Institut national
des langues
et civilisations orientales

INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS ORIENTALES

URL « Équipe de Recherche Textes, Informatique, Multilinguisme-Traitement Automatique des Langues »

Mémoire de Master 2

Analyse textuelle de corpus de discours écologiques relatifs au smog épais en Chine au moyen d'outils informatiques de text mining

DANG QINRAN

Sous la direction de

Monsieur **Mathieu VALETTE**

Avril 2015

TABLE DES MATIERES

Résumé

La dégradation de la situation environnementale climatique suscite l'attention générale de la société chinoise, du gouvernement au peuple chinois. Dès lors, le mot « le smog épais » est devenu un mot-clé qui apparaît fréquemment sur les sites web de divers types: institutionnels, informels, médiatiques. Ayant pour objectif de comparer et d'extraire différentes attitudes et réactions du gouvernement chinois et du peuple chinois, notre recherche vise à recueillir, étudier et analyser, au moyen d'outils informatiques, des discours et des textes publiés dans ces trois types de sites sur le sujet « smog épais en Chine ».

Mots-clés : Smog épais en Chine, pollution environnementale, protection de l'environnement, problèmes atmosphériques, analyses textuelles, analyses du discours, TAL, outils informatiques

1. Introduction

La Chine connaît depuis 30 ans un important essor économique, elle s'est développée de manière spectaculaire. Ce processus d'urbanisation rapide a aggravé la situation environnementale, surtout la situation atmosphérique. Il faut même remonter à 2008, où la question de pollution de l'air constituait une source très sérieuse de préoccupation des Chinois. Depuis 2011, cette pollution atmosphérique a atteint une ampleur inédite, et a commencé à marquer l'esprit du grand public, ainsi, il était dorénavant appelé smog épais (« 雾霾（*wumai*） » en chinois). Le résultat d'une importante pollution industrielle de l'électricité, produite à partir de charbon et de l'exploitation du trafic automobile. L'environnement n'est pas la seule victime du smog, la santé des habitants chinois dans les grandes agglomérations est aussi en cause. La pollution du smog provoque une inquiétude générale, du milieu gouvernemental chinois au peuple chinois. Dès lors, le mot « le smog épais » ou bien « le brouillard de pollution » est devenu un mot-clé très fréquent dans divers types de sites, dans la presse, sur les réseaux sociaux, les forums, et les blogs etc. Alors, parmi les articles consacrés à ce sujet, existe-t-il des divergences et des convergences au niveau des attitudes entre les textes institutionnels, ceux informels et ceux médiatiques ? C'est donc dans ce contexte-là et avec cette hypothèse que nous voulons effectuer une étude de recherche sur trois types de sites au sujet du « smog épais en Chine ». Dans le but de rendre plus logique et plus fluide ce travail, nous présentons d'abord de manière brève le cadre général de la structure de cette recherche. Globalement, le travail est composé de trois grandes parties principales :

- Construction du corpus ;
- Analyses du corpus ;
- Conclusion d'analyses.

Trois sites de types différents ont été choisis pour construire le corpus: le *gov* (le

site officiel du gouvernement central chinois), le *sohu* (un site web portail privé) et le *people* (site de presse institutionnel, l'organe du gouvernement chinois). En ce qui concerne l'analyse du corpus, nous avons étudié et comparé les trois différents types de discours en externe parmi les trois sous-corpus et en interne parmi les rubriques sur trois critères principaux : la fréquence, la cooccurrence ainsi que la catégorie lexicale de la répartition des termes autour des mots-clés. Ces études sont à la fois statistiques et linguistiques. Et finalement, la divergence au niveau des attitudes, et des réactions de différents groupes face au brouillard de pollution en Chine se révèleront avec la conclusion que nous en tirerons à la fin des analyses.

2. Construction du corpus

2.1. Sources du corpus

Après le contexte général de ce travail, il est temps de vous présenter plus précisément la première partie mentionnée ci-dessus : ***la construction du corpus***. Dans cette partie, nous expliquerons d'abord la nature des trois sites, leur position, leur fonction, leur audience, etc. ; ensuite, comme le corpus est assez conséquent, nous faisaon l'état des détails sur le corpus et sur les trois sous-corpus (le gov, le sohu et le people qui sont construits à partir des rubriques) : *la quantité, la sous-quantité, la taille totale, la taille moyenne du groupe de sous-corpus, la taille moyenne du texte,* etc. et entre temps, nous présentons l'outil informatique utilisé pour la récupération en grande quantité des textes, pour le prétraitement et pour les analyses.

Dans l'intention de mettre en œuvre la divergence sur l'opinion et les réactions du discours, nous avons choisi trois sites symboliques qui représentent les paroles différentes : *gov, people et sohu*. Ils sont respectivement institutionnel, médiatique, et informel.

2.1.1. Présentation du *gov*

Le premier site www.gov.cn est le site officiel du gouvernement chinois, qui joue un rôle de porte-parole des dirigeants chinois, selon l'introduction du site *gov*, ce site est administré par le Conseil d'État de la Chine, il fournit des informations sur le gouvernement chinois, tant au niveau central qu'au niveau provincial, telle que *les nouvelles, les politiques et les services offerts par les organismes de services publics*, etc. Ce site existe en deux langues : *le chinois et l'anglais* (Ici, on se focalise juste sur les textes rédigés en chinois). Il a au total 7 rubriques : *les nouvelles, les premiers, les politiques, le Conseil d'État, les services, les conseils, les statistiques, la situation du pays*. Parmi tous les sites institutionnels chinois, le site *gov* occupe la première place, il est accessible au peuple chinois et au monde entier, « *...provides a platform for the people around the world to interact with the Chinese government, in particular, contact with the Chinese Premier*[1] ». Voilà les raisons pour lesquelles nous avons ciblé ce site pour le membre institutionnel.

2.1.2. Présentation du *sohu*

Quant au deuxième site, l'idée est de sélectionner un site dont la nature est en opposition avec le site *gov.* Après avoir consulté les principaux sites web portail les plus populaires en Chine, le site *sohu* (www.sohu.com) a été pris. *Sohu* est le plus grand site web de Chine, selon l'introduction anglaise du *sohu*. Il fournit aux millions chinois un large éventail de choix en matière d'information, divertissement et de communication[2]. Il y a environ 50 rubriques diversifiées, qui offrent non seulement des nouvelles internes et externes à la Chine, mais aussi d'autres informations étroitement liées à la vie quotidienne du peuple chinois. Les textes du site sont tous et uniquement rédigés en chinois. Quant aux utilisateurs principaux, il est populaire parmi les

[1] Cf : http://english.gov.cn/Page/Uuid/e15c646a-446e-11e4-8156-03a6019c7a4e

[2] Version française traduite par moi-même de l'introduction du site *sohu* cf : http://investors.sohu.com/

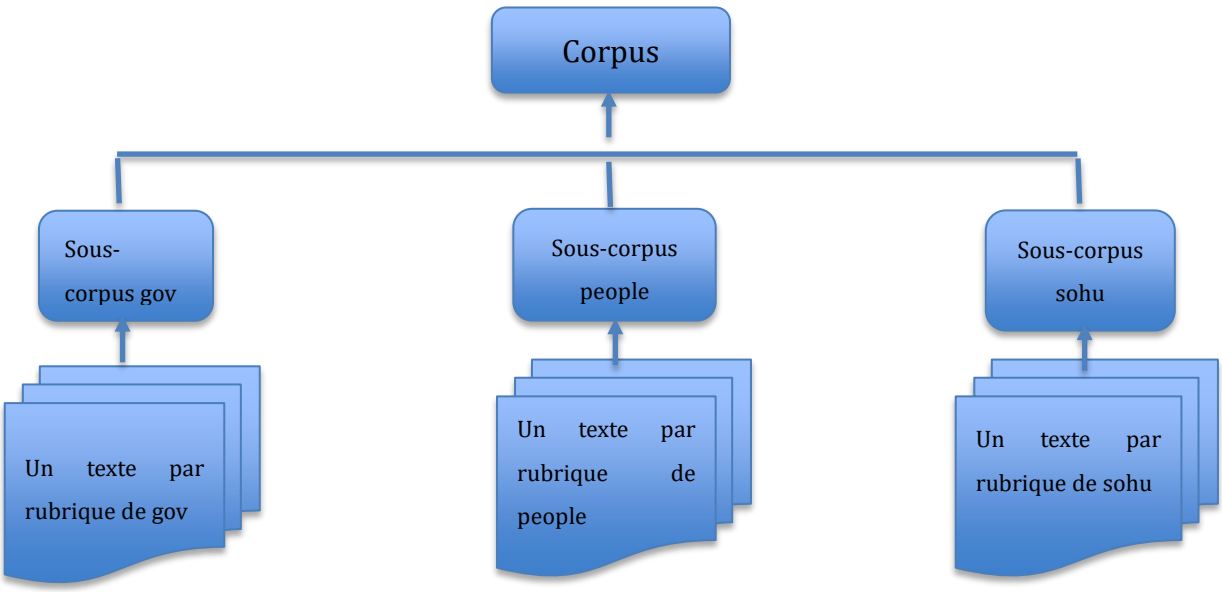
internautes chinois, et leurs âges varient d'une dizaine d'années à une cinquantaine d'années.

2.1.3. Présentation du *people*

Comme les deux sites précédents se différencie l'un avec l'autre par leur nature, il est préférable que le triosième site combine leurs caractéristiques pour jourer un rôle de transitition. Le site people (www.people.com.cn) remplit parfaitemnt cette cette condition et constitue ainsi le troisième site internet faisant l'objet de nos recherhers. Ce site doit son origine à *China Daily*, un journal officiel du parti communiste chinois, ce dernier est l'un des dix plus grands journaux du monde, le rôle principal du site *people* est de diffuser les nouvelles, non seulement sur le parti et le gouvernement chinois, mais aussi des informations qu'on peut trouver dans d'autres sites web portail privés. Il abrite environ 46 rubriques, dont celle de *地方 (locaux)，领导(dirigeants chinois)* marque sa nature institutionnelle comme le site *gov*, alors que les restes sont similaires comme le site *sohu*, tels que *le blog* et *le forum*, permettant aux utilisateurs de discuter assez librement dedans avec les dirigeants chinois. En plus de ces points communs, ce site propose 16 langues et s'implante à l'étranger à travers ses plusieurs filiales. Par rapport aux sites précédents, *people* facilite la diffusion des paroles du gouvernement chinois, afin de faire mieux connaître la Chine au monde, la caractéristique multilingue du *people* montre aussi le niveau d'ouverture du site.

2.2. Présentation des détails du corpus

2.2.1. Présentation de la structure du corpus



2.2.2. Taille du corpus et les trois sous-corpus

Comme nous le présente ce schéma, le corpus a été divisé en trois sous-corpus composés de textes venant des différentes rubriques des trois sites. Plus précisément, le site *gov* possède 8 rubriques, *people* 18, *sohu* 16. Le contenu de chaque rubrique est concrétisé par un texte d'environ 5500 mots, le corpus a au total 23100 mots répartis dans 42 rubriques sur 3 sites. En fait, cela ne représentent pas l'intégralité des rubriques des trois sites (sauf celles du site *gov*), et les rubriques qui seront indiquées sont celles, d'après nous, plus intéressantes et représentatives. Pour rendre plus clairs la répartition et les détails du corpus, les informations seront rangées dans un seul tableau. Les rubriques similaires et les rubriques spécifiques des sites seront soulignées en différentes couleurs. Cet arrangement consiste à nous faciliter la tâche de repérer les rubriques similaires, et ce, dans le but de les comparaer l'une avec l'autre. Voici le tableau détaillant les rubriques appartenant aux trois sites :

Nom du site	Nom de la rubrique	Nb de mots	Rubriques spécifiques	Rubriques similaires gov/people	Rubriques similaires gov/sohu	Rubriques similaires people/sohu	Rubriques similaires gov/people/sohu

GOV (Site institutionnel)	FUWU (service)	environ 5500	✓				
Rubriques totales : 8	GUOWUYUAN (conseil d'État)	environ 5500	✓				
Mots totaux : 44000	LDHD (les activités des dirigeants chinois)	environ 5500	✓				
	GZDT (Travail)	environ 5500	✓				
	LIANGHUI (Deux sessions[3])	environ 5500		✓			
	NEWS	environ 5500					✓
	ZHENGCE (politique)	environ 5500		✓			
	ZHUANTI (sujet spécial)	environ 5500	✓				
People (Site médiatique)	AUTOMOBILE	environ 5500				✓	
Rubriques totales : 18	BBS (forum)	environ 5500				✓	
Mots totaux : 115,500	BJ (Pékin)	environ 5500	✓				
	BLOG	environ 5500				✓	
	CCNEWS (nouvelles des entreprises gouvernementales)	environ 5500		✓			
	News	environ 5500		✓			
	EDU	environ				✓	

[3] Deux sessions : Les deux sessions annuelles de l'APN (Assemblée populaire nationale de Chine) et de la CCPPC (Conférence consultative politique du Peuple Chinois). Cf : http://www.weibo.com/p/1001603815970580892284?from=page_100206_profile&wvr=6&mod=wenzhangmod&sudaref=www.google.fr

	(éducation)	5500					
	ENERGY	environ 5500	✓				
	ENV (environnement)	environ 5500	✓				
	EZHENG (forum du people)	environ 5500	✓				
	FINANCE	environ 5500				✓	
	HEBEI	environ 5500	✓				
	HEALTH	environ 5500				✓	
	LEGAL	environ 5500	✓				
	LIANGHUI (Deux sessions)	environ 5500		✓			
	LINGDAO (dirigeants chinois)	environ 5500		✓			
	MILITARY	environ 5500				✓	
	PAPER	environ 5500	✓				
	POLITICS	environ 5500		✓			
	TRAVEL	environ 5500				✓	
	WORLD	environ 5500	✓				
SOHU (Site informel)	AUTOMOBILE	environ 5500				✓	
Rubriques totales : 16	BB (bébé)	environ 5500	✓				
Mots totaux : 88,000	BBS (forum)	environ 5500				✓	
	BLOG	environ 5500				✓	
	BUSINESS	environ				✓	

		5500					
	CITY	environ 5500	✓				
	EDU (éducation)	environ 5500				✓	
	FASHION	environ 5500	✓				
	GONGYI (bien public)	environ 5500	✓				
	HEALTH	environ 5500				✓	
	IMMOBILIER	environ 5500	✓				
	MILITARY	environ 5500				✓	
	NEWS	environ 5500					✓
	SZ (Suzhou)	environ 5500	✓				
	TECHNOLOGY	environ 5500				✓	
	TRAVEL	environ 5500				✓	
Rubriques totales des trois sites : 42							
Mots totaux : 231,000							

Tableau de toutes les rubriques des trois sous-corpus et les rubriques similaires

2.2.3. Présentation des rubriques similaires du corpus

Ce tableau nous a permis de mieux renndre compte de la répartition des rubriques dans les trois sites : les rubriques spécifiques de chaque site ; les rubriques similaires entre deux sites et trois sites ; le nombre des mots par rubrique ; le nombre total des rubriques ; et des mots sur le fond du tableau. Voici un résumé du tableau sur la répartition du corpus:

1. Les rubriques spécifiques (en jaune):

- *gov* :

- FUWU (service)
- GUOWUYUAN (Conseil d'État)
- LDHD (activités des dirigeants chinois)
- ZHUANTI (sujets spéciaux) ;

■ *people* :

- BJ (Pékin)
- HEBEI (une province près de Pékin)
- CCNEWS (nouvelles des entreprises gouvernementales)
- ENERGY (énergy)
- ENV (environnement)
- LEGAL (légal)
- PAPER (jounaux)
- WORLD (monde) ;

■ *sohu* :

- BB (bébé)
- CITY (ville)
- FASHION (mode)
- GONGYI (bien public)
- IMMOBILIER (immobilier)
- SZ (Suzhou) ;

2. Les rubriques similaires *gov/people* (en rouge) :

- LIANGHUI (Deux sessions)
- ZHENGCE (politiques)
- News
- LINDAO/LDHD (activités des dirigeants chinois)

3. Les rubriques similaires *gov/sohu* : NEWS

4. Les rubriques similaires *people/sohu* (en bleu):

- AUTOMOBILE (automobile)

- BBS (forum)
- BLOG
- EDU (éducation)
- FINANCE
- HEALTH (santé)
- MILITARY (militaire)
- Travel (Voyage)

5. Les rubriques similaires *gov/people/sohu* (en vert):
   - NEWS (nouvelles)

2.3. Présentation de l'outil informatique : Gromoteur

Gromoteur est un outil donnant aux linguistiques accès facile aux corpus textuels. Il permet de recueillir les textes des pages web ou d'importer les fichiers textuels locaux, de les prétraiter, analyser ainsi qu'exporter.

2.3.1. Première fonctionnalité du Gromoteur: récupération des textes depuis site web

La fonctionnalité de récupération des textes a été utilisée pour construire le corpus. Comme nous avons déjà les sites cibles et le sujet de recherche, il suffit de donner au *Gromoteur* l'adresse du site ainsi que les mots-clés de recherche. Prenons le site *gov* comme un exemple concret. Pour collecter les textes du site www.gov.cn qui parlent du sujet de 雾霾 (*smog épais*), il faut d'abord entrer une enquête "*雾霾 site: www.gov.cn*" au robot d'indexation (le *spider*) du *Gromoteur*, cette enquête permet à l'outil de connaître où il doit aller pour chercher les textes, et les mots-clés 雾霾 (*smog épais*) sur lesquels qu'il faut se baser. Après, il s'agit de contraindre la collecte en donnant des critères, tels que *combien de pages prendre, les pages qu'il faut éviter de prendre, et quelles rubriques prendre,* etc.

2.3.2. Deuxième fonctionnalité du Gromoteur: Prétraitement du corpus brut

Tous les textes récupérés sont des textes bruits en format site web, pour obtenir un corpus propre répondant aux besoins du travail, il est obligatoire de « nettoyer » ces textes en enlevant les informations inutiles, telle que les médadatas. La fonctionnalité prétraitement du *Gromoteur*- qui s'appelle "*select*" – a été très utile. Elle peut analyser la structure du site web et permet de laisser juste le contenu dont nous avons besoin : les textes eux-mêmes. Par exemple : Après avoir analysé le format et les codes sources du site *gov*, nous avons remarqué que le contenu est entouré de balise " <*p*>", et la plupart des textes dans ce site sont conçus comme celui-ci. Il est probablement pertinent de choisir le " <*p*>" comme critère d'exclusion pour le site *gov*. Partant de cette hypothèse, nous avons testé le premier sous-corpus et cela a réussi. Ainsi de suite, la réussite a continué sur les deux autres sites. À part de l'utilisation du *Gromoteur*, le terminal a été mis en usage avec des expressions régulières qui permettent d'affiner encore plus le corpus, telles que le "*cat*" qui permet de concaténer les fichiers en un seul, le "*sort*" pour ranger les fichiers, et le"*sed*" pour supprimer la tabulation avant la phrase, etc. La méthode de combinaison du *Gromoteur* et des expressions régulières accélère efficacement la procédure du travail.

```
283 <p>    最近大部分地区遭遇雾霾天气，给人们带来诸多不便。俗话说"秋冬雾，伤人刀"，大雾天气不仅能见
    度低，严重影响着路面交通，更为各种疾病打开了方便之门，对人们的健康造成很大的影响。阴霾天气不但影响心脏、呼吸道，还可能会伤
    肺。专家提醒，雾天灰尘颗粒较大，细菌、病毒也飘浮在空气中，人们出行最好佩戴口罩，同时居家也必须做好防范工作，谨防疾病来
    袭。</p>
284
285 <p>    <strong>雾天易发呼吸道疾病</strong></p>
286
287 <p>    雾霾天气极易使支气管哮喘、支气管炎、过敏性鼻炎和变异性咳嗽等呼吸道疾病复发。</p>
288
289 <p>    雾霾天气时，空气中漂浮着粉尘、烟尘，尘螨也可能悬浮在雾气中，支气管哮喘患者吸入这些过敏
    原，就会刺激呼吸道，出现咳嗽、闷气、呼吸不畅等哮喘症状。人们每次呼吸，往肺部深处吸入大约50万个微粒，可是在受到污染的空气
    中，吸入的微粒比这多100倍！</p>
290
291 <p>    许多人有晨练的习惯。专家强调，大雾天气无论如何也要停止晨练。人们晨练时，人体需要的氧气量
    增加，而雾中的有害物质会侵害呼吸道造成供氧不足，从而产生呼吸困难、胸闷、心悸等不良症状。</p>
```

Code source du site *gov*

2.3.3. Troisième fonctionnalité du Gromoteur : Regroupement des textes en fonction de rubrique

Nous avons mentionné les rubriques dans chaque site dans la partie introduisant la structure du corpus. Pour les regrouper en fonction de leur nom, nous nous servons de

la fonctionnalité de regroupement du Gromoteur, qui permet de catégoriser les trois sous-corpus en fonction des rubriques qui y sont insérées. Pour réaliser le regroupement, il suffit de tirer les caractéristiques et les points communs partagés des adresses du site, puis créer et entrer dans le *Gromoteur* une expression régulière dépendante de la formulation des adresses correspondantes, par exemple :

- l'adresse de la rubrique de FUWU (service) du site *gov* :

http://www.gov.cn/fwxx/jk/2011-12/14/content_2020255.htm

- l'adresse de la rubrique de GUOWUYUAN (Conseil d'État) du site *gov* :

http://www.gov.cn/guowuyuan/2015-04/02/content_2842074.htm

- l'adresse de la rubrique de NEWS (nouvelles) du site *gov* :

http://www.gov.cn/xinwen/2015-04/04/content_2842853.htm

Nous pouvons ainsi généraliser une expression comme " *www.gov.cn/(.*?)* ", étant donné que l'adresse commence toujours avec "*www.gov.cn*", et c'est à partir du premier slash que la formulation commence à manifester le nom de la rubrique. Cette fonctionnalité est efficace dans la mesure où elle évite toute autre procédure qui produit plusieurs d'aller-retour, d'exportation et d'importation, juste pour regrouper le corpus en rubrique.

2.4. Présentation de l'outil informatique de lemmatisation et étiquetage

Notre corpus a été partiellement construit après la procédure de prétraitement et de regroupement, il s'agit pour l'étape suivante de segmenter et d'annoter le corpus selon les critères lexicales de la langue chinoise, puisque la lemmatisation et l'étiquetage permettent d'effectuer les analyses linguistiques sur le corpus. En plus, les analyses d'algorithme nécessitent aussi ces informations de calcul des données statistiques, et le résultat final des analyses résultent de ces deux grandes parties. Développé par des chercheurs chinois, l'outil proposé par *Institute of Applied Linguistice Ministry of Education in China* a été choisi pour assumer la tâche. Tout critère établi dans l'outil est relativement plus exhaustif et plus exact. Cet outil exige un fichier de texte pur et

de l'encodage UTF-8. Il existe deux versions: version en ligne avec une limite de 10,000 mots à chaque fois et version téléchargeable sans limite et avec la possibilité d'importer plusieurs fichiers en gros. Nous en avons utilisé la deuxième.

Voici le tableau des codes d'étiquette proposés par l'outil chinois en version traduite. Il y a au total 49 codes qui catégorisent de manière plus ou moins complète: *les morphèmes, les noms, les verbes, les adjectifs, les adverbes, le temps, la location, l'institution, les ponctuations, et les sous-catégories* de ces aspects, etc.

