après leurs études universitaires pourtant il n'est pas le premier lemme figurant sur la liste des hautes fréquences. Afin d'obtenir de meilleurs résultats et plus de détails sur ce thème, et dans l'objectif d'exposer les avantages et les fonctions proposées par *Hyperbase*, nous avons décidé d'étudier *travail* via plusieurs ongles :

##### **6.1.1. Étude de thème *travail* via les fréquences**

---

<sup>21</sup> Brunet E., 2011, *Hyperbase*, manuel de référence, p.44, selon : <http://ancilla.unice.fr/bases/manuel.pdf> [consulté le 04/01/2016].



**Figure 2 :** graphique du mot *travail*

Selon le graphique ci-dessous, *travail* est présent dans tous les discours, soit ceux qui remontent comme la *psychologie* (traite le sujet du stress au milieu professionnel) et la *gestion*, soit ceux qui appariassent en déclin. Pourquoi ces occurrences massives de tel terme? Probablement parce que ces étudiants pensent déjà au terrain de l'emploi, après avoir quitté l'université ou plutôt son usage polyvalent lui permet de remplacer de nombreuses formes tel que : fonction, mission, poste, affaire, action, etc. Le grand chiffre des occurrences du mot *travail* est probablement dû aussi à son usage associé à certains substantifs (plan du travail), ou des adjectifs (bon travail, travail collectif...) pour designer autres expressions.

### 6.1.2. Recherche via les occurrences du mot *travail*

Cette recherche sur le mot *travail* apporte les informations les plus performantes dans une analyse thématique. Ci-dessous, un lot extrait du tableau exposant l'environnement du *travail*.

Écart	Mot	Écart
10,11	Travail	4,9
10,11	Condition	4,79
10,11	Charge	4,77
10,11	Complicque	4,76
10,08	Au	4,59
9,75	De	4,51
9,35	Enseignant	4,5
9,01	Le	4,44
8,04	Modernes	4,39
7,83	Défavorables	4,39
7,41	Stress	4,37
7,11	Notre	4,35
6,68	Poste	4,3
6,62	Environnement	4,27

**Tableau 2 :** Environnement thématique (ordre hiérarchique) du mot *travail*

Il s'agit d'un calcul de spécificité particulier ressemblant au calcul des corrélats qui repère à chaque fois deux mots associés dans le même graphique.<sup>22</sup> Mais cette fois-ci la recherche ne se réduit pas à deux mots comparés, mais à l'assortiment de tous les mots qui peuvent être en lien avec le mot connu comme *pôle* (ici *travail*) vu qu'ils se trouvent dans le même milieu. Nous n'allons pas commenter le tableau d'une façon de parler du lien entre le mot *travail et son voisinage*, car cela sera développer plus loin (voir figure 3, p.11), mais plutôt de repérer des liens entre les voisins de ce mot-pôle pour introduire de divers mini-thèmes. Par exemple, la liaison entre *poste* et *environnement* est le lieu. Les deux mots indiquent un endroit de rencontre pour travailler. Le terme *environnement* peut aussi avoir un lien avec *conditions* mais cette fois-ci pour indiquer quelque chose d'abstrait qui est le milieu ou l'entourage humain où le travail est effectué, etc.

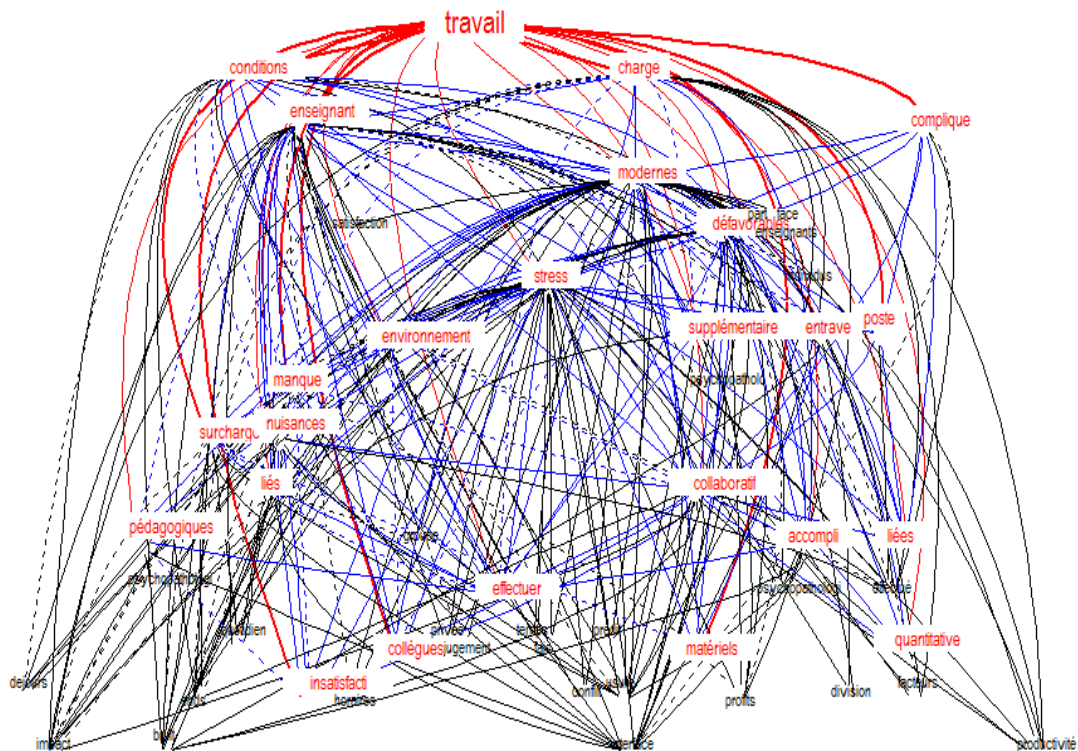
### 6.1.3. Co-occurrences du mot-pôle (*travail*) en absence des mots outils