Numéro	Code d'étiquette		Catégorie	Description de code
	Code d'étiquette niveau 1	Code d'étiquette niveau 2		
1	**a**		adjectif	adjectif
2		**aq**	adjectif de qualité	adjectif - qualité
3		**as**	adjectif d'état	adjectif - state
4	**c**		conjonction	conjonction
5	**d**		adverbe	adverbe
6	**e**		excitation	excitation
7	**f**		différence	difference
8	**g**		morphème	la première lettre du caractère chinois "gen" qui signifie le morphère en linguistique
9		**ga**	morphème d'adjectif	la première lettre du caractère chinois "gen" qui signifie le morphère en linguistique-adjectif
10		**gn**	morphème de nom	la première lettre du caractère

				chinois “gen” qui signifie le morphère en linguistique-noun
11		**gv**	morphème de verbe	la première lettre du caractère chinois “gen” qui signifie le morphère en linguistique-verb
12	**h**		tête	head
13	**i**		idiom	idiom
14		**la**	idiom d’adjectif	idiom-adjectif
15		**lc**	idiom de conjonction	idiom-conjonction
16		**ln**	idiom de nom	idiom-noun
17		**lv**	idiom de verbe	idiom-verb
18	**j**		abréviation	la première lettre du caractère chinois”jian” qui signifie l’abréviation
19		**ja**	abréviation d’adjectif	la première lettre du caractère chinois”jian” qui signifie l’abréviation-adjectif
20		**jn**	abréviation de nom	la première lettre du caractère chinois”jian” qui signifie l’abréviation-noun
21		**jv**	abréviation de verbe	la première lettre du caractère chinois”jian” qui signifie l’abréviation-verb
22	**k**		queue	comme d’habitude
23	**m**		numéro	numeral
24	**n**		nom	noun
25		**nd**	nom de direction	noun-direction

26		**ng**	nom général	noun-general
27		**nh**	nom des humains	noun-humain
28		**ni**	nom d'institution	noun-institution
29		**nl**	nom de location	noun-location
30		**nn**	nom de nation	noun-nation
31		**ns**	nom d'espace	noun-space
32		**nt**	nom de temps	noun-time
33		**nz**	nom spécifique	noun-la première lettre de "zhuan" qui signifie spécifique
34	**o**		onomatopéique	onomatopoeia
35	**p**		préposition	preposition
36	**q**		quantité	quantity
37	**r**		pronom	pronoun
38	**u**		auxiliaire	auxiliary
39	**v**		verbe	verb
40		**vd**	verbe de direction	verb-direction
41		**vi**	verbe intransitif	verb-intransitive
42		**vt**	verbe transitif	verb-transitive
43		**vl**	verbe de lien	verb-linking
44		**vu**	verbe d'auxiliaire	verb-auxiliary
45	**w**		autres	comme d'habitude
46		**wp**	ponctuations	comme d'habitude
47		**ws**	string autre que le chinois	"W"-string
48		**wu**	icône inconnu	"W"-unknown
49	**x**		autre que le morphère	comme d'habitude

Tableau d'explication des codes de la catégorie lexicale

3. État de l'art

3.1. Définition de l'analyse de contenu

Il y a au total 10 définitions qui sont attribuées à l'analyse de contenu par de nombreux spécialistes. Entre 1940 et 1968 Holsti (1968) a proposé de différentes définitions sur l'analyse de contenu. D'après Kaplan (1943), l'analyse de contenu est la « sémantique statistique des discours politiques ». Six ans après en 1949, Janis considère l'analyse de contenu comme une technique de classification matériel verbal ou écrit sur les « jugements » apportés par les analystes avec des « règles explicitement formulées ». Et puis trois ans après, Berelson (1952) et Cartwright (1953) définissent l'analyse de contenu comme « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste des communications, ayant pour but de les interpréter » (Berelson) ou « de toute conduite symbolique » (Cartwright). En 1968, Holsti définit cette démarche comme « méthode de recherche à buts multiples développée spécifiquement pour explorer un large éventail de problèmes dans lesquels le contenu de la communication sert de base à l'inférence […] par l'identification systématique et objective des caractéristiques spécifiques des messages » (pp. 597 et 601). Jusqu'ici, nous avons les cinq définitions de l'analyse de contenu avant 1970. À partir de 1974, deux spécialistes ont proposé deux définitions à l'analyse de contenu. D'abord, Unrug propose une définition qui lui en fait un «ensemble de techniques d'exploitation de documents, utilisées en sciences humaines […] consistant à mettre en fiche les principaux concepts utilisés ou les principaux thèmes abordés dans un texte scientifique […] possédant une signification ou un ‘‘contenu’’ sémantique […] » (p.9). Dans la même année pour Mucchilli, il pense que « elle se veut une méthode capable d'effectuer l'exploitation totale et objective des données informationnelles » (p. 17), et de son point de vue, l'analyse du contenu constitue essentiellement la recherche du « signifié » (p.23), autrement dit le sens d'un texte. Il considère l'analyse de contenu comme « rechercher les informations qui s'y trouvent, dégager le sens ou les sens de

ce qui y est présenté, formuler, classer tout ce que contient ce document ou cette communication » (p.26). Giorgi (1975a) désigne l'analyse de contenu comme « la recherche de la signification de la situation telle qu'elle existe pour le sujet » en se servant d'« analyse descriptive » et « analyse qualitative ». Puis, avec les propositions de Ghiglione et de Matalon (1978), on considère le sens précis de ce qui est dit dans l'analyse de contenu. La définition la plus récente est « l'investigation psychosociologique [est] à la recherche de la signification des faits, des attitudes, de l'exprimé [...] » (p.152) et « l'étude psychosociologique n'est pas terminée avec l'analyse statistique, elle ne fait que commencer » (p.158) proposée par Clapier-Valladon (1980a).

En s'appuayant sur l'évolution des définitions de l'analyse de contenu, voici une synthèse des grandes caractéristiques[4] de cette technique :

- objective/objectivité :
- exhaustive et systématique
- quantitative[5]
- qualitative[6]
- centrée sur la recherche de la signification du matériel analysé
- générative ou inférentielle

Si on conclut et synthétise les idées des deux auteurs sur l'analyse de contenu (d'un document ou d'une communication), l'analyse de contenu est une technique de recherche ayant pour objectif principal d'interpréter et de mettre en évidence, par le résultat d'analyse quantitative ou qualitative, l'attitude, la position idéologique et tout ce qui concerne les activités cognitives de l'auteur ou du locuteur.

---

[4] René L'Écuyer, Ph.D. (1990) « Méthodologie de l'Analyse Développementale de Contenu », Québec, Presse de l'Université du Québec,pp.9-12.

[5] Pour la forme quantitative, l'essentiel est de calculer la fréquence des éléments linguistiques choisis.

[6] Quant à la forme qualitative, les éléments linguistiques identifiés comme particuliers ou spécifiques sont à étudier attentivement. Malgré la différence de ces deux formes d'analyse, elles sont souvent utilisées simultanément lors de l'analyse de contenu. Ce qui est le cas réel de notre recherche.

De même pour notre but d'analyse, toute la recherche et l'analyse quantitative et qualitative sur trois types de textes de presse, toute procédure pour les catégoriser et comparer de manière objective, a pour objectif principal de dégager l'attitude et la position prise par les trois différents sites. Même si la position idéologique est fixée par leur nature (institutionnelle, informelle, et médiatique), les résultats de l'analyse de contenu de ces textes confirment notre hypothèse ou bien ils retrouvent d'autres sens qui ne sont pas représentés par leur nature mais quand même importants.

3.2. Types d'analyse de contenu :

Selon Mucchieli, il existe trois modes d'analyse de contenu :

Mode 1 : L'analyse logico-esthétique

Prenant en compte la forme du discours, tel que le vocabulaire, la longueur des phrases, l'ordre des mots, les figures de style, les hésitations, etc. L'analyse logico-esthétique vise à étudier l'influence de la structure du discours sur la perception du sens et l'état d'esprit du locuteur.

Ce mode d'analyse de contenu a été appliqué sur les textes des rubriques de « blogs » partagés du *people* et du *sohu*, où la forme du discours utilisée par les bloggeurs pour exprimer leurs sentiments personnels sur le « smog épais » diffèrent des autres textes. En se servant des vocabulaires émotionnels et des figures de style variées, le sens et l'état d'esprit des bloggeurs peut se refléter à travers la forme de la communication.

Mode 2 : L'analyse sémantique structurale

Différent que l'analyse logico-esthétique, ce mode d'analyse ne travaille pas sur la forme du discours, ni les éléments linguistiques mêmes, mais sur les principes et toutes

les relations qui organisent les éléments du discours sans prendre en compte leur contenu. Ces principes sont sous-jacents, ils peuvent être les systèmes de relations, les règles d'enchaînement, d'association et d'exclusion.

Mode 3 : L'analyse logico-sémantique

Ce type d'analyse consiste à dégager les informations en relation avec le sens principal du contenu : la préférence des thèmes, les prises de position, les mots-clés et les arguments servis pour justifier ceux précédents. Ainsi l'analyse logico-sémantique est subdivisée en trois types :

- **L'analyse thématique**. La notion de cette analyse est évidente à la vue du nom. Ce type d'analyse a pour objectif de déterminer les thèmes abordés du discours. Les unités sémantiques sont dégagées en fonction de leur signification et catégorie pour constituer l'univers du discours.

- **L'analyse du positionnement**. La position idéologique prise par l'auteur est reflétée par le jugement de l'auteur. Par le biais du calcul de la fréquence des éléments linguistiques plus souvent abordés qui représentent les jugements de l'auteur, on peut détecter la direction (positive, négative ou neutre) du jugement de l'auteur.

- **L'analyse fréquentielle**. À travers le calcul sur la fréquence des énoncés, on repère les énoncés les plus répandus dans le discours. Plus il est fréquent plus plus il est significatif aux yeux de l'auteur.

L'analyse logico-sémantique est le plus souvent utilisée lors de notre recherche. Les trois types subdivisés sont tous abordés dans notre analyse de manière soit combinatoire soit parallèle. Pour que les opinions, la prise de position et l'idée

principale des trois types de discours soient explicitement exprimés, nous devons déterminer les thèmes dans un univers de discours, détecter la position par les jugements prononcés dans le discours, et aussi de repérer les énoncés les plus fréquents. Tout dépend de l'analyse logico-sémantique.

3.3. Étapes de l'analyse de contenu

Synthèse comparative des étapes de l'analyse de contenu à partir de la nomenclature de six auteurs[7]

MUCCHIELLI (1974,1979)	D'UNRUG (1974)
1- Découpage en « unités informationnelles »	1- Définition des unités : - à l'avance ou - issues, induites du texte
2- Regroupement et classement des unités en « thèmes » ou « catégories » plus larges par « analogie de sens »	2- Établissement de catégories = classement - critères formels = stylistique ou - critères sémantiques = thématique
3- Quantification et traitement statistique	3- Comparaisons : intergroupe, intragroupe, à une norme, diverses parties entre elles
4- Analyse qualitative ou descriptive	4- Interprétation : cause et signification des résultats 5- Validation

[7] René L'Écuyer, Ph.D. (1990) « Méthodologie de l'Analyse Développementale de Contenu », Québec, Presse de l'Université du Québec, pp.54-56.

BARDIN (1977)	CLAPIER-VALLADON (1980a, b)*[8]
1- Préanalyse a- lectures flottantes b- choix des documents c- formulation des hypothèses et des objectifs d- repérage des indices et élaboration des indicateurs e- préparation du matériel	1- Lecture exhaustive du corpus : à plusieurs reprises, pour se familiariser avec le matériel, identifier les « idées forces », « certains thèmes » ou « mot bases » conduisant au premier repérage des « unités de texte »
2- Exploitation du matériel : le CODAGE a- découpage : choix des unités d'enregistrement et de contexte b- énumération : choix des règles de comptage c- catégorisation : choix et classification - procédures sans catégorie préalable - procédures avec catégorie préalable	2a- Démarche classificatoire 1 : identification des premiers thèmes de base pour la classification = thésaurus du corpus 2 : identification des thèmes définitifs = grille d'analyse - choix des unités d'enregistrement - découpage du discours - répartition du contenu dans les catégories de la grille 2b – Analyse quantitative :

[8] L'identification de 11 phrases ou étapes appliquées plus particulièrement à l'étude des récits de vue (POIRIER, CLAPIER-VALLADON, RAYBAUT, 1983, pp. 150-220) complète ce qui est résumé ici dans ce tableau de CLAPIER-VALLADON : 1- préanalyse ; 2- clarification du corpus ; 3- compréhension du corpus : lexique-thésaurus ; 4- organisation du corpus : grille d'analyse ; 5- organisation catégorielle ; 6- sommation des récits ; 7- analyse quantitative ; 8- compte rendu final ; 9- différenciation typologique ; 10-11- contrôles et commentaires.

3- Traitement des résultats, inférences et interprétations a- analyse quantitative et qualitative b- inférence c- interprétation	compilation, tests statistiques, comparaisons Analyse qualitative 3- Interprétation : - analyse des relations entre les diverses composantes (catégories) du matériel obtenu ; permet de dégager des « thèmes centraux », des « patterns », des « typologies » - 2 niveaux - descriptif - interprétatif en référant au contenu latent (au sens psychanalytique) pour découvrir le sens, la signification
VAN KAAM (1959)	GIORGI (1975a, b)
1- Établissement d'une liste de tous les énoncés du matériel obtenu = listing 2- Elaboration des premiers regroupements les plus évidents = « rough preliminary grouping » (séparer ce qui se ressemble de ce qui ne se ressemble pas) 3- a) « Réduction » des premiers groupements en groupes de plus	1- Découverte du sens général par la lecture de tout le matériel 2- Établissement d' « unités de significations » ou des divers « constituants » de base dans une deuxième lecture 3- a) Élimination des unités redondantes

en plus homogènes « sans vider la formulation présentée par le sujet », c'est-à-dire sans déformer le sens initialement donné par le sujet b) « Élimination » d'éléments non classifiables 4- Tentative d'identification des constituants de base 5- « Identification finale des constituants par application » : chaque énoncé est revu en se demandant en quoi il correspond à cette catégorie et dans quelle mesure il irait mieux ou non dans un autre	b) Regroupement des unités ressemblantes c) Clarification du sens des divers constituants par comparaison entre eux 4- Recherche du sens profond 5- Description scientifique : quantitative vs qualitative

Notre recherche et études se réfèrent aux étapes principales présentées dans le tableau résumé par L'ÉCUYER (1985a et 1987). Des lectures préliminaires sur les textes du « smog épais en Chine » constituent la prémière étape pour composer les textes de coprus. Grâce à cette phase préparatoire, nous avons pu avoir une vue d'ensemble sur les types de textes et formuler généralement notre problématique de recherche, à l'issue desquels trois types de sites ont été ciblés pour collecter le matériel et construire notre corpus. Après avoir défini les trois grands types d'unités informationnelles (institutionnelle, informelle et médiatique), il faut classifier à l'intérieur de chaque type les unités de texte. Les rubriques définies par les trois sites nous donnent la possibilité de découper les textes à l'intérieur de chaque « unité informationnelle» en fonction des thèmes. Ces mesures de réorganisation des textes permettent d'effectuer à la fois une analyse générale parmi les trois types de textes des trois sites, et une analyse profonde entre les rubriques homogènes de ces sites. Après avoir classifié le corpus, il faut le traiter en analysant les messages transmis de manière quantitative et qualitative. Ainsi, nous calculons la fréquence des éléments linguistiques en relation avec notre objectif et analysons les unités lexicales qui semblent spécifiques ou impliqués au mot-clé: le calcul de la fréquence des co-occurrences du mot-clé « 雾霾 smog épais » est pour repérer ceux qui sont le plus en relation avec notre sujet de recherche, et puis l'analyse au niveau lexicale et catégorielle de ces éléments consiste à mettre en évidence la position idéologique, et l'idée principale de ce type de textes. Souvent les résultats d'analyse se présentent sous forme de tableaux+graphique. La prise en compte des conclusions de notre recherche se manifeste par l'interprétation des résultats d'analyses, elle montre clairement les différentes opinions et positions prises dans chaque groupe : l'attitude et la réaction du gouvernement chinois et du peuple chinois.

3.4. La statistique descriptive

« La statistique descriptive peut être définie comme l'ensemble des méthodes de dénombrement, de classement, de synthèse et de présentation des données quantitatives

relatives à un ensemble d'individus » (Luc ALBARELLO, Étienne BOURGEOIS, Jean-Luc GUYOT, 2007, p.11). Après avoir dénombré, rangé et classé les données, le traitement statistique descriptif des unités lexicales en fonction de la catégorie définie sera appliqué pour mettre au propre et en évidence les résultats d'analyse. Deux modes classiques de présentation des données sont adoptées souvent par les statisticiens : un tableau de données ou de fréquence et un graphique, car par rapport à des descriptions énonciatives exhaustives des résultats sur de nombreuses données, il est préférable d'appliquer « un tableau des données+un graphique », des commentaires d'explications complémentaires sont nécessaires afin de mieux interpréter les résultats.

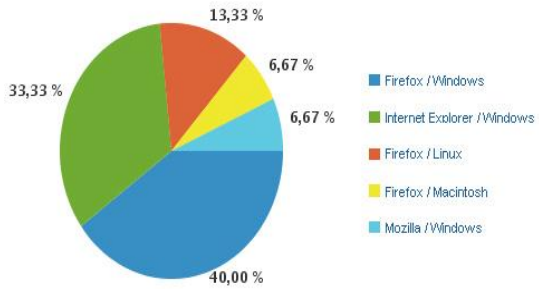
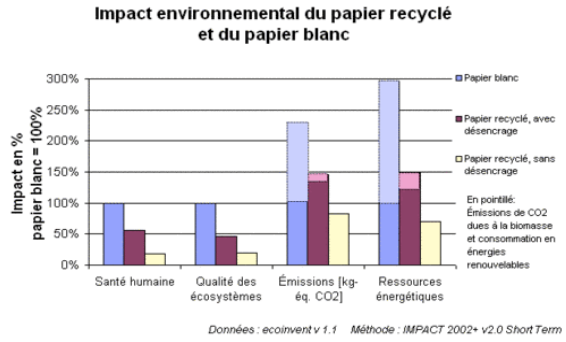
D'après les auteurs[9], un tableau de fréquence doit comprendre au minimum :
- un titre qui indique le plus clairement possible de quoi on étudie ;
- la liste complète des valeurs de la variable[10] ;
- les fréquences brutes ;
- le nombre total d'individus étudiés

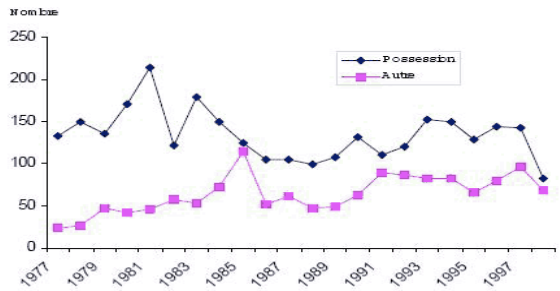
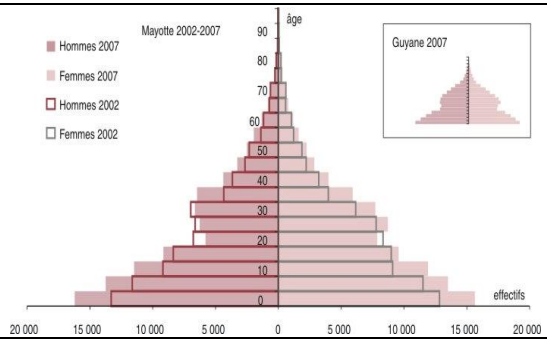
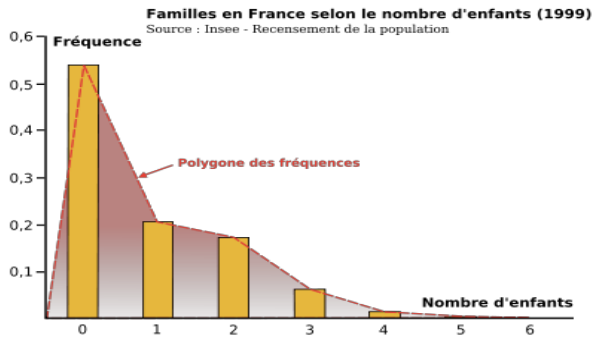
et peut être rajouter des éléments supplémentaires :
- les fréquences relatives ;
- les pourcentages ;
- les fréquences et pourcentages cumulés.

En complément du tableau des données, le graphique nous permet de « voir, au premier coup d'œil, les caractéristiques principales de la distribution » (2007, p.132) : soit la comparaison des écarts entre les différents variables, soit la tendance de l'évolution des variables (de bas en haut ou l'inverse).

[9] Luc ALBARELLO, Étienne BOURGEOIS et Jean-Luc GUYOT
[10] Il y a trois types de variables : nominale (données nominales), ordinale (relation d'ordre entre les catégories) et métrique (données numériques).

3.4.1. Les types des graphiques

Les types des graphiques	Modèle du graphique	Utilisé à variable de type
le diagramme circulaire, autrement dit « camembert » ou « tarte »		variable nominale et variable ordinale
le diagramme à barres ou diagramme en bâtonnets		variable nominale et variable ordinale

la ligne brisée		variable ordinale
l'histogramme		variable métrique
le polygone de fréquence		variable métrique

Nous appliquons la méthode statistique descriptive sur l'analyse du corpus. Basé sur des critères différents de regroupement, le corpus est classé en trois grands types de sous-corpus et plusieurs sous-types pour chaque sous-corpus. La classique modalité de statistique descriptive « un tableau de fréquence/des données + un graphique » a été davantage adoptée durant notre recherche statistique du corpus. Pour les éléments des tableaux : Chaque tableau possède un nom qui explique le contenu majeur des variables, sur lesquels on calcule la fréquence brute, et parfois le pourcentage ou la fréquence et le pourcentage cumulés, le nombre total des individus étudiés a été présenté chaque fois dans le tableau. Quant aux graphiques : Comme la plupart des données sont de type soit nominal soit ordinal, trois types de graphiques ont été largement utilisés : « camembert », « diagramme à barres » et « ligne à barre ». Cette modalité de statistique descriptive a été mis en œuvre dans les trois sous-corpus et dans certains des sous-types de chaque sous-corpus pour étudier la répartition du mot-clé «雾霾 smog épais » et des co-occurrents du mot-clé, etc..

4. Analyses du corpus

4.1. Présentation de l'outil informatique : Nexico

Les analyses textuelles et statistiques sont réalisées à l'aide de *Nexico*. *Nexico* est un outil de traitement inclut dans *Gromoteur*, il a une version simplifiée de *Lexico 3*, qui vise à analyser le corpus de manière statistique. Cet outil peut calculer, pour un terme choisi et pour tous les *tokens*, la fréquence totale d'un mot dans tout le corpus, la fréquence du mot dans une section[11] choisie, la cooccurrence du terme, et la spécificité basée sur l'implantation rapide de la distribution hypergéométrique cumulative. Il peut produire des tableaux et des schémas concernés les statistiques obtenues, qui sont exportables et destinés à aider les chercheurs à effectuer les analyses. Pour notre part de recherche, nous nous appuyons surtout sur les statistiques de la fréquence totale, la

[11] Ici, il s'agit de la rubrique d'un sous-corpus.

cooccurrence et la concordance (si nécessaire) que le *Nexico* offre.