Précisons que le calcul des co-occurrences se fait sur l'ensemble de tous les mots se rassemblant autour du mot-pivot. Sur le graphe plus bas, nous avons choisi le seuil 3 qui donne un chiffre de 53 termes dans l'environnement du mot-pôle, pour avoir une présentation

---

<sup>22</sup> Brunet E., 2011, *Hyperbase*, manuel de référence, p.48 selon l'adresse: <http://ancilla.unice.fr/bases/manuel.pdf> [consulté le 04/01/2016].

lisible avec le plus grand nombre de thèmes possibles. Après avoir éliminé les mots outils, nous avons matérialisé la liste des co-occurents comme dans le graphique ci-dessus.



**Figure 3 :** Environnement du mot *travail* en éliminant les mots outils

Selon ce graphe les termes qui se trouvent dans l'environnement du mot-pôle sont, à titre d'exemple : « conditions, charge, stress, défavorable, matériel, collègue,», etc.

Les termes qui sont liés directement au mot-pôle sont ses mots-amis, avec lesquelles il entretient une liaison solide. Ceux dont le lien est indirect et donc faible avec *travail* sont les proches de ses proches. La couleur rouge représente les liens directs et celle en bleu montre le lien des mots liés au pôle. Une troisième couleur noir, n'apparaît pas en cas de graphe simplifié, indique les autres cas. Les mots sont ainsi différenciés par plusieurs couleurs ; le rouge concerne les nœuds fréquentés, le noir pour les nœuds isolés (ayant moins de cinq liaisons). L'épaisseur des traits et la taille des caractères sont proportionnelles à leur importance dans le réseau. Cette structure *pôle-amis-proches d'amis* rend la lecture du graphe complexe sachant que nous pouvons avoir une image encore plus simplifiée si nous sollicitons ceci à *Hyperbase*.

Les résultats proposés par *Hyperbase*, dans le cas d'une recherche thématique sont semblables mais chaque fonction nous donne plus à voir et à commenter, et c'est le but, à notre avis, de la diversité des options du logiciel. Par exemple, le graphique via les occurrences attire

l'attention du chercheur et du lecteur pour l'analyser d'un point de vue quantitatif. L'environnement thématique d'un mot-pôle montre la question de la capacité des mots à se lier entre eux fortement ou légèrement. C'est une fonction cantonnée pour repérer le chemin (court/long) entre les mots selon la relation (forte/faible) qu'ils entreprennent. C'est, donc, au chercheur de choisir l'option la plus représentative pour son étude.

## **Conclusion**

Via la lexicométrie, la linguistique est devenue propriétaire des résultats chiffrés. Selon D. Mayaffre, D. Peschanski, a constaté une absence d'« effectivité »<sup>23</sup> dans les études linguistiques en Histoire, par exemple, et a réclamé des « résultats »<sup>24</sup>.

Aujourd'hui, la lexicométrie nous semble indispensable, non seulement vu la longueur du corpus, mais également pour une lecture séparée des résultats. Notre objectif à travers cet exposé est de montrer ce « que cachent les données textuelles »<sup>25</sup>, si l'on reprend l'expression de F. Rastier. Cette méthodologie de recherche permet au chercheur de donner libre cours à sa propre interprétation des résultats et sa créativité sous toutes les réserves d'usage.

La méthode statistique a remporté de nombreux succès non seulement auprès des domaines des sciences de la nature mais aussi dans les sciences humaines et sociales. Il est donc temps de réunir les différents domaines de recherche et permettre aux chercheurs de porter un regard plus large et plus perçant en leur facilitant l'accès aux connaissances d'autres spécialités.

---

<sup>23</sup> Mayaffre D., 2000, p.749.

<sup>24</sup> Ibid.

<sup>25</sup> Article de Rastier F., 9es JADT en 2008.

## Bibliographie

- Brunet E., 2011, *Ce qui compte*, Paris, Edition Champion.
- Lebart L., Salem A., 1994, *Statistique textuelle*, Paris, Dunod.
- Mayaffre D., 2015, *Textes augmentés et discours numériques*, Actes du colloque international du 14-19 septembre 2015 : Texte et discours en confrontation dans l'espace européen pour un renouvellement épistémologique et heuristique, Université de Lorraine, Metz.
- Mayaffre D., 2000, *Le poids des mots*, Paris, Honoré Champion,
- Muller C., 1992, *Principe et méthodes de statistique lexicale*, Paris, Honoré Champion,
- Muller C., 1992, *Principe et méthodes de la statistique linguistique*, Paris, Honoré Champion.
- Tournier M., 2000, *Lexicométrie*, in P. Charaudeau et D. Maingueneau, Dictionnaire d'Analyse des discours, Paris, Seuil.