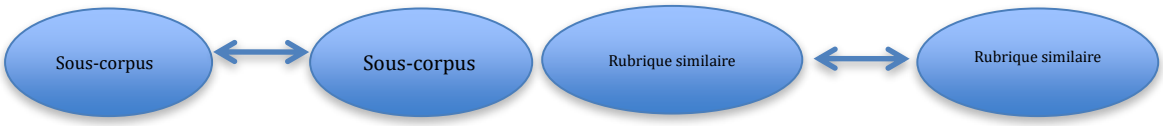
4.2. Choix des mots-clés

En chinois, le mot *雾霾 (smog épais)* est composé de deux caractères différents, comme ces deux caractères sont tous idéographiques, ils ont chacun un sens lors qu'ils sont séparés. L'outil de segmentation que j'ai utilisé a séparé ces deux caractères afin qu'il forme chacun une unité lexicale. Le 雾 (brume) et le 霾 (smog) sont donc les deux mots-clés que nous avons choisis pour mener la recherche suivante.

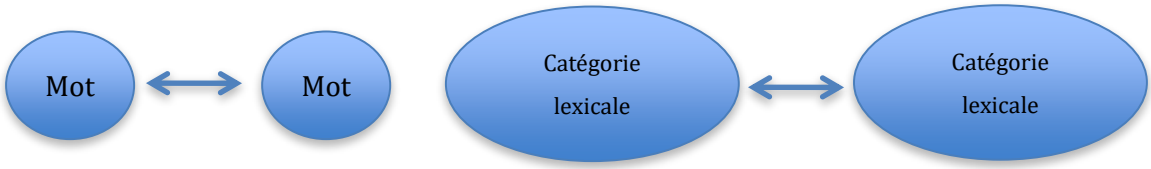
4.3. Structure générale des analyses

Puisqu'il s'agit de plusieurs aspects et de plusieurs éléments dans la phase d'analyse, nous préférons concrétiser d'abord la structure de l'analyse avec le schéma et un tableau qui donnerait une vision globale de cette étape cruciale du travail :

Deux groupes de comparaison :



Deux types de comparaison :



	Comparaison des mots dans un même sous-corpus	**Comparaison externe entre sous-corpus**	**Comparaison interne entre rubriques similaires**
Catégorie lexicale	*Mots les plus fréquents du sous-corpus*	*Mots les plus fréquents du sous-corpus*	*Mots les plus fréquents du sous-corpus*
	VS	VS	VS
	Mots les plus fréquents du sous-corpus	Mots les plus fréquents du sous-corpus	*Mots les plus fréquents du sous-corpus*
Sens	*Mots les plus co-occurrents avec deux mots-clés*	*Mots les plus co-occurrents avec deux mots-clés*	
	VS	VS	
	Mots les plus co-occurrents avec deux mots-clés	*Mots les plus co-occurrents avec deux mots-clés*	

Une série d'analyses sera déployée dans chaque sous-corpus, mais aussi dans les trois sous-corpus et enfin dans les rubriques similaires entre deux sous-corpus sur:

- *comparaison des mots les plus fréquents avec les mots les plus co-occurrents avec les deux mots-clés* ;
- *comparaison de la répartition des mots (les plus fréquents et les plus co-occurrents) en fonction de leur catégorie lexicale* ;
- *comparaison des mots les plus fréquents* entre sous-corpus et entre rubrique similaire;
- *comparaison des mots les plus co-occurrents avec deux mots-clés « smog » et « brouillard » entre sous-corpus*.

4.4. Analyse méthodologique

Pour chaque sous-corpus et chaque groupe de rubriques, une capture d'image du *Nexico* a été utilisée pour créer un tableau qui présente en détails les mots les plus fréquemment utilisés dans leur catégorie lexicale, un autre tableau qui contient les

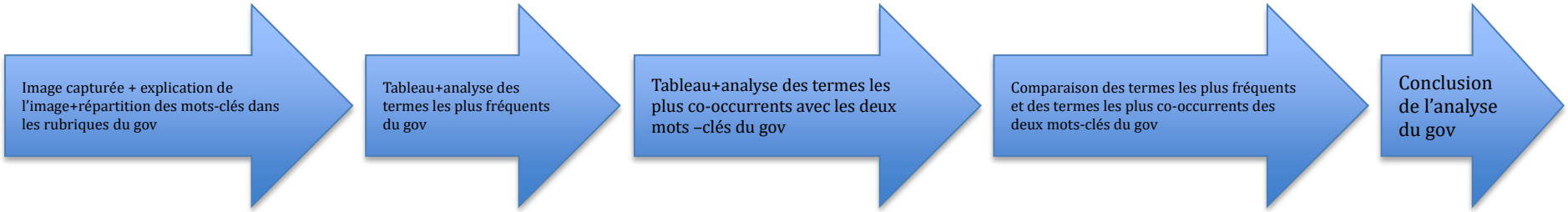
termes les plus co-occurrents avec les mots-clés (*smog* et *brouillard*) accompagnés aussi de leurs catégories lexicales, un pseudo-modèle de "*résultat+schéma+tableau*" sera appliqué sur le groupe des trois sous-corpus et l'autre groupe « parmi les rubriques similaires». En ce qui concerne l'ordre de l'explication de la capture d'image insérée dans le tableau, nous présenteron de haut en bas et de gauche à droite le contenu et les éléments la capture d'image, nous expliquerons enfin la signification de la courbe. Quant aux aspects du contenu du tableau, nous vous faisons part de la fréquence totale des 34[12] premiers mots dans tout le sous-corpus, la fréquence des 30 termes les plus co-occurrents des deux mots-clés ; nous développerons ensuite la comparaison de la répartition sur la catégorie lexicale des mots les plus fréquents et de ceux les plus co-occurrents avec les deux mots-clés à l'externe entre sous-corpus, à l'interne entre rubriques similaires.

4.5. Analyses du sous-corpus *gov*

4.5.1. Processus des analyses du sous-corpus *gov*

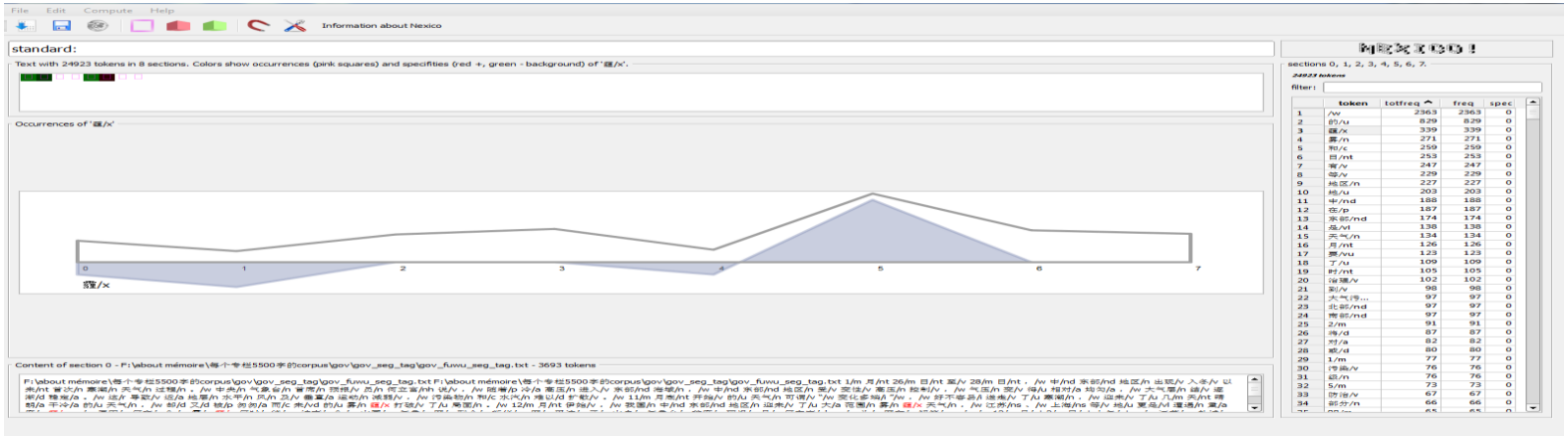
[12] Lors de l'analyse de la fréquence des mots, nous avons constaté que la fréquence des mots change en fonction de la rubrique d'appartenance, la décision a été prise pour fixer une valeur générale de choisir les 34 premiers mots parmi les plus fréquents et les 30 premiers mots qui sont co-occurrents avec les deux mots-clés.

Ci-dessous, une vue globale sur le processus et l'organisation de la première partie d'analyse sur le sous-corpus gov :



4.5.2. Explication et Analyses de l'image du sous-corpus *gov*

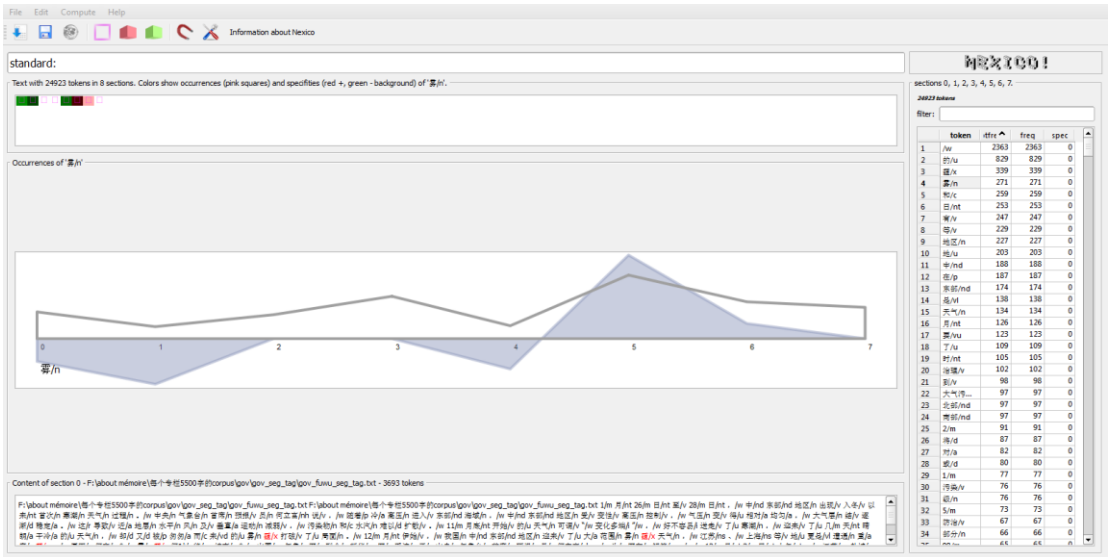
Débutons les analyses avec le premier sous-corpus sur le site *gov* :



Capture d'image sur la répartition du 霾/n (le smog) du gov

Voici le premier tableau d'analyses sur le mot-clé 霾(smog) du sous-corpus *gov*. Il y a deux endroits indiquant le nombre de *textes*, autrement dit combien de *rubriques*, et le nombre de *tokens* dans ce sous-corpus : en haut à gauche (*Text with 24923 tokens in 8 sections, Colors show occurrences (pink squares) and specifities (red+, green-background)*) of 霾/n (le smog) avec 8 carrés, et en haut à droite (*sections 0*[13]*,1, 2,3,4,5,6,7 24923 tokens*). Dans la configuration de base, il y a deux courbes, une grise et l'autre bleue. La grise montre le nombre total d'occurrences recherchées et la courbe bleu montre la spécificité du mot par rapport à la section. La courbe grise est donc toujours plus haute ou égale à zéro. La courbe bleue peut être négative. Quand à la courbe bleue, elle est au dessus de la ligne horizontale, par rapport à ce que nous attendions, le terme recherché est sur-employé dans la section, si la courbe bleue est au dessous de la ligne horizontale, le terme est sous-employé dans la section, si la courbe bleu est au même niveau de la ligne horizontale, cela veut dire que l'emploi du terme est plus ou moins normal.

En plus de la répartition du « smog », nous allons aussi observer la répartition du 雾/n « brouillard » dans le sous-corpus *gov* avec l'image capturée du *Nexico* et le tableau suivant:



Capture d'image sur la répartition du 雾／n (brouillard) du gov

13 Le *gromoteur* compte les textes à partir de 0 au lieu de 1.

En dessous de la capture d'écran, un cadre est présent pour montrer le contenu ainsi que le nombre des *tokens* de la section choisie. Alors que le tableau à droite possède au total quatre éléments principaux : le *token* (le terme segmenté), le *totfrq* (la fréquence totale du terme dans tous les textes du sous-corpus), le *freq* (la fréquence du terme dans la section choisie), le *spec* (la spécificité du terme dans tous les textes du sous-corpus). Avant de recueillir les termes les plus co-occurrents avec le mot-clé 霾/*x* (*smog*), nous révèlerons d'abord les 34 mots les plus fréquents dans le sous-corpus *gov* d'après leur nombre de totfreq calculé par le *Nexico* (le *Nexico* permet aussi de mettre en ordre croissant le *totfreq* et le *freq*).

La troisième colonne dans l'image du gromoteur nous fait part du résultat de calcule de la spécificité du mot en question, dont la formule de calcule est $SP=1/10^x$ ($x>0$), quand $x<0$, $SP=10^x$, le chiffre indiqué dans la colonne de « SPEC » dans la capture d'écran est la valeur de *x*. Cette méthode de calcul signifie que la probabilité du mot en question est $1/10^x$ dans la section du sous-corpus. Autrement dit, il y a une relation inversement proportionnelle entre la valeur absolue de x et de taux de fréquence du mot en question : quand $x>0$, plus que la valeur de *x* augmente, le taux de fréquence du mot en question baisse corrélativement, et *vice versa*. Le chiffre indiqué dans la colonne de « SPEC » dans l'image capturée est la valeur de *x*. C'est pour cette raison majeur que l'on s'intéresse à la situation où les deux mots-clés sont dans une position de sur-employés ou sous-employés, en d'autres termes, la valeur de *x* est supérieure ou inférieur à 0.

Comme la courbe grise du mot 霾/*x(smog)* montre le nombre total d'occurrences recherchées, la courbe bleu signifie l'évolution de l'état d'emploi des deux mots-clés dans toutes les sections du sous-corpus *gov*, présentons l'information transmise par les deux courbes dans le tableau comme ci-dessous :

Section	0 FUWU (service)	1 GWY (Conseil d'Etat)	2 GZDT (Travail)	3 LIANGHUI (Deux sessions)	4 LDHD (Activités des dirigeants chinois)	5 NEWS (Nouvelle)	6 ZHENGCE (Politique)	7 ZHUANTI (Sujets spécifiques)

L'état de l'emploi du mot « smog »	-4 (sous-employé)	-8 (sous-employé)	0 (normal)	0 (normal)	-4 (sous-employé)	11 (sur-employé)	0 (normal)	0 (normal)
L'état de l'emploi du mot « brouillard »	-3 (sous-employé)	-6 (sous-employé)	0 (normal)	0 (normal)	-4 (sous-employé)	20 (sur-employé)	2 (sur-employé)	0 (normal)

Tableau de la répartition du 霾／n (le smog) et du 雾／n (le brouillard) du gov

À travers ces deux images et le tableau, on peut constater que l'état de la répartition est bien similaire chez ces deux mots-clés, ce phénomène prouve d'une autre manière que la relation corrélative entre ces deux mots comme ce que nous avons mentionné en haut : en chinois, ces deux caractères font en effet un seul mot qui signifie le « smog épais de pollution ».

Si nous appuyons sur ce principe de départ pour analyser l'état de l'emploi des deux mots-clés dans les rubriques du *gov*, nous pouvons ainsi dégager les rubriques « *fuwu* »(service public), « *guowuyuan* »(Conseil d'État) et « *lingdaohuidong* » (Activités des dirigeants chinois) où la valeur absolue du chiffre de spécificité est inférieur à 10, autrement dit, la probabilité de présence de un des deux mots-clés est plus de $1/10^{10}$.

4.5.3. Présentation et analyse des mots les plus fréquents du *gov*

Après avoir comparé dans l'ensemble de l'état de la présence des mots-clés du *gov*, nous allons préciser l'analyses en rentrant dans les détails, débutons l'analyse des mots les plus fréquents du *gov*. Pour ce faire, nous avons utilisé le résultat dans le tableau à droite de la capture d'écran fourni par Nexico.

Explication des titres de chaque colonne du tableau :

⇒ *token : mot segmenté comme unité lexicale ;*

⇒ *totfreq : la fréquence total du mot (token) dans tous les textes du sous-corpus ;*

⇒ *cooc : combien de fois que ce terme est tombé à côté du mot-clé (dans le seuil de*

5) ;

⇒ *traduction : le sens du mot en français*

Numéro	Token	Totfreq	Traduction
1	/w	2363	espace
2	的/u	829	auxiliaire
3	霾/x	339	smog épais
4	雾/n	271	brouillard
5	和/c	259	et
6	日/nt	253	le jour
7	有/v	247	avoir
8	等/v	229	attendre
9	地区/n	227	la région
10	地/u	203	auxiliaire
11	中/nd	188	le centre
12	在/p	187	dans
13	东部/nd	174	l'est
14	是/vl	138	être
15	天气/n	134	le temps
16	月/nt	126	le mois
17	要/vu	123	vouloir
18	了/u	109	auxiliaire
19	时/nt	105	l'heure
20	治理/v	102	régulariser
21	到/v	98	arriver
22	北部/nd	97	le nord
23	南部/nd	97	le sud
24	大气污染/v	97	la pollution atmosphérique
25	2/m	91	numéro
26	将/d	87	marqueur de future
27	对/a	82	correct
28	或/d	80	ou
29	1/m	77	numéro
30	级/n	76	niveau
31	污染/v	76	polluer
32	5/m	73	numéro
33	防治/v	67	protéger
34	部分 / n	66	la partie

Tableau des termes les plus fréquents du gov

À travers ces 34 mots les plus fréquents, les textes du *gov* sont relativement structurels, car le calcul de la proportion des mots, au niveau de sa catégorie lexicale avec son étiquette, montre bien qu'il y a 16/34 de mots[14] dans le groupe nominal, 9/34 dans le groupe verbal, et les 9 mots restants sont composés soit des conjonctions, soit des auxiliaires, soit des adverbes. Parmi ces mots, nous constatons en outre qu'en plus des deux mots-clés le *霾／x(smog)* et le *雾／n(brouillard)*, les noms (termes nominaux) indiquent la région et la direction de la région comme le *地区／n (la région)*， *东部／nd (l'est)*，*北部／nd (le nord)*， *南部／nd (le sud)*, ils reflètent directement la localisation de la Chine où les gens subissent le smog épais, l'année dernière, le reportage de la presse annonçait beaucoup plus fréquents que la région du nord[15] constitue la région qui a le plus souffert de la pollution industrielle venant des usines, alors qu'au fur et à mesure, ce phénomène n'est plus expulsif qu'au nord, la pollution atmosphérique répandait et touchait aussi le sud et le l'est. Quant aux verbes, la moitié des verbes ont un lien très étroit avec le « smog épais », ils sont souvent utilisés et associés avec le « smog épais », tels que le *治理／v* (régulariser), le *污染／v* (polluer), et le *防治／v* (prévenir), ces verbes représentent les actions prises par le gouvernement chinois, qui font preuve qu'en face de la situation actuelle en Chine sur la pollution atmosphérique, le gouvernement ne reste pas les bras croisés, au contraire, il est bien au courant de la situation soucieuse, et s'est attelé à prendre des mesures contre la pollution.

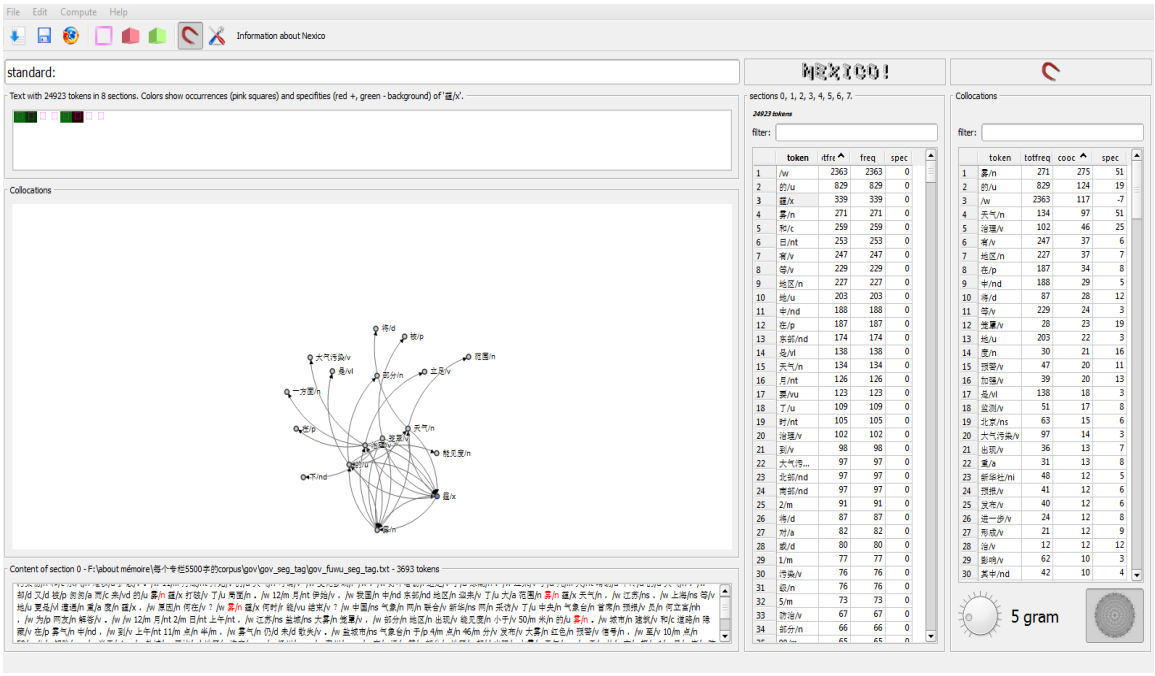
4.5.4. Présentation et analyses des mots les plus co-occurrents avec les deux mots-clés

4.5.4.1. Présentation et analyses des mots les plus co-occurrents avec 霾/n (le smog)

Dans le second temps, avec les résultats statistiques du *Nexico*, nous allons regarder ensemble les termes les plus co-occurrents avec le premier mot-clé *霾/x* (*smog*) :

[14] En fait, le mot *霾* peut être aussi considéré comme un nom comme le *雾*, la relation entre ces deux mots est concordante, et la combinaison de ces deux caractères signifie le *smog épais*.

[15] Ce n'est pas seulement le nord, mais aussi le nord-ouest de la Chine.



Capture d'écran du Nexico La collocation du 霾／x(smog) du gov

Nous pouvons remarquer dans le tableau ci-dessus sur la collocation du *霾／x(smog)* qu'il y a aussi le *totfreq* et le *spec*, le premier indique pareillement le nombre total d'un terme choisi, alors que le deuxième change. En plus de ces deux indices, le tableau présente le nombre de la cooccurrence (le *cooc*), c'est à dire qu'il y a combien de fois de cooccurrences lorsque ce terme est tombé à côté du mot-clé. (Ici, le *Nexico* a choisi par défaut le seuil de 5, signifiant qu'on choisit les 5 mots à gauche et à droite du mot-clé. Le seuil est changeable si on tourne le petit bouton 

en bas à droite du tableau).

Rangeons les données dans un tableau selon le nombre de la cooccurrence des termes en ordre croissant, les 30 premiers mots les plus co-occurrents du mot-clé *霾／x* (*smog*) sont :

Numéro	Token	Totfreq	Cooc	Traduction
1	雾 / n	271	275	le brouillard
2	的 / u	829	124	auxiliaire
3	／w	2363	117	espace
4	天气 / n	134	97	le temps

5	治理 / v	102	46	régulariser
6	有 / v	247	37	avoir
7	地区 / n	227	37	la région
8	在 / p	187	34	dans
9	中 / nd	188	29	centre
10	将 / d	87	28	marqueur de future
11	等 / v	229	24	attendre
12	笼罩 / v	28	23	couvrir
13	地 / u	203	22	auxiliaire
14	度 / n	30	21	degré
15	预警 / v	47	20	alerter
16	加强 / v	39	20	renforcer
17	是 / vl	138	18	être
18	检测 / v	51	17	détecter
19	北京 / ns	63	15	Pékin
20	大气污染 / n	97	14	la pollution atmosphérique
21	出现 / v	36	13	apparaitre
22	重 / a	31	13	lourd
23	新华社 / ni	48	12	Presse de XINHUA (équivalent d'AFP)
24	预报 / v	41	12	prédire
25	发布 / v	40	12	publier
26	进一步 / v	24	12	marcher un pas de plus
27	形成 / v	21	12	se former
28	治 / v	12	12	régler
29	影响 / v	62	10	influencer
30	其中 / nd	42	10	parmi

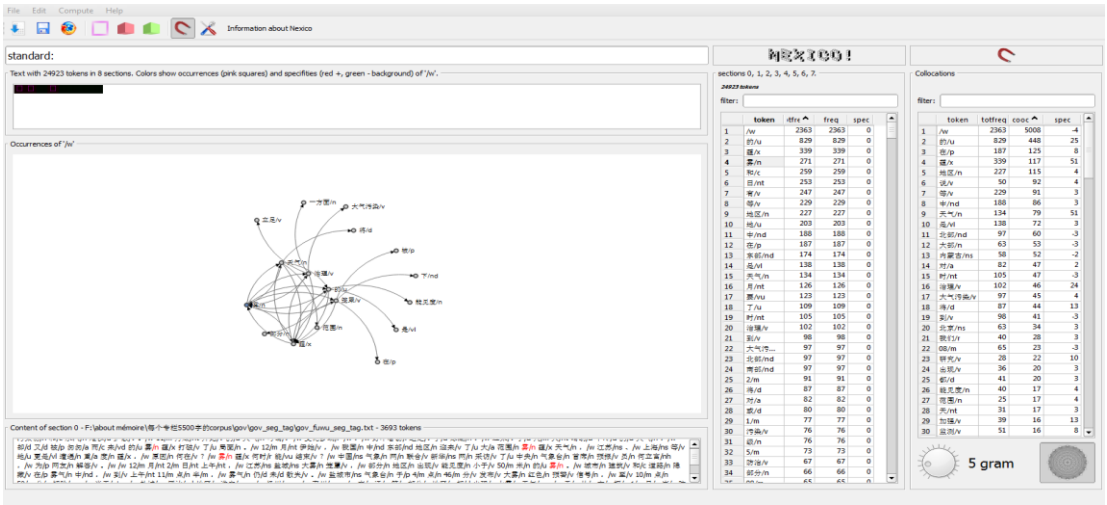
Tableau des termes les plus co-occurrents du 霾 / x (smog) du gov

La répartition statistique de la catégorie avec l'étiquette lexicale manifeste qu'il y a 8/30 de mots sont nominaux, 16/30 de mots sont verbaux, d'autres catégories comme l'adjectif, la préposition et l'auxiliaire n'occupent que 6/30. Le mot *雾 / n (brouillard)* est au premier rang en raison de leur relation concordante en chinois. Après, comparés avec le tableau précédent, il y a des mots qui ont été répétés : dans les co-occurrents 雾 / n (brouillard), *地区 / n*(la région), *天气 / n* (le temps)，*治理 / v*(régulariser et *大气污染 / n (la pollution atmosphérique).* En plus de ces mots, dans les co-occurrents de *霾 / x* (le smog), de nouveaux mots apparaissent : comme verbes le *笼罩 / v (couvrir)*， *预警 / v(alerter)*， *加强 / v(renforcer)*， *检测 / v (détecter)*， *预报*

/v(prévenir)，发布／v(publier)， 治／v(régler)，影响／v(influencer), comme nom d'espace le *北京／ns* (Pékin)，nom d'institution le *新华社／ni* (Presse XINHUA), et comme adjectif le *重／a* (lourd). Parmi les co-occurrents, l'adjectif *重／a* (lourd) décrit l'état de la pollution du smog épais en Chine. Les verbes parlent tous du « smog épais », soit de lui-même, soit des mesures que le gouvernement chinois a prises pour le détecter, le publier, ou le régulariser. Les co-occurrents du *霾／x* (smog) englobent plusieurs aspects des mesures que le gouvernement a pris mis en évidence par ces verbes, en même temps, le mot « *Pékin* » indique de manière précise l'endroit marqué par le smog épais en Chine. Le *新华社／ni* (Presse XINHUA) est la *Presse XINHUA,* elle est une institution équivalente de l'AFP dans quelque sorte, ce qui les distingue le plus c'est que la Presse XINHUA est sous la direction du gouvernement chinois, il publie non seulement des nouvelles comme font d'autre presse, et une certaine partie est aussi contribué aux nouvelles des dirigeants chinois ainsi qu'au gouvernement. Parmi les co-occurrents l'adjectif *重／a* (lourd) décrit l'état de la pollution du smog épais en Chine.

4.5.4.2. Présentation et analyses des mots les plus co-occurrents avec 雾/n (le brouillard)

Continuons-nous avec le deuxième mot-clé du sous-corpus *gov*: le *雾／n* (le brouillard), nous étalons le contenu de l'analyse conformément à la méthode du premier mot-clé le *霾／x* (smog). Donc, voici l'image capturée sur la collocation du *雾／n* (le brouillard) et le tableau des termes les co-occurrents du *雾／n* (le brouillard) analysé par le *Nexico* :



***Capture d'écran du Nexico La collocation du 雾／n* (brouillard)**

Numéro	Token	Totfreq	Cooc	Traduction
1	／w	2363	5008	espace
2	的／u	829	448	auxiliaire
3	在／p	187	125	dans
4	霾／x	339	117	le smog
5	地区／n	227	115	la région
6	说／v	50	92	parler
7	等／v	229	91	attendre
8	中／nd	188	86	le centre
9	天气／n	134	79	le temps
10	是／vl	138	72	être
11	北部／nd	97	60	le nord
12	大部／n	63	53	la plupart
13	内蒙古／ns	58	52	Inner Mongolie
14	对／a	82	47	correct
15	时／nt	105	47	l'heure
16	治理／v	102	46	régulariser
17	大气污染／n	97	45	la pollution atmosphérique
18	将／d	87	44	marqueur de future
19	到／v	98	41	arriver
20	北京／ns	63	34	Pékin
21	我们／r	40	28	nous
22	08／m	65	23	numéro
23	研究／v	28	22	étudier
24	出现／v	36	20	apparaitre

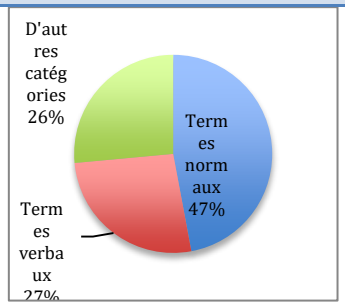
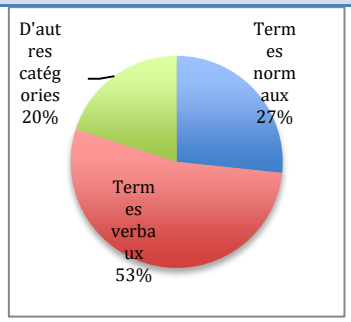
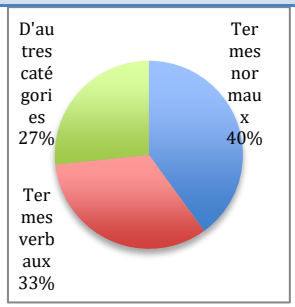
25	都 / d	41	20	tous
26	能见度 / n	40	17	la portée de vue
27	范围 / n	25	17	la zone couverte
28	天 / nt	31	17	la jounée
29	加强 / v	39	16	renforcer
30	监测 / v	51	16	détecter

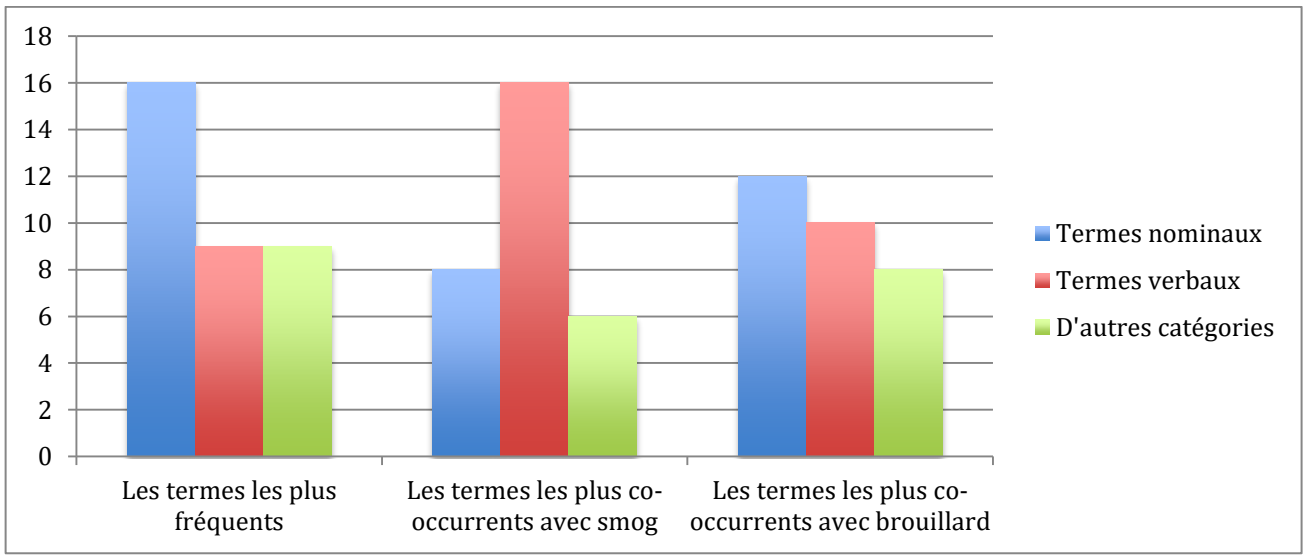
Tableau Les termes les plus co-occurrents du 雾 / n (brouillard) du gov

La proportion des termes nominaux est 12/30, celle des verbaux est 10/30, d'autres est 8/30. La répartition des catégories lexicales est relativement identique que les deux derniers (les plus fréquents et les plus co-occurrents du *smog*) du sous-corpus *gov*. Comparé avec l'analyse précédente, à part son collaborateur le *霾 / n* (le smog) qui se situe au premier rang de la co-occurrente, il y a aussi des termes co-occurrents du *雾 / n* (le brouillard) qui sont croisés avec ceux du *霾 / n* (smog), tel que *地区 / n (région)*， *天气 / n (temps)*，*治理 / v (régulariser)*，*大气污染 / n (pollution atmosphérique)*，*北京 / ns (Pékin)*，*加强 / v (renforcer)*，*监测 / v (détecter)*. Par rapport au "*Tableau Les termes les plus fréquents du sous-corpus gov*", c'est aussi les termes comme le 地区 / n (région)，le 天气 / n (temps)，le 治理 / v (régulariser et le 大气污染 / n (pollution atmosphérique) qui chevauchent. Ainsi, nous pouvons dire qu'il y a un lien très étroit entre le « smog épais » et ces quatre termes dans le sous-corpus *gov*.

4.5.5. Graphique synthétique des analyses du sous-corpus *gov*

Dans l'intention de présenter globalement la première phase d'analyse du sous-corpus *gov*, nous verrons les statistiques et les données dans un seul tableau avec des graphiques visées :

Catégorie lexicale	**La proportion des termes les plus fréquents du sous-corpus**	**La proportion des termes les plus co-occurrents avec « smog »**	**La proportion es termes les plus co-occurrents avec « brouillard »**
Termes nominaux (bleu)	16/34	8/30	12/30
Termes verbaux (rouge)	9/34	16/30	10/30
D'autres catégories (jaune)	9/34	6/30	8/30
Graphique			



Graphique synthétique des analyses du sous-corpus gov

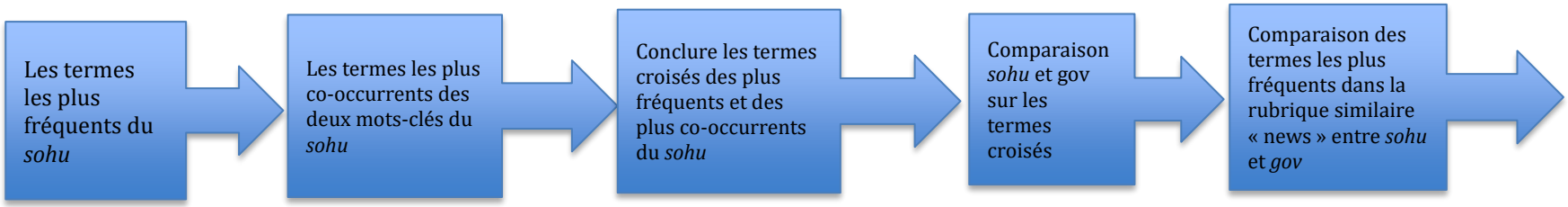
4.5.6. Bilan d'analyse du sous-corpus *gov*

Le bilan d'analyses du sous-corpus gov sera présenté sur trois égards : l'organisation des textes, la répartition des mots selon la catégorie lexicale et le contenu principal du sous-corpus. En ce qui concerne l'organisation des textes du site *gov*, ces derniers sont relativement structurés et organisés de manière régulière, que ce soit au niveau des mots croisés (les nominaux et les verbaux) répétitifs, au niveau de la proportion similaire de différentes catégories lexicales des termes ou au niveau du degré de la corrélation avec les mots-clés. Quant au contenu, reflété par les termes les plus fréquents et les termes les plus co-occurrents avec les mots-clés, les textes se centralisent au sujet du « smog épais en Chine», des régions souffrant en Chine, et surtout des mesures que le gouvernement chinois a pris face à cette situation urgente. Voilà des caractéristiques conclues du sous-corpus *gov*.

4.6. Analyses du sous-corpus *sohu*

Maintenant, nous utilisons la même méthodologie pour le deuxième sous-corpus informel: *sohu.* Par rapport au *gov* qui est un site institutionnel du gouvernement chinois, le site de *sohu* est un site informel privé. Ayant pour objectif de vous présenter les résultats et les analyses de l'ordre du global en détail, nous listons d'abord les termes les plus fréquents du *sohu*, ensuite, les co-occurrents des deux mots-clés, puis nous en tirer les termes croisés dans les deux groupes (les plus fréquents + les plus co-occurrents), da manière à mettre en contraste le résultat du *gov* et du *sohu* au niveau de l'utilisation des mots des textes en fonction de leur nature et de responsabilités. L'ordre de l'analyse et de l'explication sont comme ci-dessous :

4.6.1. Processus des analyses du sous-corpus *sohu*



4.6.2. Présentation et analyses des mots les plus fréquents du sou-corpus sohu

Contrairement à la première partie d'analyse du *gov*, nous ne commençons pas l'analyse avec les termes les plus fréquents du sous-corpus *sohu*, car ce dernier dispose de 2 fois plus de rubriques que celles du *gov*. Une fois dégagé le résultat des analyses sur les termes les plus fréquents et ceux les co-occurrents avec les deux mots-clés, nous allons mettre en contraste ces deux sous-corpus pour effectuer la comparaison. Après, des analyses seront menées sur l'état de l'emploi des deux mots-clés dans toutes les rubriques choisies dans *sohu*. Donc, voici le premier tableau sur les termes les plus fréquents du *sohu*.

Numéro	Token	Totfreq	Traduction
1	/w	8363	espace
2	的/u	2469	auxiliaire
3	霾/x	821	le smog
4	雾/n	662	le brouillard
5	在/p	497	dans
6	了/u	404	auxiliaire
7	是/vl	394	être
8	空气/n	357	l'air
9	和/c	274	avec
10	有/v	237	avoir
11	中/nd	208	dans…
12	污染/v	208	polluer
13	一/m	193	un
14	也/d	181	aussi
15	不/d	178	non
16	等/v	170	etc.
17	我/r	158	je
18	年/nt	157	l'année
19	就/d	153	adverbe
20	天气/n	153	le temps
21	北京/ns	153	Pékin
22	要/vu	147	falloir
23	多/a	143	nombreux
24	对/a	137	envers
25	这/r	135	ceci
26	人/n	130	personne
27	5/m	125	cinq
28	上/nd	122	dessus
29	可以/vu	119	pouvoir
30	2/m	117	deux
31	能/vu	115	pouvoir
32	都/d	115	tous
33	为/p	114	pour
34	会 / vu	111	pouvoir

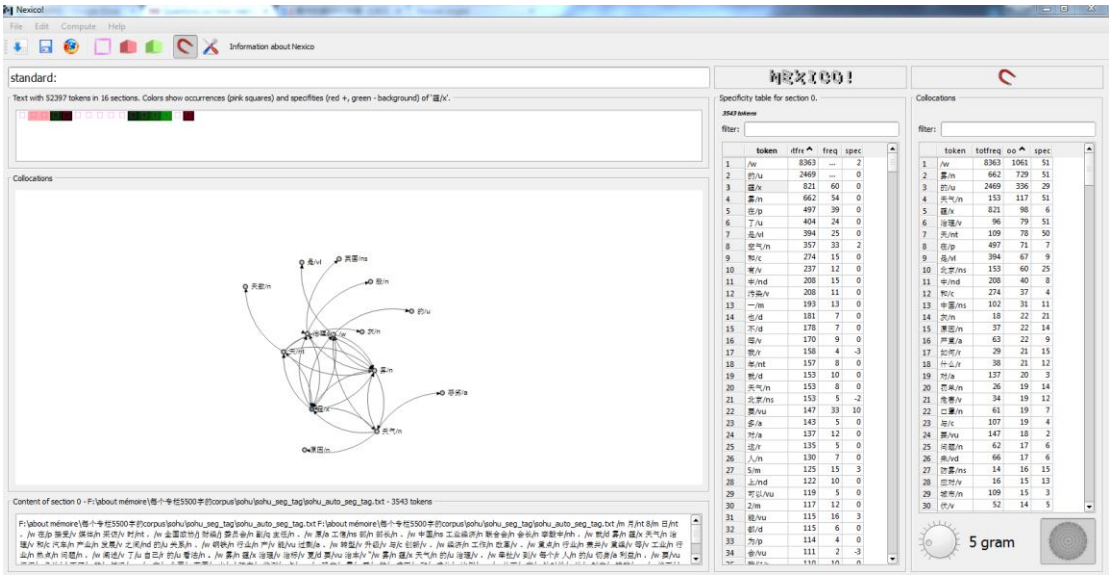
Tableau des termes les plus fréquents du sous-corpus sohu

Nous allons analyser la répartition des termes en fonction de la catégorie lexicale

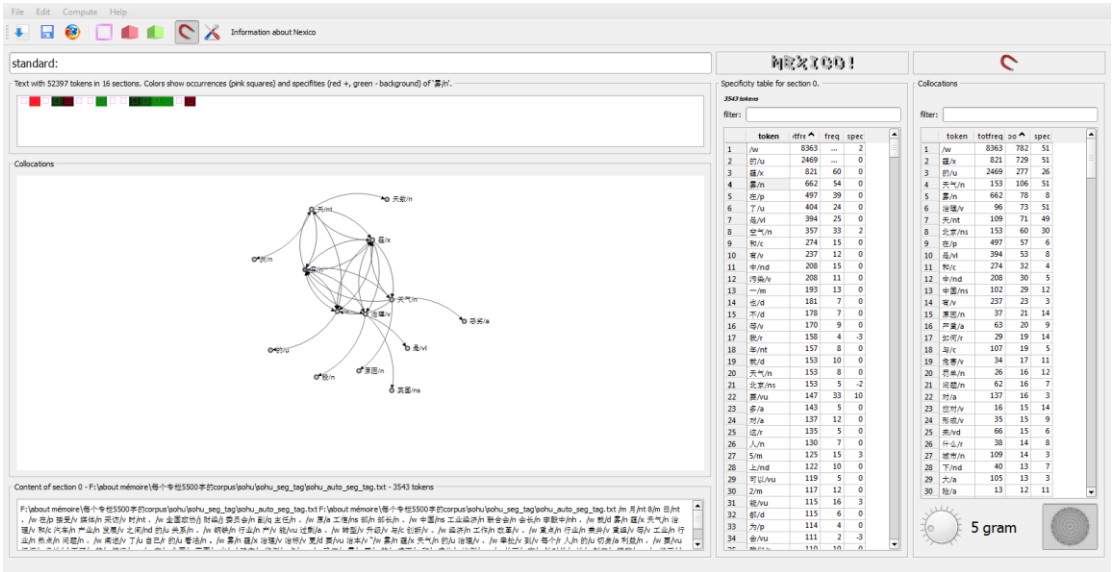
et de leurs liens avec les deux mots-clés selon le sens du mot. Il y a en total 12/34 noms ou termes nominaux, 7/34 verbes ou termes verbaux. Même si cette répartition est similaire que celle du *gov,* les deux sous-corpus ne partagent que les deux mots-clés, *le 天气／n (le temps) et 北京／ns (Pékin).* Alors, au niveau des mots nominaux, par rapport au *gov*, *sohu* ne dispose pas de mots de direction qui indique les régions souffrant du « smog épais » en Chine, par exemple, le *地区／n (la région)， 东部／nd (l'est)，北部／nd (le nord)， 南部／nd (le sud)，*mais le terme *北京／ns (Pékin)* partagé dans les deux sous-corpus montre du doigt directement sur la ville qui subit du smog épais, en autre terme, au niveau géographique, *sohu* localise l'endroit plus précis et ne montre qu'un victime, alors que *gov* parle non seulement plus largement dans l'ensemble en Chine, il précise aussi deux victimes : Pékin et Inner Mongolier. Quant aux verbaux, par rapport au *gov*, qui met accent sur les actions et les mesures prises contre le « smog épais » par le gouvernement chinois (prouvées par les verbes 治理／v (régulariser)，防治／v(prévenir)), la moitié des mots du *sohu* sont des verbes de modalité, tels que le *要／vu(vouloir), 可以／vu(pouvoir), 能／vu(pouvoir)，会／vu(falloir),* il n'y a que un verbe *污染／v(polluer)* qui est lié au « smog épais » sans prise en compte du contexte de l'article. En plus, nous avons remarqué pour la première fois que le pronom « je » est apparu dans la liste des termes les plus fréquents du *sohu*. Ce phénomène pourrait-il résulter des certaines rubriques de *blog*/*forum* manquantes dans le sous-corpus *gov* ? Comme tous les textes du *gov* sont soit des reportages soit des rapports gouvernementaux, l'auteur doit rédiger les textes d'un point de vue très objectif, sans aucune émotion personnelle, alors que dans le *blog* ou *forum*, les gens sont libres d'exprimer leur propre opinion sur un sujet ou un thème, ce qui augmente la fréquence des termes émotionnels et personnels, comme le « je ». Avec cette hypothèse, nous avons ainsi consulté la fréquence de « je » dans les rubriques de *blog* et de *forum* du *sohu*, effectivement, le « je » est placé au cinquième rang des termes les plus fréquents dans la rubrique de forum, cette recherche explique et confirme la raison pour laquelle que l'occurrence du « je » dans les termes les plus fréquents.

4.6.3. Présentation et analyses des mots les plus co-occurrents avec les deux mots-clés

Dans un second temps, voici les termes les plus co-occurrents avec les deux mots-clés 霾 / x(smog) et 雾 / n (brouillard) du *sohu*:



Capture d'écran Les termes les plus co-occurrents avec 霾 / x(smog) du sohu



Capture d'écran Les termes les plus co-occurrents avec 雾 / n(brouillard) du sohu

Listons les co-occurrents avec la traduction du français des deux images dans le tableau ci-dessous :

Les co-occurrents du *smog* du *sohu*					**VS**	**Les co-occurrents du *brouillard* du *sohu***			
Numéro	Tokens	Totfreq	Cooc	Traduction		Token	Totfreq	Cooc	Traduction
1	/w	8363	1061	espace		／w	8363	782	espace
2	雾／n	662	729	le brouillard		霾／n	821	729	le smog
3	的／u	2469	336	auxiliaire		的／u	2469	2872	auxiliaire
4	天气／n	153	117	le temps		天气／n	153	106	le temps
5	霾／x	821	98	le smog		雾／n	662	78	le brouillard
6	治理／v	96	79	régulariser		治理／v	96	73	régulariser
7	天／n	109	78	le ciel		天／nt	109	71	le ciel
8	在／p	497	71	dans		北京／ns	153	60	Pékin
9	是／vl	394	67	être		在／p	497	57	dans
10	北京／ns	153	60	Pékin		是／vl	394	53	être
11	中／nd	208	40	dans ...		和／c	274	32	avec
12	和／c	274	37	avec		中／nd	208	30	dans...
13	中国／ns	102	31	la Chine		中国／ns	102	29	la Chine
14	灰／n	18	22	le cendre		有／v	237	23	avoir
15	原因／n	37	22	la raison		原因／n	37	21	la raison
16	严重／a	63	22	grave		严重／a	63	20	grave
17	如何／r	29	21	comment		如何／r	29	19	comment
18	什么／r	38	21	quoi		与／c	107	19	avec
19	对／a	137	20	envers		危害／v	26	17	abimer
20	罚单／n	26	19	l'amande		罚单／n	26	16	l'amande
21	危害／v	26	19	abimer		问题／n	62	16	le problème
22	口罩／n	61	19	le masque protecteur		对／p	137	16	envers
23	与／c	107	19	avec		应对／v	16	15	affronter
24	要／vu	147	18	vouloir		形成／v	35	15	se former

25	问题 / n	62	17	le problème		来 / vd	66	15	venir
26	来 / vd	66	17	venir		什么 / r	38	14	quoi
27	防雾 / ns	14	16	lutter contre le brouillard		城市 / n	109	14	la ville
28	应对 / v	16	15	affronter		下 / nd	40	13	dessous
29	城市 / n	109	15	la ville		险 / a	13	12	dangereux
30	伏 / n	52	14	volt		大 / a	105	13	grand

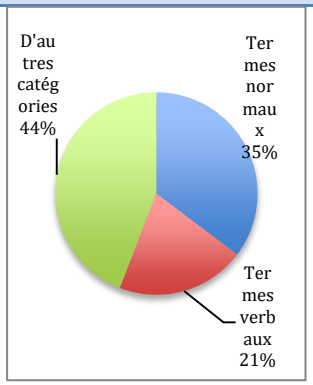
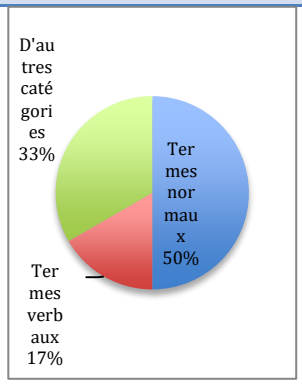
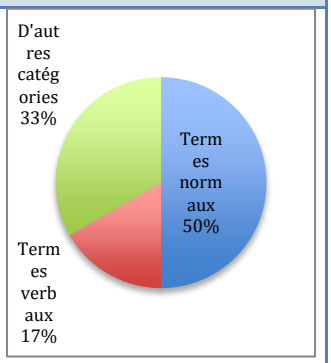
Tableau des termes les plus co-occurrents avec 霾 / n (le smog) et ceux avec 雾 / n (le brouillard) du sohu

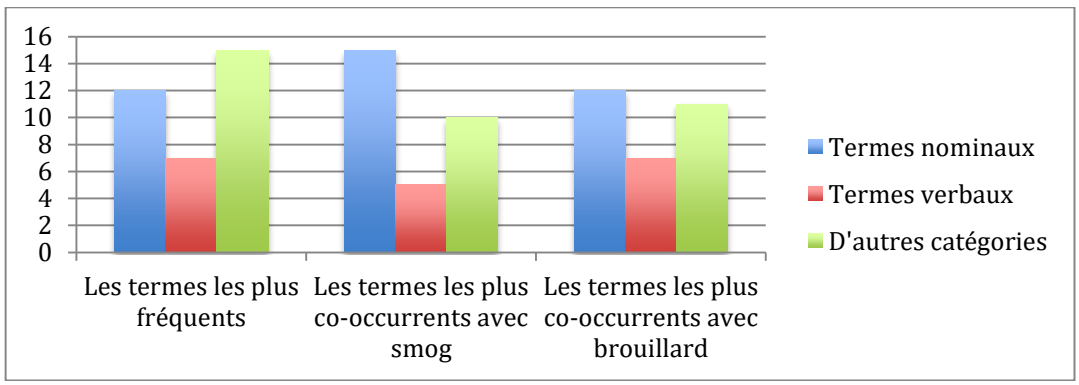
Si nous synthétisons les données statistiques sur les termes les plus co-occurrents avec les deux mots-clés, pour les co-occurrents de *霾/n (le smog)*, nous avons compté 15/30 nominaux et 5/30 verbaux, pour ceux de *雾/n (le brouillard)*, 12/30 nominaux et 7/30 verbaux. En gros, il y a plus de noms que les verbes, ce qui est à l'opposé du *gov*. En plus, nous pouvons en tirer les termes les plus intéressants par rapport à notre sujet : *中国/ns(la Chine)，灰/n(le cendre)，原因/n(la raison)，罚单/n(l'amande)，口罩/n(le masque protecteur)，伏/n(volt)，城市/n(la ville)，治理/v(régulariser)，危害/v(abimer)，防雾/v (lutter contre le brouillard)， 应对/v (affronter), 严重/a(grave)*. Sauf le verbe *治理/v(régulariser)* et *l'adjectif 严重/a(grave)* qui sont déjà apparus dans les co-occurrents *gov*, le reste est spécifique dans *sohu*. En effet, le contenu du sous-corpus *sohu* est beaucoup plus varié en raison du choix des rubriques, mais nous pouvons quand même mettre en contraste ces deux sous-corpus avec ces mots spécifiques. Par rapport au *gov*, qui met en accent sur les actions et les mesures que le gouvernement a prises face au smog épais, *sohu* souligne plutôt la raison *(原因)* d'où vient le smog épais, la nocivité *(危害)* et la protection que les gens prend dans la vie quotidienne en ville *(城市)*. Le 伏 / n(volt), étant l'unité de mesure électrique, relèvent une origine principale du smog épais de la Chine. D'après un reportage du web, cet épais brouillard est la conséquence d'une importante pollution dont les causes principales sont l'industrie, la production d'électricité pour se chauffer - 70% de l'énergie est produite à partir de charbon - et l'explosion du trafic automobile. Depuis 2005, le nombre de voiture a triplé[16] et aujourd'hui, toutes les deux secondes, une voiture fait son apparition en Chine. Les polluants atmosphériques ont ainsi suivi la même progression. Il faut noter que la plupart des automobiles véhiculent dans les villes de la Chine, c'est pour cette raison-là que le terme 城市 / *n(la ville)* et 罚单 / *n(l'amande)* sont mentionnés dans la rubrique d' « automobile » du sous-corpus *sohu*, mais absent dans *gov.* En plus, l'effet toxique du brouillard de pollution inquiètent beaucoup les habitants chinois, ils sont obligés d'apporter les masques protecteurs (口罩) pour se protéger. *Sohu,* en tant que site informel médiatique privé s'intéresse à ces mesures protectrices prises par le peuple chinois.

16 Cf :http://www.asef-asso.fr/mon-air-exterieur/l-actu-ville-sante/2185-pollution-en-chine-pourquoi-comment-quelles-consequences

4.6.4. Graphiques synthétiques des analyses du *sohu*

Rangeons la proportion des termes avec des graphiques dans un tableau comme ce que nous avons fait pour le premier sous-corpus :

Catégorie lexicale	**La proportion des termes les plus fréquents du sous-corpus *sohu***	**La proportion des termes les plus co-occurrents avec « smog » du *sohu***	**La proportion des termes les plus co-occurrents avec « brouillard » du *sohu***
Termes nominaux	12/34	15/30	12/30
Termes verbaux	7/34	5/30	7/30
D'autres catégories	15/34	10/30	11/30
Graphiques			



Graphique synthétique des analyses du sous-corpus sohu

4.6.5. Présentation et analyses de l'état de la répartion des mots-clés du sous-corpus *sohu*

Après ces analyses globales sur *sohu* et la comparaison générale du *sohu* et du *gov*, nous allons rentrer dans le détail du sous-corpus afin de comparer la rubrique partagée dans les deux sous-corpus : *news*. Avec les images capturées du Nexico nous regardons la répartition des mots dans cette rubrique, ensuite, nous rangeons les mots les plus fréquents dans la rubrique news avec leurs catégories lexicales sur le sous-corpus *sohu*.

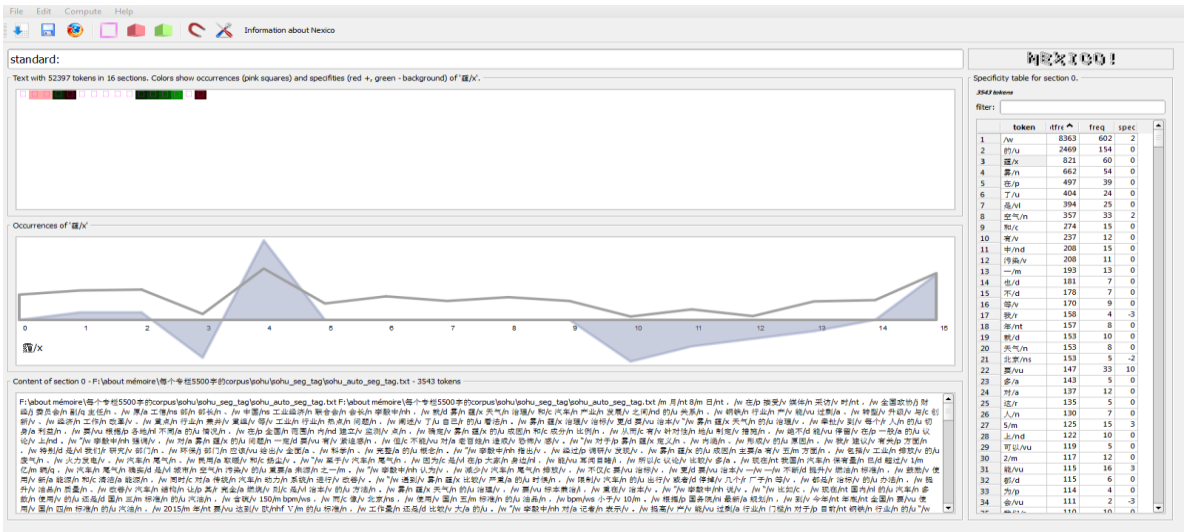


Image capturée du Nexico sur la répartition du 霾／n (smog) du sous-corpus sohu

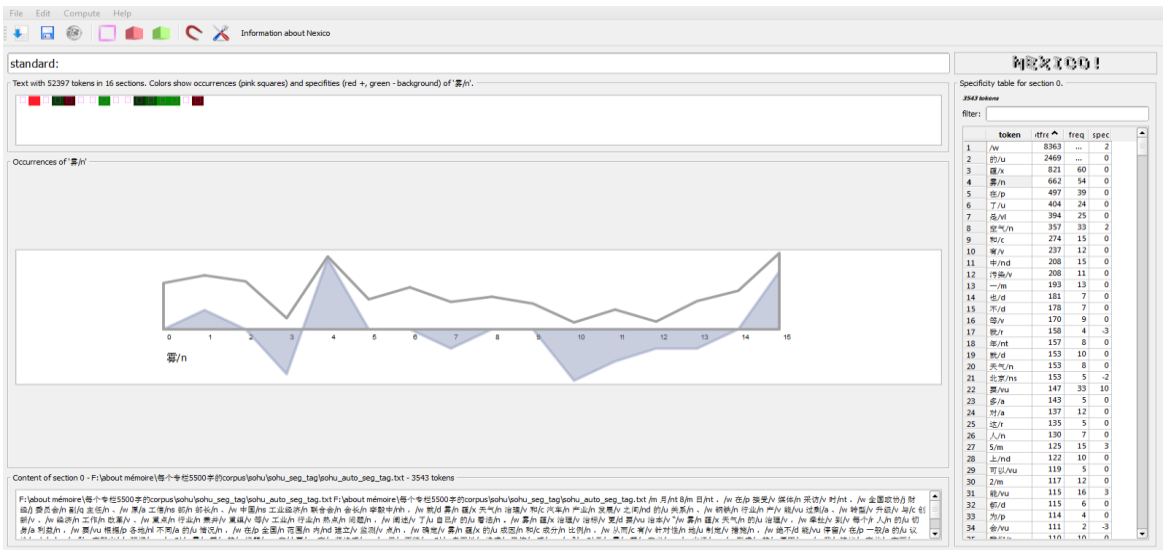


Image capturée du Nexico sur la répartition du 雾／n (brouillard) du sous-corpus sohu

Il y a 52397 tokens répartis dans 16 sections du *sohu*, ce dernier possède à peu près deux fois plus de *tokens* et de *sections* par rapport au *gov*, ainsi, nous voyons dans les deux images que la courbe bleue est beaucoup plus variée, le tableau ci-dessous conclut les caractéristiques de la courbe bleue sur la répartition du « smog » et du « brouillard »:

Section	0 AUTO (automobile)	1 BB (bébés)	2 LUNTAN (bbs)	3 BLOG	4 BUSINESS	5 CITY	6 EDU (éducation)	7 FASHION (mode)	8 FOCUS	9 GONGYI (bien-être public)	10 HEALTH (santé)	11 IT (informatique)	12 MILITARY	13 NEWS	14 SZ (Suzhou)	16 TRAVEL
L'état de l'emploi du mot «smog »	0 (normal)	2 (sur-employé)	-7 (sous-employé)	2 (sur-employé)	-10 (sous-employé)	21 (sur-employé)	0 (normal)	0 (normal)	0 (normal)	0 (normal)	0 (normal)	-8 (sous-employé)	-3 (sous-employé)	-3 (sous-employé)	12 (sur-employé)	0 (normal)
L'état de l'emploi du mot « brouillard »	0 (normal)	3 (sur-employé)	-5 (sous-employé)	0 (normal)	-7 (sous-employé)	11 (sur-employé)	0 (normal)	0 (normal)	-3 (sous-employé)	0 (normal)	0 (normal)	-11 (sous-employé)	-5 (sous-employé)	-3 (sous-employé)	9 (sur-employé)	0 (normal)

Tableau La répartition du 霾／n (smog) et du 雾／n(brouillard) du sous-corpus sohu

4.6.6. Analyses et comparaison de la rubrique « News » partagée du *sohu* et du *gov*

Les images au-dessus permettent de comparer de manière générale les statistiques de cette rubrique dans les deux différents sous-corpus, et de voir s'il y a une divergence dans cette rubrique nommée pareil.

Nom du sous-corpus	Mot-clé	Totfréquence dans le sous-corpus	Fréquence dans la rubrique **news**	Le pourcentage fréquence/totfréquence	L'état de la répartition du mot-clé
GOV	brouillard	339	69	69/339≈0.20	sur-employé
	smog	339	99	99/339≈0.29	sur-employé
SOHU	brouillard	662	33	33/662≈0.05	sous-employé
	smog	821	44	44/823≈0.05	sous-employé

Tableau de comparaison de l'utilisation des mots-clés entre gov et sohu dans la rubrique news

Avec ces données statistiques, la taille du sous-corpus *sohu* est à peu près 2 fois plus grande que celle du *gov* (le chiffre affiché dans la fréquence totale), alors que le chiffre de la fréquence des deux mots-clés dans la rubrique de « *news* » du *sohu* est presque 2 fois moins grand que celui du *gov*. Il faut noter que le pourcentage du "*fréquence/fréquence totale*" est sous un même sous-corpus, qui montre en quelque sorte que le résultat du pourcentage est significatif. Si nous comparons le résultat du *gov* avec celui du *sohu*, le pourcentage d'apparition des deux mots-clés dans la rubrique « news » du sous-corpus *gov* est presque 4 fois plus de celui du *sohu*, malgré la taille du dernier est 2 fois plus grande que l'autre. Autrement dit, le *gov* utilise beaucoup plus de « brouillard » et de « smog » dans la rubrique « news » par rapport au *sohu*. En plus, avec les deux tableaux[17] précédents qui manifestent la répartition des deux mots-clés dans les rubriques des deux sous-corpus, les deux mots-clés ont été respectivement sur-employés (*gov*) et sous-employé (*sohu*) dans la rubrique « news », ceci prouve d'une autre manière l'état de l'utilisation des mots-clés que nous venons de comparer.

Ensuite, il s'agit de contraster les termes les plus fréquents dans la rubrique de « news » dans les deux sous-corpus :

[17] Cf : P25 « Tableau de la répartition du 霾/n(le smog) et du 雾/n (le brouillard) du *gov* » et P45 « Tableau de la répartition du 霾/n(le smog) et du 雾/n (le brouillard) du *sohu* ».

Les termes les plus fréquents dans « News » du SOHU					VS	Les termes les plus fréquents dans « News » du GOV			
Numéro	Token	Totfreq	Freq	Traduction		Token	Totfreq	Freq	Traduction
1	/w	8363	575	espace		霾/x	339	99	le smog
2	的/u	2469	135	auxiliaire		日/nt	253	78	la journée
3	霾/x	821	44	le smog		雾/n	271	69	le brouillard
4	空气/n	357	36	l'air		的/u	829	60	auxiliaire
5	污染/v	208	35	la pollution		有/v	247	58	avoir
6	在/p	497	35	dans		地/u	203	56	adverbe
7	2/m	117	33	le numéro		地区/n	227	51	la région
8	5/m	125	33	le numéro		等/v	229	44	attendre
9	雾/n	662	33	le brouillard		中/nd	188	43	le centre
10	pm/ws	79	31	particule fine par mètre cube		东部/nd	174	37	l'est
11	北京/ns	153	30	Pékin		月/nt	126	35	le mois
12	大雾/n	35	21	le brouillard		时/nt	105	31	l'heure
13	人/n	130	21	l'être humain		摄/v	36	30	photographier
14	和/c	274	21	avec		新华社/ni	48	30	la Presse XINHUA
15	了/u	404	20	auxiliaire		北部/nd	97	29	le nord
16	中/nd	208	20	le centre		在/p	187	28	dans
17	是/vl	394	18	être		南部/nd	97	27	le sud
18	影响/v	70	16	influencer		和/c	259	24	avec
19	多/a	143	16	nombreux		笼罩/v	28	23	couvrir
20	天气/n	153	16	le temps		级/n	76	23	le degré
21	天/nt	109	15	la journée		08/m	65	22	le numéro
22	等/v	170	15	attendre		10/m	47	21	le numéro
23	颗粒/n	62	14	le particule		西部/nd	49	19	l'ouest

24	有/v	237	14	avoir		或/d	80	19	ou
25	高速/a	15	13	la grande vitesse		部分/n	66	18	la partie
26	我们/r	110	13	nous		1/m	77	18	le numéro
27	就/d	153	13	adverbe		到/v	98	18	arriver
28	一/m	193	13	un		11/m	23	17	le numéro
29	质量/n	77	12	la qualité		5/m	73	17	le numéro
30	都/d	115	12	tous		发/v	22	16	distribuer
31	烟花/n	19	11	le feu d'artifice		度/n	30	16	le degré
32	伦敦/ns	23	11	Londre		华北/nl	55	16	le nord
33	物/n	49	11	l'objet		至/v	56	16	arriver
34	但/c	89	11	mais		将/d	87	16	adverbe

Tableau Les termes les plus fréquents dans la rubrique « News » du sohu VS ceux du gov

Les deux couleurs soulignent les termes pareils (vert) et les termes différents (rouge) entre *gov* et *sohu*. Ici nous ne regardons que les mots nominaux et les verbaux. Au niveau de répartition des mots en fonction de la catégorie lexicale, pour *sohu*, la proportion des nominaux est de 14/34, celle des verbaux est de 5/34, alors que pour *gov*, les nominaux sont de 15/34 et les verbaux sont de 6/34. Les deux sous-corpus sont relativement à la même hauteur en cette matière. Quant à la répartition des mots-mêmes, il y a juste les deux mots-clés qui sont partagés dans « News » sur ces deux sous-corpus, les termes différents sont d'ailleurs bien nombreux, tels que les *空气/n (l'air), 污染/v(polluer), pm/ws (particule fine par mètre cube), 北京/ns (Pékin),大雾/n (brouillard), 影响/v (influencer)，天气/n (le temps)，颗粒/n (la particule)，高速/a (la grande vitesse), 我们/r (nous)，质量/n (la qualité), 烟花/n (le feu d'artifice), 伦敦/ns(Londres)* sur *sohu*, il semble que dans la rubrique « News » du *sohu,* on s'intéresse à plusieurs aspects sur le brouillard de pollution, surtout le composant, le type, l'origine du smog épais reflétés par les termes nominaux (*pw, la particule, le brouillard*), on mentionne aussi deux origines d'où vient cette pollution atmosphérique : *高速/a (en grande vitesse)* et *烟花/n (le feu d'artifice)*, car la concordance de « la grande vitesse » du *sohu* dans « News » désigne en fait les automobiles en autoroute, comme vous vous en souvenez, nous avons dit dans la partie précédente que l'émanation des automobiles font partie des origines du smog épais, en plus, certains experts ont confirmé que le feu d'artifice pollue l'atmosphère et peut produire le brouillard de pollution ; En plus de ces deux mots, le pronom « nous » dans « News » du *sohu* joue un rôle similaire que « je », qui s'exprime un point de vue ou des sentiments personnels, ce terme correspond à la nature de *sohu*. Le terme « Londres » apparaît 11 fois dans « News », d'après le résultat de recherche sur sa concordance, on parle des origines et des nocivités produits par le smog épais dans l'époque à Londres. Toutes ces informations visent à mettre la situation actuelle en Chine en alerte, et proposer des mesures protectrices face à cette pollution grave atmosphérique. Pour *gov*, comme nous avons déjà rencontré les occurrences et les ont expliquées, nous ne le répétons plus ici.

4.6.7. Bilan d'analyse du sous-corpus *sohu*

Par rapport au *gov, sohu* est beaucoup moins strict et maîtrisé au niveau de l'organisation des textes. Cette caractéristique émerge dès la mise en propre du corpus, les formes d'organisation du contenu varient en fonction de la rubrique. En plus, si nous comparons l'occurrence des mots, les termes les plus fréquents ne sont pas toujours croisés avec ceux des plus co-occurrents des deux mots-clés. Cependant, la répartition des mots selon la catégorie lexicale du *sohu* est relativement similaire, qui est à l'opposé du *gov* est le nombre des nominaux est supérieur à celui des verbaux. En outre, les termes nominaux du *gov* indiquent souvent les régions subissant du brouillard de pollution, alors que chez *sohu*, les nominaux sont multifonctions, qui nous font part des origines, de la nocivité, des composants et du type du smog et des mesures de protection venant des habitants chinois. Quant aux verbaux, le *gov* met en accent sur les mesures prises par le gouvernement chinois, alors que les verbes du *sohu* sont souvent des verbes de modalité, qui expriment la volonté ou une espérance du peuple chinois. Une autre caractéristique du *sohu* est qu'il y a des pronoms personnels comme « je » et « nous », cela relève la nature privée du *sohu*, où les gens sont libres d'exprimer leur point de vue ou leurs sentiments sur un sujet, qui n'est pas le cas du site institutionnel.

3.7. Analyse du sous-corpus *people*

4.7.1. Processus des analyses du sous-corpus people

Il y a deux raisons essentielles qui nous poussent à appliquer le même processus d'analyse sur *people* comme sur *sohu*. Premièrement, au niveau du nombre des rubriques, le sous-corpus *people* en possède d'une vingtaine, voire plus que *sohu*, cela alourdit objectivement les tâches de travail analytique en détail sur les rubriques; ensuite, en tant que dernier sous-corpus, nous allons combiner les résultats du *people* avec ceux des deux premiers sous-corpus, cette tâche nécessite l'accomplissement des analyses sur le *people* lui-même. Cette organisation permet d'éviter l'aller-retour du

travail analytique sur *people*.

4.7.2. Présentation et analyses des mots les plus fréquents du *people*

Voici donc les termes les plus fréquents dans le sous-corpus *people* :

Numéro	Token	Totfreq	Traduction
1	/w	10288	espace
2	的/u	2188	auxiliaire
3	霾/x	1040	smog
4	雾/n	826	brouillard
5	在/p	454	dans
6	了/u	396	auxiliaire
7	有/v	394	avoir
8	是/vl	387	être
9	日/nt	361	le jour
10	和/c	328	avec
11	污染/v	290	polluter
12	中/nd	279	dans…
13	北京/ns	270	Pékin
14	等/v	267	etc.
15	空气/n	245	l'air
16	一/m	241	un
17	年/nt	229	l'année
18	天气/n	226	le temps
19	中国/ns	213	la Chine
20	大/a	203	grand
21	地/u	201	adverbe
22	地区/n	195	la région
23	月/nt	193	le mois
24	也/d	183	aussi
25	不/d	179	non
26	天/nt	170	la journée
27	将/d	170	adverbe
28	5/m	169	numéro
29	被/p	166	préposition
30	对/a	166	envers
31	为/p	161	pour
32	时/nt	160	l'heure
33	多/a	151	nombreux
34	2/m	150	numéro

Tableau des termes les plus fréquents du people

Parmi les 34 termes les plus fréquents du *people*, il y a 10/34 nominaux et 4/34 verbaux, le reste qui sont plutôt des termes nécessaires dans les textes de presse, occupe plus de moitié de portion par rapport à celle des nominaux et des verbaux. À part des deux mots-clés, il n'y a que *污染/v(polluer)，北京/ns(Pékin)，空气/n(l'air)，中国/ns(la Chine) et 地区/n(la région)* sont liés au sujet de recherche, dont « polluer » et « la région » sont deux termes croisés avec ceux du sous-corpus *gov*, « l'air » et « Pékin » sont les mots croisés avec *sohu*. Le terme « *la Chine* » est premièrement apparu dans le corpus, après avoir consulté la répartition du terme « la Chine » dans le sous-corpus *people*, nous avons eu les rubriques suivantes dans lesquelles « la Chine » est fréquemment présent :

Rubriques[18]	Classement	Fréquence dans la rubrique
bbs	13	16
blog	12	16
news	19	12
edu	19	12
energy	32	10
ezheng (forum)	17	23

Tableau de la répartition du 中国/ns(la Chine) dans les rubriques du people

On parle du smog épais de la Chine par rapport à d'autres pays après avoir consulté la concordance du *中国/ns(la Chine)* dans les rubriques mentionnées, par exemple l'Angleterre (Londres) et la France (Paris), la comparaison incarne notamment sur trois aspects : *pour tirer une leçon à travers des expériences échouée des autres pays* ; *pour conclure les origines du brouillard de pollution en analysant les cas d'étrangers* ; *pour prendre des mesures spécifiques en comparant la situation de la Chine avec celle d'autrui*. Avec le tableau ci-dessus, nous remarquons que les types des rubriques varient d'une à l'autre, les rubriques « informelles » et « institutionnelles » sont tous présentes, tels que le *bbs*, le *forum* et le *lingdao*. Ce phénomène témoigne de

[18] Les rubriques dans lesquelles le termes « la Chine » est fréquemment présent.

l'attention prêtée au smog épais par le gouvernement ainsi que par le peuple chiois. On essaie tous de chercher les solutions pour lutter contre la pollution de l'air, la comparaison et l'apprentissage sur des problèmes similaires dans d'autres pays favorisent la connaissance et la prise de mesures contre le smog épais en Chine.

4.7.3. Présentation et analyses des mots les plus co-occurrents avec les deux mots-clés du people

Ensuite, nous allons lister les termes les plus co-occurrents des deux mots-clés dans *people* et voir qu'est-ce qu'il donnera comme résultat :

Les mots les plus co-occurrents du *smog* du *people*					**VS**	**Les mots les plus co-occurrents du *brouillard du people***			
Numéro	Token	Totfreq	Cooc	Traduction		Token	Totfreq	Cooc	Traduction
1	/w	10288	1249	espace		/w	10288	957	espace
2	雾 / n	826	833	le brouillard		雾 / n	826	833	le brouillard
3	的 / u	2188	305	auxiliaire		的 / u	2188	272	auxiliaire
4	天气 / n	226	160	le temps		天气 / n	226	145	le temps
5	天 / nt	170	93	le ciel		天 / nt	170	82	le ciel
6	是 / vl	387	73	être		是 / u	387	62	être
7	在 / p	454	67	dans		治理 / v	111	58	régulariser
8	治理 / v	111	63	régulariser		在 / p	454	52	dans
9	有 / v	394	60	avoir		北京 / ns	270	35	Pékin
10	霾 / n	1040	58	le smog		对 / p	166	34	envers
11	中 / p	279	49	dans...		将 / d	170	32	adverbe
12	将 / d	170	46	adverbe		有 / v	394	32	avoir
13	出现 / v	91	40	apparaître		严重 / a	108	30	grave
14	了 / u	396	40	particule finale		影响 / v	94	29	influencer
15	治 / v	44	39	résoudre		出现 / v	91	28	apparaître
16	北京 / ns	270	39	Pékin		中国 / ns	213	28	la Chine
17	严重 / a	108	37	grave		灯 / n	45	27	le feu de circulation
18	对 / p	166	37	envers		地区 / n	195	27	la région
19	度 / n	95	35	le degré		中 / p	279	25	dans...
20	地区 / n	195	33	la région		笼罩 / v	31	23	couvrir
21	地 / u	201	33	auxiliaire		问题 / n	92	20	le problème
22	中国 / ns	213	33	la Chine		地 / u	201	20	auxiliaire
23	预警 / v	66	32	alerter		专家 / n	68	17	l'expert

24	影响 / v	94	32	influencer		关于 / p	23	16	sur
25	重 / a	125	27	lourd		遭遇 / v	23	16	subir
26	多 / a	151	27	nombreux		持续 / v	39	16	durer
27	笼罩 / v	31	26	couvrir		市民 / n	42	16	le citoyen
28	口罩 / n	105	26	le masque protecteur		之 / u	81	16	auxiliaire
29	来 / vd	99	23	venir		口罩 / n	105	16	le masque protecteur
30	也 / d	183	22	aussi		要 / vu	135	16	vouloir

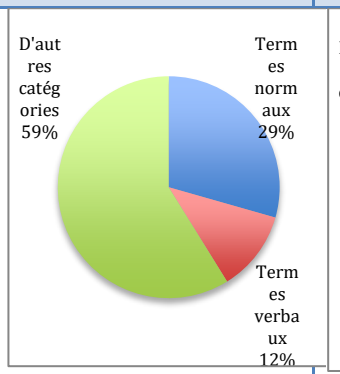
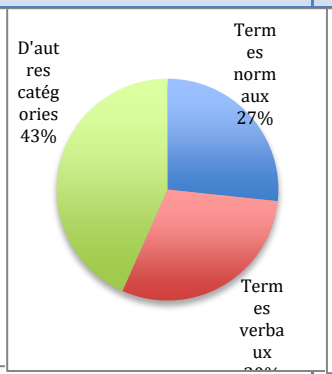
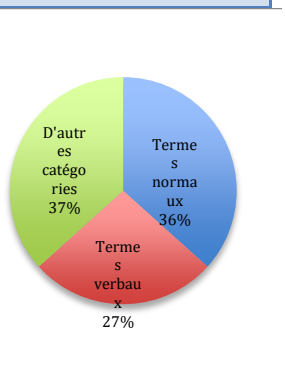
Tableau des termes les plus co-occurrents avec les deux mots-clés du people

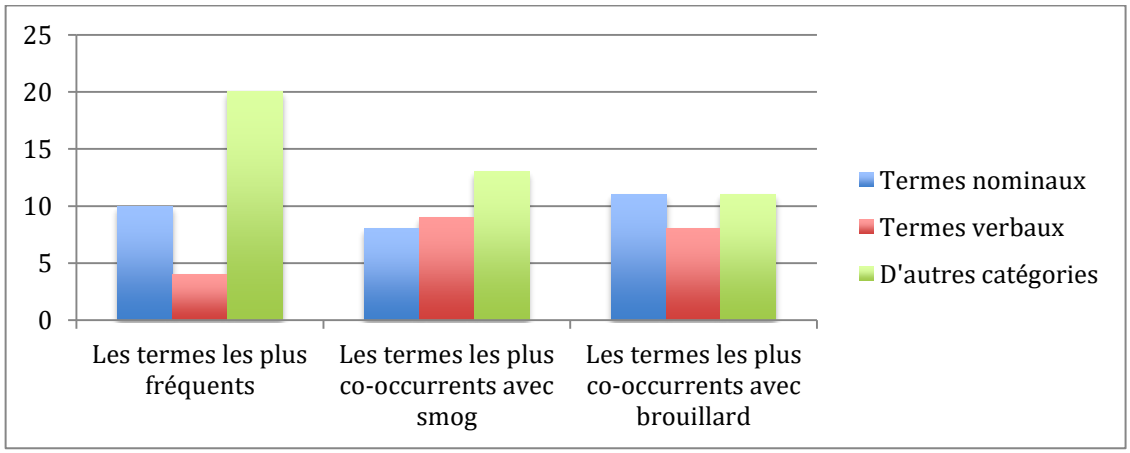
Parmi les co-occurrents du 霾 / n(smog), les nominaux occupent 8/30, les verbaux occupent 9/30, alors que la proportion des co-occurrents nominaux et celle des co-occurrents verbaux du 雾 / n(brouillard) est respectivement de 11/30 et 8/30. Par rapport à *gov* (V[19]>N) et à *sohu* (V<N), la répartition de la catégorie lexicale des mots du *people* n'est pas si distinguée, car il n'y a pas de borne bien nette qui dit « qui est supérieur à qui ». En matière du sens, les termes apparus dans *people* font une synthèse de tout le corpus, ce dernier trait se reflète par des mots partagés avec les deux premiers sous-corpus, par exemple, il y a comme termes nominaux : *问题 / n(le problème)*，*口罩 / n(le masque protecteur)* croisés entre *people* et *sohu*，*地区(la région)* chevauchés entre *gov* et *people* et *北京 / n(Pékin)*，*天气 / n(le temps)* partagés dans les trois sous-corpus; comme verbaux : il y a *笼罩 / v(couvrir)*，*预警 / v(alerter)*，*影响 / v(influencer)*，*出现 / v(apparaître)*，*治 / v(résoudre)* croisés dans *gov* et *people*, et *治理 / v(régulariser)*，*污染 / v(polluer)* chevauchés dans les trois sous-corpus. Il est à résumer que sur le sujet « smog épais en Chine », le people synthétise le côté " *les mesures assurées et prises par le gouvernement chinois*" manifestées par les termes verbaux croisés avec ceux du *gov*, et le côté " *les problèmes et la nocivité que produit le smog épais pour la société chinoise* " introduit par les termes nominaux avec ceux du *sohu*. Au niveau de la nature du sous-corpus, ces constats confirment la nature du *people* : à la fois institutionnelle et médiatique comme nous l'avons déjà mentionné dans la partie où nous présentons la nature de chaque corpus. La seule divergence qui diffère le *people* des deux autres sous-corpus se fait marquer par deux mots : *专家 / n(l'expert)* et *中国/ns(la Chine)*. Nous avons eu le pronom personnel *我 / n(je)* dans le sous-corpus *sohu* et expliqué que ce mot désigne un sentiment personnel et subjectif des gens face au sujet de « *smog épais en Chine* ». Alors que le mot « je » est contraire en matière d'énoncé du mot *专家 / n(l'expert)*, qui nous transmet plutôt des messages scientifiques et objectifs sur ce problème, sur son origine, sa nocivité et d'autres aspects concernées. Pour le terme *中国/ns(la Chine)*, nous avons déjà analysé dans les termes les plus fréquents, nous ne le répétons plus ici.

[19] V est l'abréviation des verbes ou des termes verbaux, et N est l'abréviation des noms et des termes nominaux. V<N signifie que le nombre des verbes (termes verbaux) est inférieur que celle des noms (termes nominaux). V>N est le contraire.

4.7.4. Graphiques synthétiques des analyses du *people*

Seront intégrés dans un tableau les informations sur la proportion des termes les plus fréquents et la proportion des termes les plus co-occurrents avec les deux mots-clés du *people* :

Catégorie lexicale	**La proportion des termes les plus fréquents du *people***	**La proportion des termes les plus co-occurrents avec « smog » du *people***	**La proportion des termes les plus co-occurrents avec « brouillard » du *people***
Termes nominaux	10/34	8/30	11/30
Termes verbaux	4/34	9/30	8/30
D'autres catégories	20/34	13/30	11/30
Graphique			



Graphiques synthétiques des analyses du *people*

4.7.5. Présentation et analyses de l'état de la répartition des deux mots-clés dans les rubriques du *people*

Le résultat sera présenté à la base des deux images capturées sur l'état de la répartition des deux mots-clés dans les rubriques du *people*

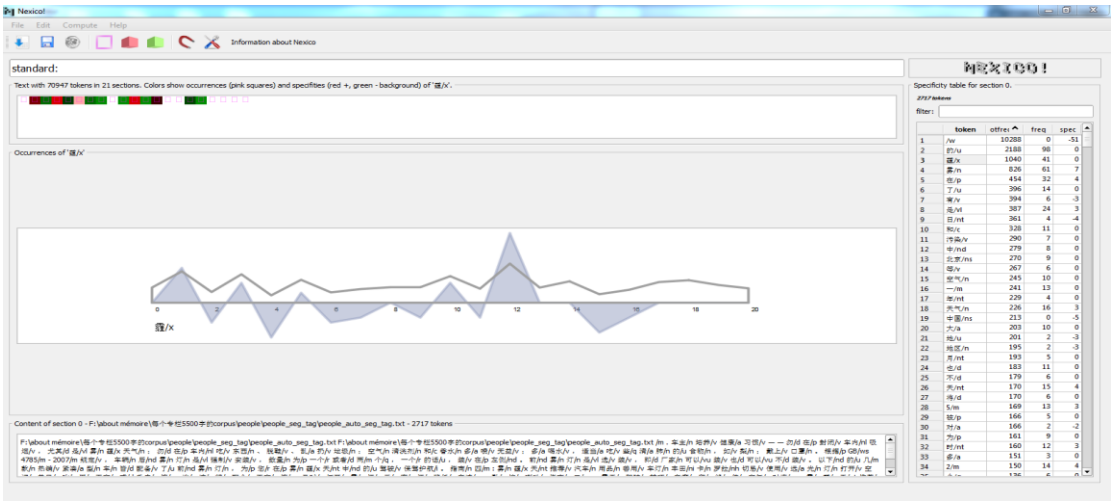


Image capturée du Nexico La répartition du 霾／n(smog) du sous-corpus people

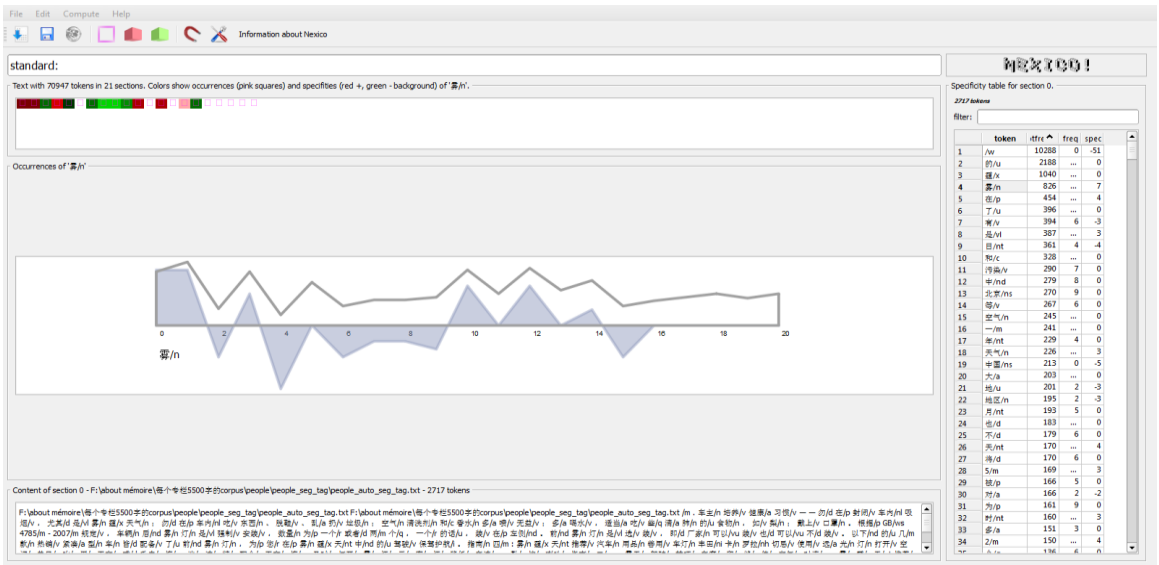


Image capturée du Nexico La répartition du 雾／n(brouillard) du sous-corpus people

Il y a au total 70947 tokens répartis dans 21 sections (rubriques) dans *people*, par rapport au *sohu*, *people* a 5 sections de plus et à peu près 20000 tokens ; et par rapport au *gov*, il a 13 sections de plus et environ 3 fois plus de tokens. L'information du *people* est donc la plus enrichie parmi les trois. Commençons avec l'état de la variation des deux courbes sur les deux mots-clés dans toutes les rubriques sélectionnées dans cette partie :

Section[20]	0 AUTO	1 BBS	2 BJ	3 BLOG	4 CCNEWS	5 NEWS	6 EDU	7 ENERGY	8 ENV	9 EZHENG	10 FINANCE	11 HEALTH	12 HEBEI	13 LEGAL	14 LIANGHUI	15 LINGDAO	16 MIL	17 PAPER	18 POLITICS	19 TRAVEL	20WORLD
L'état de répartition sur l'emploi du « smog »	0 (normal)	7 (sur-employé)	-4 (sous-employé)	4 (sur-employé)	-7 (sous-employé)	2 (sur-employé)	-4 (sous-employé)	-3 (sous-employé)	0 (normal)	-3 (sous-employé)	4 (sur-employé)	-3 (sous-employé°	14 (sur-employé)	0 (normal)	0 (normal)	-6 (sous-employé)	-3 (sous-employé)	0 (normal)	0 (normal)	0 (normal)	0 (normal)
L'état de répartition sur l'emploi du « brouillard »	7 (sur-employé)	7 (sur-employé)	-4 (sous-employé)	4 (sur-employé)	-8 (sous-employé)	0 (normal)	-4 (sous-employé)	-2 (sous-employé)	-2 (sous-employé)	-3 (sous-employé)	5 (sur-employé)	0 (normal)	5 (sur-employé)	0 (normal)	2 (sur-employé)	-4 (sous-employé)	0 (normal)	0 (normal)	0 (normal)	0 (normal)	0 (normal)

Tableau de l'état de répartition du 雾／n(brouillard) et du 霾／n(smog) du people

[20] Les rubriques sont respectivement : Automobile, Forum, Pékin, Blog, Les nouvelles des entreprises gouvernementales, Les nouvelles, Éducation, Énergie, Environnement, Forum gouvernemental, Finance, Santé, Hebei, Légalité, Deux Sessions, Les dirigeants chinois, Militaire, Le journal, Politique, Voyage et Le monde.

Avec le résultat d'analyse sur la répartition des deux mots-clés dans les rubriques,

- Sur-employés : *bbs, blog, news, finance et hebei* ;
- Sous-employés : *bj, ccnews,edu, energy, ezheng, health, lingdao, mil*

Les deux mots-clés sont soit sur-employés soit sous-employés dans ces rubriques relevées et mentionnées ci-dessus dans le tableau. Comme la valeur absolue du chiffre de spécificité n'est pas très grande (aucune valeur absolue est supérieure à 10) , nous pouvons dire que le smog épais (雾霾) est bien fréquent dans ces rubriques relevées. Si nous observons le résultat d'analyse sur l'état d'emploi des mots-clés, nous pouvons remarquer que le sujet « *smog épais en Chine* » intéressent tous les milieux, que ce soit au niveau gouvernemental (*ccnews, news, ezheng, lingdao*), ou social (*auto, bbs, blog, bj,hebei*), voire du domaine de santé(*health*), de finance (*financce*), d'éducation(*edu*), d'énergie(*energy*), et militaire(*mil*). Il s'agit presque de tous les aspects de la vie quotidienne des Chinois. Il faut noter que *people* ouvre deux rubriques respectivement pour la capitale de la Chine (*bj*) et la province du Hebei (*hebei*), l'importance de la capitale est tellement grande que l'on ouvre une rubrique spécifique, quant à la rubrique de Hebei, nous pouvons l'expliquer dans deux niveaux, une de sa position géographique : la province du Heibei se trouve juste à côté de Pékin, qui apparaît comme une périphérie supplémentaire de la capitale, il joue un rôle similaire comme l'île de France à Paris ; l'autre de son industrie développée : la présence de charbon en fait d'une part la province au développement dynamique, d'autre part, Hebei est la province la plus polluée de Chine[21], les poussières de charbon font partie principale de la pollution atmosphérique. Voilà pourquoi nous en tirons les deux rubriques dans lesquelles les deux mots-clés ont été bien marqués.

4.7.6. Bilan d'analyses du sous-corpus *people*

En résumé, *people* se marque par trois caractéristiques :

- La caractéristique qui prévaut du *people* est sa généralité qui synthétise les deux

[21]cf :http://www.lemonde.fr/planete/article/2014/02/20/a-xingtai-ville-la-plus-polluee-de-chine-on-ne-voit-plus-le-ciel_4370513_3244.html/

points relevés du *sohu* et du *gov* au niveau du sens des mots croisés, ceci confirme la nature du *people* qui est à la fois insititutionnelle comme *gov* et médiatique comme *sohu*:

- ''les mesures assurées et prises par le gouvernement chinois'' manifestées par les termes verbaux croisés avec ceux du *gov*;
- ''les problèmes et la nocivité que produisent le smog épais dans le territoire chinois'' introduisent par les termes nominaux chevauchés avec ceux du *sohu*.

- La deuxième caractéristique se traduit par deux nouveaux mots exclusifs du people : *专家／n(l'expert) et 中国／n(la Chine)*. Le premier terme en opposé du terme « je » du *sohu*, qui transmet des messages plutôt objectifs et scientifiques sur le « smog épais en Chine » ; alors que le deuxième met en contraste le problème environnemental de la Chine avec d'autres pays, comme l'Angleterre, pour en tirer des expériences et trouver des solutions plus adaptées face à nos spécificités. Ce point n'est relevé par aucun des deux autres sous-corpus.

- Voici la troisième caractéristique : Au niveau de la catégorie lexicale, le *people* n'est pas si distingué par rapport aux deux autres sous-corpus à cet égard, cela se témoigne notamment par l'équivalence similaire du nombre des verbaux (8/30) et des nominaux (9/30) dans le groupe « la proportion des termes les plus co-occurrents avec smog du *people*[22]».

4.8. Analyses et comparaison en interne entre rubriques similaires parmi *gov*, *people* et *sohu*

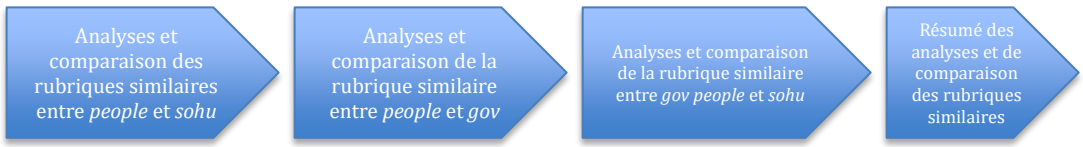
Comme le *people* est un site qui partage les caractéristiques institutionnelles du *gov* et informelles du *sohu*, nous allons analyser et comparer la performance de la présence des mots-clés des entre rubriques similaires autour du *people* parmi les trois sous-corpus.

Il y a des rubriques de nature institutionnelle partagées avec celles du *gov*, telles que le « *ccnews* » (les nouvelles du Parti Communiste Chinois) et le « *lingdao* »(les

[22] Cf : La graphique de la proportion des termes avec leur catégorie lexicale du people.

dirigeants), les rubriques d'origine médiatique informelles sont aussi présentes, comme « *auto* »(automobile),« *bbs* »(forum),« *blog* »,« *finance* », « *edu* »(éducation),« *health* »(santé),« *mil* »(militaire) et « *travel* » (voyage). Afin de mettre en comparaison ces deux types de rubriques, nous commençons avec la comparaison du *people* et du *sohu* avec les rubriques similaires, puis le *people* et le *gov*.

4.8.1. Processus des analyses et des comparaisons en interne entre les rubriques similaires parmi *gov*, *people* et *sohu*



4.8.2. Analyses et comparaison des rubriques similaires entre *people* et *sohu*

Il y a plusieurs rubriques chez *people* comme chez *sohu*, et le nombre des mots sont tous autour de 5500, mettons d'abord en contraste l'état de répartition des deux mots-clés des rubriques similaires entre ces deux sous-corpus.

	Mots-clés	Auto	BBS	Blog	Business	Edu	Health	Military	Travel
SOHU	smog	normal	sous-employé	sur-employé	sous-employé	normal	normal	sous-employé	normal
	brouillard	normal	sous-employé	normal	sous-employé	normal	normal	sous-employé	normal
PEOPLE	smog	normal	sur-employé	sur-employé	sur-employé	sous-employé	sous-employé	sous-employé	normal
	brouillard	sur-employé	sur-employé	sur-employé	sur-employé	sous-employé	normal	normal	normal

Tableau L'état de répartition des deux mots-clés dans les rubriques similaires entre sohu et people

À travers ce tableau, nous constatons que les deux mots-clés sont bien présents dans quatre rubriques de nature médiatique même informelle partagées du *sohu* et du

people : « *BBS* », « *Blog* », « *Business* » et « *Military* », car au moins trois quarts des cases sont remplies par sur-employés/sous-employés. Nous ciblons donc ces quatre rubriques et regardons la performance des termes les plus fréquents dans ces quatre rubriques similaires.

4.8.2.1. Analyses et comparaison de la rubrique « BBS » partagée du *people* et du *sohu*

Commençons l'analyse et la comparaison avec la première rubrique « BBS » partagée du *people* et du *sohu*.

Les mots les plus fréquents dans rubrique BBS du *sohu*				**VS**	**Les mots les plus fréquents dans rubrique BBS du *people***		
Numéro	Token	Freq	Traduction		Token	Freq	Traduction
1	/w	515	espace		/w	530	espace
2	的/u	210	auxiliaire		的/u	144	auxiliaire
3	了/u	83	auxiliaire		霾/x	85	le smog
4	我/r	59	je		雾/n	72	le brouillard
5	在/p	40	dans		了/u	42	auxiliaire
6	着/u	38	auxiliaire		是/vl	41	être
7	尸/n	30	le cadavre		有/v	24	avoir
8	丧/v	30	perdre		我/r	20	je
9	不/d	29	non		在/p	18	dans
10	人/n	25	l'être humain		人/n	17	l'être humain
11	霾/x	25	le smog		就/d	17	adverbe
12	年/nt	24	l'année		中国/ns	16	la Chine
13	雾/n	23	le brouillard		治理/v	14	régulariser
14	是/vl	21	être		啊/e	13	exclamation
15	看/v	20	voir		就是/u	13	auxiliaire
16	胡斐/nh	19	le nom personnel		什么/r	13	quoi
17	一/m	19	un		一个/q	13	un
18	都/d	18	tous		好/a	12	bon
19	就/d	18	adverbe		汽车/n	12	la voiture
20	他/r	15	il		大/a	12	grand
21	还/d	15	adverbe		没有/v	11	ne pas avoir
22	和/c	15	avec		还/d	11	en plus

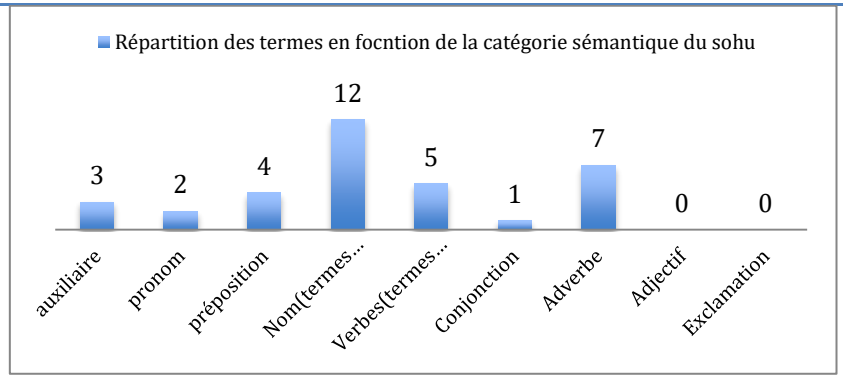
23	去年/nt	14	l'année dernière		都/d	11	tous
24	已经/d	14	déjà		天气/n	11	le temps
25	让/p	14	préposition		要/vu	10	vouloir
26	寝室/n	13	la chambre		也/d	10	aussi
27	也/d	13	aussi		人人/n	9	tout le monde
28	出现/v	12	apparaître		这样/r	9	comme ça
29	空气/n	12	l'air		不/d	9	non
30	当/p	11	préposition		和/c	9	avec
31	去/v	11	aller		吧/e	8	exclamation
32	被/p	11	préposition		能/vu	8	pouvoir
33	中/d	11	dans...		对/p	8	envers
34	这/r	11	cela		最/d	8	le plus

Tableau des mots les plus fréquents dans la rubrique « BBS » du sohu et ceux du people

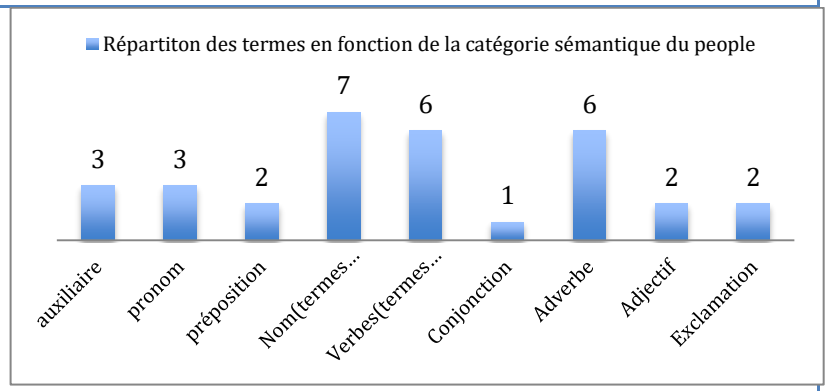
Calssons d'abord les termes de la rubrique « BBS » des deux sou-corpus selon leur catégorie lexicale :

Les termes dans la rubrique « BBS » du *sohu* :	**Les termes dans la rubrique « BBS » du *people* :**
Auxiliaire : 3	**Auxiliaire : 3**
Pronom : 2	**Pronom : 3**
Préposition : 4	**Préposition : 2**
Nom (termes nominaux):12	**Nom (termes nominaux):7**
Verbes (termes verbaux) : 5	**Verbes (termes verbaux) : 6**

- **Conjonction : 1**
- **Adverbe : 7**



- **Conjonction : 1**
- **Adverbe : 6**
- **Adjectif : 2**
- **Exclamation : 2**



⇒ ***Analyse de la rubrique « BBS » partagée entre people et sohu sur la répartition des termes en fonction de la catégorie lexicale***

Voici le résultat d'analyse sur la répartition des termes dans la rubrique « BBS » selon la catégorie lexicale. En termes du nombre des mots, à part le nombre de termes nominaux (*sohu* a presque deux fois plus que *people*) et les deux catégories supplémentaires (l'exclamation et l'adjectif) dans le *people*, la proportion des autres termes est relativement identique. En tant que deux catégories exclusives dans la rubrique « BBS » du *people*, l'exclamation et l'adjectif (surtout les deux mots d'exclamation : *啊／吧*) font passer des émotions humaines des internautes sur le sujet « smog épais en Chine ». En consultant la concordance des deux mots d'exclamation, le *啊／e* et le 吧／e jouent un même rôle dans le texte de cette rubrique, qui sont l'équivalent de l'adverbe « tellement » en français pour exprimer un sentiment fort ou un souhait sur un sujet d'intérêt.

⇒ ***Analyse des termes de la rubrique « BBS » partagés entre*** **people** ***et*** **sohu** ***au niveau de la signification***

Après la comparaison en matière de la répartition selon la catégorie lexicale, nous allons comparer les mots croisés et les mots différents entre *people* et *sohu* dans la rubrique « BBS » :

- ***Mots croisés de la rubrique « BBS » dans people et sohu***: 雾／n(brouillard)，霾／n(smog)，我／r(je).

À part les deux mots-clés, il n'y a que le pronom personnel « je » qui est partagé dans la rubrique « BBS » des deux sous-corpus. Effectivement, il est normal que qu'on puiss rencontrer le pronom « je » dans cette rubrique de forum, car les internautes sont là pour exprimer leur point de vue personnel. Ce mot « je » n'est donc pas un terme bien représentatif et significatif.

- ***Mots exclusifs dans la rubrique « BBS » du sohu*** : 空气/n (l'air).

Nous avons déjà vu dans les différents sous-corpus ce mot « l'air » soit comme un des termes les plus fréquents ou comme un des termes les plus co-occurrents des deux mots-clés. Alors qu'entre *sohu* et *people* dans leur rubrique similaire : « BBS », ce mot n'apparaît que dans celle du *sohu*. En fait, il semble qu'excepté les deux mots-clés et à part ce mot qui concerne le sujet de notre recherche, le reste est composé, soit des termes nécessaires dans les textes médiatiques, soit des termes plus ou moins hors sujet.

- ***Mots exclusifs dans la rubrique « BBS » du people*** : 中国 / n (la Chine)，治理 / v(Régulariser)，汽车 / n (l'automobile)，天气 / n (le temps)，人人 / n (tout le monde).

Alors que le mot *中国 / n (la Chine)* a été analysé dans la partie d'analyse sur les termes les plus co-occurrents des deux mots-clés du *people*, le verbe *治理 / v(Régulariser)* et le nom *天气 / n (le temps)* sont tous apparus et expliqués dans l'études de recherche du *gov*. Nous nous focalisons ainsi sur les deux noms restant : « l'automobile » et « tout le monde ». Le rôle du premier nom relève une des origines essentielles du brouillard de pollution en Chine : le gaz d'échappement vient de la circulation des véhicules automobiles ; Il s'agit quant au deuxième d'annoncer au grand public que tout le monde est censé assumer la responsabilité de protéger l'environnement atmosphérique, le gouvernement n'est pas le seul qui en est responsable. Chacun d'entre nous doit agir pour contribuer à la protection environnementale.

⇒ ***Petit résumé des résultats d'analyses sur la comparaison de la rubrique « BBS » partagée du* people *et du* sohu**

Deux types de comparaison ont été effectués dans la rubrique « BBS » partagée du *sohu* et du *people*. Par rapport au *people*, la comparaison au niveau de la catégorie lexicale met l'accent sur l'émotion humaine des internautes du *sohu* avec les mots

d'exclamation comme 啊 / e et le 吧 / e (équivalent de ''tellement'' pour exprimer une émotion forte) ; la comparaison à l'égard du sens des mots du *people* met en responsable tout le monde pour protéger l'environnement atmosphérique, y compris le gouvernement chinois. Sauf les mots-clés et le mot « l'air », nous n'avons pas trouvé de termes intéressants au sujet du smog épais dans « BBS » du *sohu*.

Le travail de contrastif continue sur la deuxième rubrique similaire : le « Blog ». Nous allons nous servir de la même méthode pour cette phase de travail.

4.8.2.2. Analyses et comparaison de la rubrique « Blog » partagée du *people* et du *sohu*

Voici le tableau contrastif des termes les plus fréquents dans la rubrique « Blog » du *sohu* et du *people*.

Les mots les plus fréquents dans rubrique Blog du *sohu*				**VS**	**Les mots les plus fréquents dans rubrique Blog du *people***		
Numéro	Token	Freq	Traduction		Token	Freq	Traduction
1	/w	826	espace		/w	662	espace
2	霾/n	73	le smog		的/u	147	auxiliaire
3	雾/n	63	le brouillard		霾/x	77	le smog
4	城市/n	25	la ville		雾/n	61	le brouillard
5	人/n	21	l'être humain		了/u	29	auxiliaire
6	空间/n	20	l'espace		一/m	28	un
7	投资/v	17	investir		是/vl	26	être
8	之/u	14	auxiliaire		口罩/n	21	le masque protecteur
9	人们/n	12	le gens		我/r	21	je
10	就是/r	12	auxiliaire		为/p	18	pour
11	危害/v	10	abîmer		中国/ns	16	la Chine
12	再/d	9	de plus		在/p	16	dans
13	乌鸦/n	8	le corbeau		天气/n	15	le temps
14	创造/v	8	créer		北京/ns	15	Pékin
15	再造/v	7	reproduire		篇/q	14	quantificateur
16	玛雅/ns	7	Maya		有/v	14	avoir
17	俺/r	7	pronom		人类/n	13	l'être humain
18	财富/n	7	la fortune		下/p	13	dessous
19	机会/n	7	l'opportunité		一个/q	12	un
20	自然/n	6	la nature		人/n	12	personne
21	稀缺/a	6	rare		中/p	12	dans...
22	随/p	6	préposition		空气/n	12	l'air

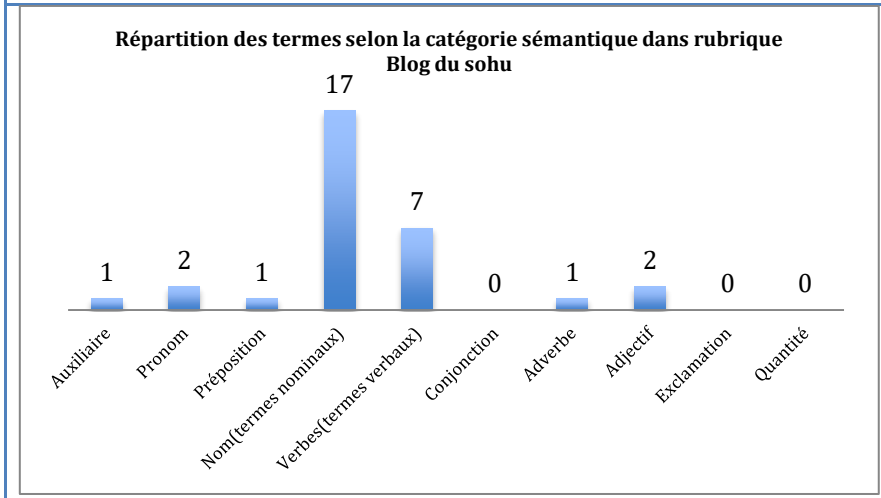
23	意思/n	6	le sens		装修/v	11	décorer
24	水资源/n	6	les ressources hydriques		发展/v	11	développer
25	城市化/v	5	urbaniser		pm/ws	11	la particule fine par mètre cube
26	房地产/n	5	l'immobilier		上/p	11	au dessus de...
27	当年/nt	5	cette époque-là		同/p	10	comme
28	藏/v	4	cacher		会/vu	10	pouvoir
29	危机/n	4	le danger		我们/r	10	nous
30	必然/a	4	inévitable		天/nt	10	le ciel
31	小镇/n	4	le bourg		对/p	10	envers
32	忘记/v	4	oublier		不/d	10	non
33	远离/v	4	éloigner de		治理/v	10	régulariser
34	生存/v	4	vivre		宠物/n	9	l'animal domestique

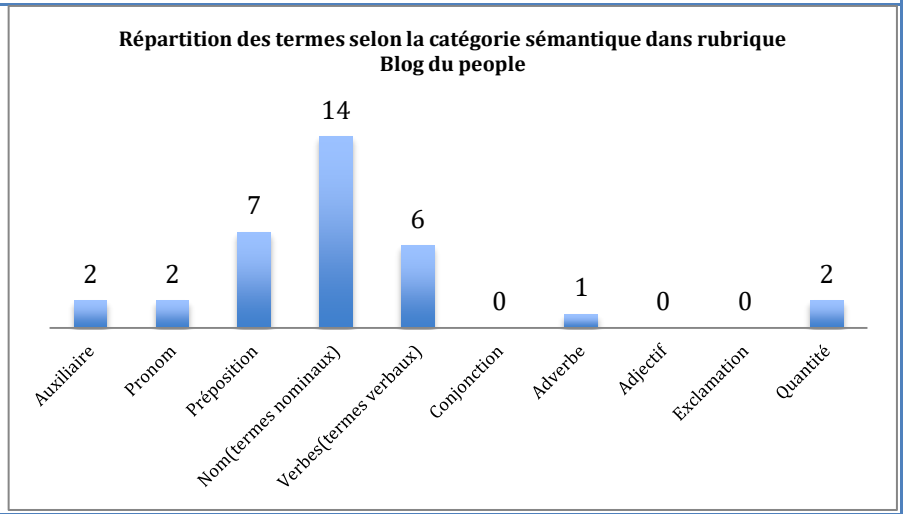
Tableau des mots les plus fréquents dans la rubrique « Blog » du sohu et ceux du people

La synthèse de la répartition des termes selon la catégorie lexicale est comme ci-dessous :

Les termes dans la rubrique « Blog » du *sohu* :	**Les termes dans la rubrique « Blog » du *people* :**
Auxiliaire : 1	Auxiliaire : 2
Pronom : 2	Pronom : 2
Préposition : 1	Préposition : 7
Nom (termes nominaux):17	Nom (termes nominaux): 14
Verbes (termes verbaux) : 7	Verbes (termes verbaux) : 6

Conjonction : 0	Conjonction : 0
Adverbe : 1	Adverbe : 1
Adjectif : 2	Adjectif : 0
Exclamation : 0	Exclamation : 0
Quantité : 0	Quantité : 2





⇒ ***Analyse de la rubrique « Blog » partagée du people et du sohu sur la répartition des termes en fonction de la catégorie lexicale***

Dans l'ensemble, l'état de la répartition est similaire aux termes dans la rubrique « Blog » partagée du *people* et du *sohu*, la proportion des termes nominaux de tous les deux sous-corpus atteint un sommet parmi les 9 catégories lexicales des termes indiqués, les termes verbaux sont placés au deuxième rang. Si nous rentrons dans les détails de chaque catégorie, les termes de conjonction et ceux d'exclamation ne se trouvent nulle part dans la rubrique de chacun des deux. Quant à la préposition, *people* en utilise beaucoup plus par rapport au *sohu*. Afin d'analyser et comparer les termes de manière plus concrète et plus transparente, nous allons étudier dans un second temps les termes au niveau du sens et du lien qu'ils entretiennent avec le sujet de recherche.

⇒ ***Analyse des termes dans la rubrique « Blog » partagée du people et du sohu au niveau de la signification***

- ***Mots croisés de la rubrique « Blog » dans people et sohu***: 雾／n(brouillard)，霾／n(smog).

On n'a que les deux mots-clés comme mots croisés dans la rubrique « Blog » des deux sous-corpus. Ce phénomène témoigne en quelque sorte que l'on parle de style différent du brouillard de pollution en Chine. Cette divergence du style s'avèrerait avec l'analyse suivante sur les termes divergents des deux sous-corpus.

- ***Mots exclusifs dans la rubrique « Blog » du sohu :*** 城市／n (la ville)，创造／v (créer)，危害／v(abîmer)，创造／v (inventer)，再造／v (réinventer)，玛雅／n (Maya)，俺／r(je)，财富／n(la fortune)，自然／n(la nature)，稀缺／a(rare)，乌鸦／n(corbeau)，水资源／n(les ressources hydriques)，城市化／n (la urbanisation)，房地产／n(l'immobilier)，藏／v(cacher)，危机／n(le danger)，小镇／n(le bourg)，远离／v (s'éloigner) et 生存／v(vivre).

Voici en haut les mots exclusifs dans la rubrique « Blog » du *sohu*. Nous

pouvons remarquer que 98% des termes sont "*les premiers vus*". On parle de l'aspect différent du « smog épais en Chine » dans leur « Blog » privé : le brouillard de pollution en Chine provient en partie de « *l'urbanisation* » et « *l'immobilier* », ces derniers ont pour objectif de faire « *la fortune* ». Alors que cette pollution de l'air le « *abîmer* » « *la nature* », « *la ville* » , elle met même « *le bourg* » en « *danger* », et rend de plus en plus « *rare* » « *les ressources hydriques* », « *je* » ressent « *le danger* » de « *vivre* » et la prédiction de « *Maya* » et voudrait « *s'éloigner* » et « *cacher* » quelque part. Et le terme « *corbeau* » est une métaphore pour décrire la modalité de la personne vivant sous le ciel pollué. Voilà le style plutôt littéraire et les angles divers que transmettent les textes de « Blog » du *sohu*.

- ***Mots exclusifs dans la rubrique « Blog » du people :*** 口罩 / n(le masque protecteur)，我 / r(je)，中国 / n(la Chine)，天气 / n(le temps)，北京 / n(Pékin)，空气 / n(l'air)，装修 / v(décorer)，发展 / v(développer)，pm / ws(la particule fine par mètre cube)，我们 / r(nous)，治理 / v(régulariser)，宠物 / n(l'animal domestique).

En ce qui concerne les mots exclusifs dans « Blog » du *people*, contrairement au sohu, le 98% des termes sont présents comme « déjà vus » dans l'analyse du sous-corpus *people,* tant en matière des termes les plus fréquents qu'au niveau des termes les plus co-occurrents. Il n'a que trois mots « *décorer* », « *pm* » et « l'*animal domestique* » figurant pour la première fois dans cette rubrique. En plus de « *pm* » qui est l'unité de mesure du smog épais, le « *décorer* » relié en fait à l'immobilier, alors que « *l'animal domestique* » s'emploie pour nous indiquer qu'il n'y pas que l'être humain qui est atteinte par cette nocivité du brouillard : les animaux ne sont pas non plus épargnés face à la pollution.

⇒ ***Petit résumé des résultats d'analyse de la comparaison des mots dans la rubrique « Blog » partagée du people et du sohu***

Il faut noter que même si tous les deux sous-corpus ont développé dans cette rubrique privée « Blog » de nouveaux aspects pour « décrire » le smog épais en Chine, la divergence liée au degré d'innovation est évidente. Ce dernier constat se base sur les occurrences des mots et le style dans chaque sous-corpus : les blogueurs du *sohu* emploient beaucoup de nouveaux termes et un style littéraire ou réthorique pour décrire leur sentiment personnel, par exemple on a le « Maya » pour exprimer l'inquiétude face à la dégradation du climat ; le « s'éloigner » exprime leur volonté de fuir l'urbanisation des villes modernes ; et le « corbeau » est un terme métaphorique qui décrit l'état frustré de la personne sous le ciel gris, etc. Au contraire du *sohu*, le style du *people* n'est pas tellement distinct même quand il s'agit d'une rubrique privée et personnelle, il paraît que le style est en cohérence avec celui du *gov* : plutôt objectif et conservateur, ce caractère se reflète par la répétition des mots et la rareté d'indication de nouveaux termes dans « BBS » du *people*.

4.8.2.3. Analyses et comparaison sur la rubrique « Business » partagée du *people* et du *sohu*

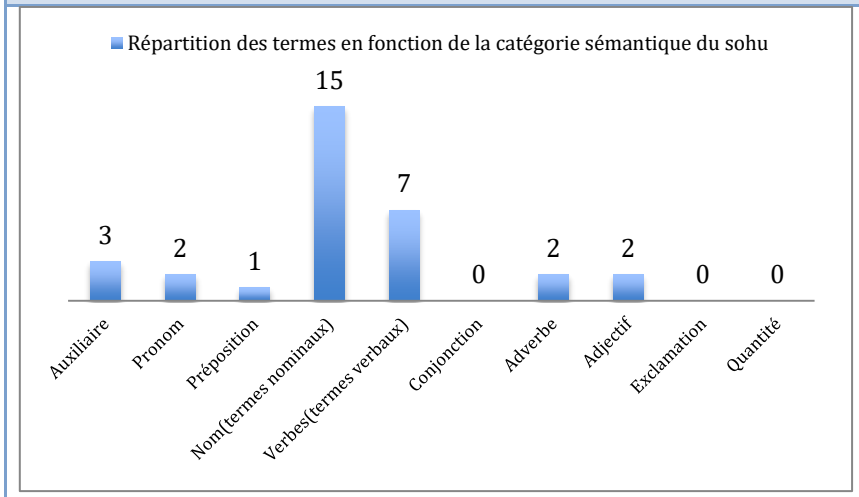
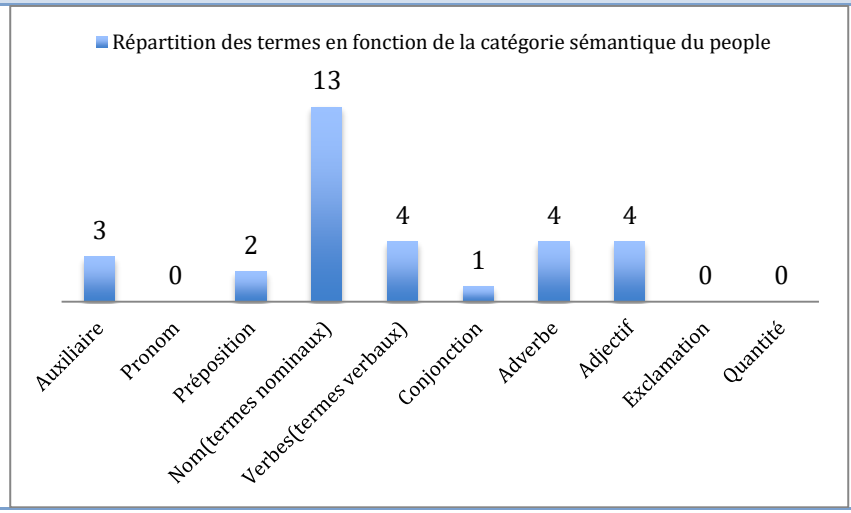
Nous arrivons à la troisième rubrique similaire : *Business*. Déployons les analyses qui s'inscrivent dans la continuïté de celles précédentes :

Les mots les plus fréquents dans Business du *sohu*				**VS**	**Les mots les plus fréquents dans Business du *people***		
Numéro	Token	Freq	Traduction		Token	Freq	Traduction
1	/w	485	espace		/w	521	espace
2	的/u	163	auxiliaire		的/u	133	auxiliaire
3	光/n	42	la lumière		霾/x	72	le smog
4	伏/v	39	se pencher sur		雾/n	63	le brouillard
5	是/u	29	être		了/u	28	auxiliaire
6	能源/n	28	l'énergie		在/p	28	dans
7	了/u	28	auxiliaire		是/vl	24	être
8	问题/n	22	le problème		口罩/n	21	le masque protecteur
9	我们/r	22	nous		污染/v	21	la pollution
10	在/p	22	dans		天气/n	19	le temps
11	西部/nd	18	l'ouest		也/d	18	aussi
12	有/v	18	avoir		空气/n	18	l'air
13	刘汉元/nh	17	nom personnel		多/a	17	nombreux
14	环保/j	16	la protection environnementale		日/nt	16	la journée
15	年/nt	16	l'année		记者/n	15	le journaliste
16	发展/v	15	développer		一/m	15	un
17	一/m	14	un		对/a	13	envers
18	霾/x	14	le smog		不/d	13	non
19	税/n	13	le taxe		和/c	13	avec
20	发电/v	13	produire de l'électricité		3/m	12	numéto
21	政策/n	13	la politique		来/vd	12	venir

22	也/d	13	aussi		天/nt	12	le jour
23	这/r	13	pronom		中/p	12	dans...
24	雾/n	13	le brouillard		等/d	12	etc.
25	解决/v	12	résoudre		体检/v	11	l'examen physique
26	我国/n	12	notre pays		月/nt	11	le mois
27	经济/n	12	l'économie		有/v	11	avoir
28	好/a	12	bon		市民/n	10	le citoyen
29	消耗/v	11	consumer		很/d	10	très
30	可以/vu	11	pouvoir		都/d	10	tous
31	要/vu	11	vouloir		要/vu	10	vouloir
32	不/d	11	non		大/a	10	grand
33	支出/n	10	la dépense		高/a	9	haut
34	刚性/a	10	nécessaire		环保/j	9	la protection environnementale

Tableau des mots les plus fréquents dans la rubrique « Business » du sohu et ceux du people

Les termes dans la rubrique « Business » du *sohu* :	**Les termes dans la rubrique « Business » du *people* :**
Auxiliaire : 3	**Auxiliaire** : 3
Pronom : 2	**Pronom** : 0
Préposition : 1	**Préposition** : 2
Nom (termes nominaux): 15	**Nom (termes nominaux)**: 13
Verbes (termes verbaux) : 7	**Verbes (termes verbaux)** : 4

Conjonction : 0	**Conjonction** : 1
Adverbe : 2	**Adverbe** : 4
Adjectif : 2	**Adjectif** : 4
Exclamation : 0	**Exclamation** : 0
Quantité : 0	**Quantité** : 0
	

⇒ ***Analyse de la rubrique « Business » partagée du people et du sohu sur la répartition des termes en fonction de la catégorie lexicale***

A part des deux catégories lexicales : les noms (les termes nominaux) et les verbes (les termes verbaux), qui sont marquées toutes les deux dans cette rubrique similaire du *sohu* et du *people*, les trois catégories du *sohu*, qui sont respectivement : les pronoms, , sont plus prononcées par rapport à celles du *people* : les prépositions, les adverbes et les adjectifs, car en général, les termes nominaux, verbaux et des pronoms ont souvent un sens significatif par rapport aux prépositions, adverbes et adjectifs, qui sont plutôt des mots de liaison qui articulent les phrases. Est-ce que cette remarque rendrait le contenu de la rubrique similaire « Business » du *sohu* plus intéressant que *people*? Nous allons vérifier cette hypothèse en combinant le résultat d'analyse des mots selon leurs sens. L'exclamation et la quantité sont toutes absentes dans les deux sous-corpus, qui rendent cette rubrique plus objective que la rubrique privée, telle que « BBS » ou « Blog ».

⇒ ***Analyse des termes de la rubrique « Business » partagée du people et du sohu au niveau de la signification***

- ***Mots croisés de la rubrique « Business » dans partagée du people et du sohu***: 雾／n(brouillard)，霾／n(smog)， 环保／j(la protection environnementale).

À part ces deux mots-clés qui sont croisés dans cette rubrique, il y a aussi un nouveau mot qui est apparu pour la première fois dans notre recherche : *环保／j(la protection environnementale)*[23], il est l'abréviation de 环境保护. En fait, l'occurrence du mot « la protection environnementale » relève un phénomène bien récent et qui montre la tendance croissante dans le domaine des finances. Avec l'expansion de l'air de pollution qui touche déjà la région du sud de la Chine, tel

[23] En chinois, *环保* est l'abréviation de 环境保护 dont deux mots en binôme peut former un groupe lexicale, autrement dit, le premier groupe composés des deux premiers mots signifie « l'environnement » (环境), et le deuxième groupe constitué des deux derniers mots signifie « la protection »（保护）. L'on n'en tire juste le premier pour reconstruire un autre mot dont la signification reste toujours le même.

que la province du Zhejiang (une province qui se trouve juste à côté de Shanghai). Les habitants cherchent toujours eux-mêmes des solutions pour diminuer la nocivité du smog épais, l'appareil purificateur de l'air est bien populaire et répond à des besoins urgents personnels. La popularité du purificateur d'air contribue à la croissance de l'économie des entreprises et à l'augmentation de ses actions. Ce type d'actions a été nommé comme 环保股(éco-action). Voilà la raison pour laquelle le mot « la protection environnementale » se manifeste dans la rubrique « Business ».

- ***Mots exclusifs dans la rubrique « Business » du sohu :*** 光／n (la lumière)，伏／n(volte)，能源／n(l'énergie)，问题／n(le problème)，我们／r(nous)，西部／nd(le nord)，发展／v(développer)，发电／v(produire de l'électricité)，政策／n(la politique)，解决／v(résoudre)，我国／n(notre pays)，经济／n(l'économie)，消耗／v(consommer).

13 mots appartenant respectivement à la catégorie nominale, verbale et pronominale, ont été choisis et considérés comme des mots exclusifs dans la rubrique « Business » du *sohu*. Certains mots se sont déjà affichés dans les parties précédentes du *sohu* comme *光／n(la lumière)*，*伏／n(volte)* et *问题／n(le problème)* ; du *people* comme *我们／r(nous)*. Les mots restant se manifestent tous pour la première fois. Deux mots de nature géographique : ‘‘notre pays’’ et ‘‘le nord’’. Nous avons rencontré le mot « la Chine» dans les analyses antérieures, alors que cette fois-ci dans la rubrique « Business » du *sohu*, on parle plus directement en utilisant la première personne pour décrire « la Chine »-« notre pays », auquel un ton plus direct et un sentiment plus attaché ont été accrochés. *西部／nd(le nord)* complète la dernière zone souffrant du smog épais qui est manquant du *gov.发电／v(produire de l'électricité)*, *经济／n(l'économie)* et *消耗／v(consommer)* composent un groupe qui relève l'origine principale qui produit le problème environnemental, surtout atmosphérique. *发展／v(développer)*, *能源／n(l'énergie)* et *解决／v(résoudre)* présente qu'afin de résoudre ce problème, on

cherche des solutions pour utiliser de nouvelles sources d'énergie pour remplacer celles traditionnelles et polluantes.

- ***Mots exclusifs dans la rubrique « Business » du people :*** 口罩 / n(le masque protecteur)，污染 / v(polluer)，天气 / n(le temps)，空气 / n(l'air)，市民 (le citoyen).

Comme ces mots sont tous apparus et analysés dans les parties antérieures, et que leurs rôles sont tous pareils, nous ne le répétons plus. Cependant, il faut noter d'une part que la composition des mots dans les deux groupes « termes exclusifs » sont tous intégrés dans les trois catégories lexicales : nominale, verbale et pronominale ; d'autre part, par rapport aux mots relevés du *sohu*, ces termes sont plutôt des termes neutres qui peuvent s'employer dans n'importe quelle rubrique. Ces deux points confirment notre hypothèse émise tout à l'heure : le contenu intégré dans la rubrique « Business » du *sohu* est plus intéressant que celui du *people*.

⇒ ***Petit résumé des résultats d'analyses de la comparaison sur la rubrique « Business » partagée du people et du sohu***

Les termes affichés dans la rubrique « Business » du *sohu* présentent deux grandes branches d'informations : il s'agit pour la première branche axant sur le smog épais lui-même, de son origine à sa nocivité jusqu'à la proposition des solutions, en ce qui concerne la deuxième branche, on relie le sujet avec la rubrique « Business », la notion de « éco-actions » dans le domaine de finance a été dérivé pour créer des contacts entre eux. Alors que les mots relevés du *people* ne sont pas revêtis de ces caractères, au contraire, ils sont répétitifs et neutres au niveau du sens, et peuvent être appliqués dans n'importe quelle rubrique. Cela rend le contenu de « Business » du *people* moins intéressant et moins représentant par rapport à celui du *sohu*.

4.8.2.4. Analyses et comparaison sur la rubrique « Military » partagée du *people* et du *sohu*

Les mots les plus fréquents dans Military du *sohu*				VS	Les mots les plus fréquents dans Military du *people*		
Numéro	Token	Freq	traduction		Token	Freq	Traduction
1	/w	242	espace		/w	402	espace
2	的/u	78	auxiliaire		的/u	129	auxiliaire
3	激光/n	50	le laser		心理/n	67	la mentalité
4	武器/n	43	l'arme		霾/x	35	le smog
5	在/p	27	dans		和/c	29	avec
6	导弹/n	17	le misile		雾/n	28	le brouillard
7	目标/n	13	le target		是/u	23	être
8	中/p	12	dans		在/p	23	dans
9	天气/n	11	le temps		新兵/n	16	les soldats fraîchement recrutés
10	等/d	11	etc.		问题/n	16	le problème
11	是/vl	11	être		对/p	16	correct
12	和/c	11	avec		起降/v	15	décoller et atterrir
13	美军/n	10	l'armée américaine		系统/n	15	le système
14	可以/vu	10	pouvoir		应急/v	14	faire face à une nécessité urgente
15	系统/n	9	le système		影响/v	14	influencer
16	一/m	9	un		能见度/n	13	la portée de vue
17	雾/n	9	le brouillard		中国/ns	13	la Chine
18	霾/x	9	le smog		飞行员/n	12	le pilote
19	影响/v	8	influencer		引导/v	12	guider
20	能/vu	8	pouvoir		新/a	12	nouveau
21	而/c	8	mais		着陆/v	11	atterrir

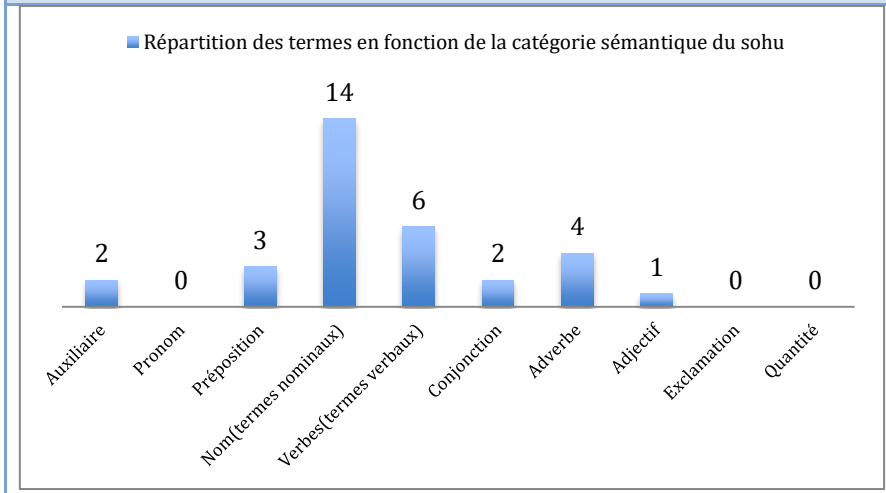
22	上/p	8	dessus		训练/v	11	entrainer
23	小/a	7	petit		飞机/n	11	l'avion
24	了/u	7	auxiliaire		一/m	11	un
25	就/d	7	adverbe		能力/n	10	la capacité
26	功率/n	6	puissance		政府/n	10	le gouvernement
27	军事/n	6	militaire		下/p	10	dessous
28	发射/v	6	tirer		地方/n	9	le pouvoir local
29	公司/n	6	l'entreprise		工作/n	9	le travail
30	会/vu	6	vouloir		等/d	9	etc.
31	将/d	6	adverbe		战士/n	8	le soldat
32	也/d	6	aussi		军事/n	8	la militaire
33	攻击/v	5	attaquer		不/d	8	non
34	光束/n	5	le rayon de soleil		也/d	8	aussi

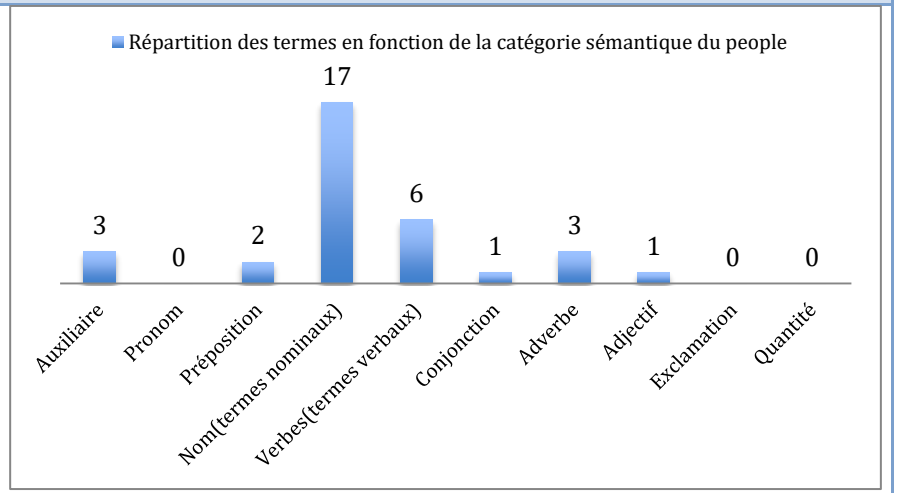
Tableau des mots les plus fréquents dans la rubrique « Military » du sohu et ceux du people

Nous arrivons dans la dernière rubrique d'intérêt « Military ».

Les termes dans la rubrique « Business » du *sohu* :	**Les termes dans la rubrique « Business » du *people* :**
Auxiliaire : 2	Auxiliaire : 3
Pronom : 0	Pronom : 0
Préposition : 3	Préposition : 2
Nom (termes nominaux): 14	Nom (termes nominaux): 17
Verbes (termes verbaux) : 6	Verbes (termes verbaux) : 6

Conjonction : 2	Conjonction : 1
Adverbe : 4	Adverbe : 3
Adjectif : 1	Adjectif : 1
Exclamation : 0	Exclamation : 0
Quantité : 0	Quantité : 0





⇒ ***Analyse de la rubrique « Military » partagée du people et du sohu sur la répartition des termes en fonction de la catégorie lexicale***

À travers les graphes ci-dessus, le nombre des termes nominaux du *people* dépasse pour la première fois celle du *sohu*, ce point met en relief la divergence la plus catégorique dans la rubrique « militaire » des deux sous-corpus. La deuxième caractéristique commune entre ces deux sous-corpus est qu'il n'y pas de pronom ni d'exclamation dans cette rubrique. À part ces deux points, la répartition des autres catégories reste presque la même.

⇒ ***Analyse des termes dans la rubrique « Military » partagée du people et du sohu au niveau de la signification***

- ***Mots croisés de la rubrique « Business » dans deux sous-corpus***:霾 / n(le smog), 雾 / n(le brouillard) et 影响 / v (influencer).

En plus des deux mots-clés, le verbe « influencer » a été marqué dans ce groupe. Nous avons parlé de ce mot qui relève l'influence du brouillard sur plusieurs aspects de la société, que ce soit au niveau de la nature qu'au niveau de la vie quotidienne des Chinois.

- ***Mots exclusifs dans la rubrique « Military » du sohu :*** 激光 / n (le laser)，武器 / n(l'arme)，导弹 / n(le missile)，目标 / n(le cible)，美军 / n(l'armée américaine)，军事 / n(le militaire)，发射 / v(tirer)，攻击 / v(attaquer).

Des termes spécifiques du domaine militaire ont été indiqués et font preuve que cette rubrique est bien différente des trois dernières. Il n'existe que deux catégories lexicales pour ces termes : les noms et les verbes. Avec les concordances de ces mots, on parle de tout type d'influence sur le lancement de l'arme : un faible niveau de visibilité causé par le smog épais peut affecter la précision en termes de lancement des armes, par exemple le missile. On a mentionné de vraies expériences

vécues par l'armée américaine sur ce type d'affecte du brouillard.

- ***Mots exclusifs dans la rubrique « Military » du people :*** 问题 / n(le problème)，起降 / v(décoller et atterir)，应急 / v(faire face à une urgence)，能见度 / n(la portée de vue)，中国 / ns(la Chine)，飞行员 / n(le pilote)，引导 / v(guider)，着陆 / v(atterir)，飞机 / n(l'avion)，政府 / n(le gouvernement)，地方 / n(le pouvoir local).

Deux différences principales se témoignent par les termes exclusifs des deux sous-corpus : premièrement, par rapport au sohu qui traite plusieurs aspects militaires affectés par l'air de pollution, people concentre plutôt l'influence du smog épais sur le décollage et l'atterrissage de l'avion ; deuxièmement, le texte de « military » du sohu ne liste que les influences du smog épais sur le militaire, alors que le contenu de « military » du *people* parle aussi des réactions et des attitudes du gouvernement chinois ainsi que locales : on prend des mesures et établit des politiques face à ces problèmes.

⇒ ***Petit résumé des résultats d'analyse de la comparaison sur la rubrique « Military » partagée du people et du sohu***

Les comparaisons des termes sur les quatre rubriques similaires du *sohu* et du *people* mettent en relief les caractéristiques spécifiques appartenant à ces deux sous-corpus. On parle tous d'un milieu spécial chez *sohu* comme chez *people* : le militaire. La diversité des aspects militaires (les armes, les missiles, l'expérience de l'armé américaine) amenée par les termes manifestés du *sohu,* en fait un site bien dynamique et bien varié au niveau du contenu, alors que pour *people*, même si nous avons rencontré des termes nouveaux dans cette rubrique spéciale, on parle toujours d'un seul aspect militaire : l'avion, dans ce sens-là, le *people* souligne son caractère formel par rapport au *sohu*. En plus, le *people* mentionne les mesures prises par le gouvernement chinois, cela synchronise le pas avec le *g*

4.8.3. Analyses et comparaison de la rubrique similaire « Lingdao » (dirigeants chinois) partagée du *gov* et du *people*

Après avoir mis en contraste les termes les plus fréquents des rubriques similaires, du *sohu* et du *people*, nous allons pratiquer la même méthode de comparaison sur les rubriques similaires entre le *gov* et le *people*.

Voici les rubriques similaires partagées dans le *gov* et le *people* avec l'état de répartition des deux mots-clés.

	Mots-clés	Lingdao (Dirigeants chinois)	Zhengce (Politics)	Lianghui (Deux sessions)
GOV	smog	-4 (sous-employé)	0 (normal)	0 (normal)
	brouillard	-4 (sous-employé)	2 (sur-employé)	0 (normal)
People	smog	-6 (sous-employé)	0 (normal)	0 (normal)
	bruillard	-4 (sous-employé)	0 (normal)	2 (sur-employé)

Tableau Les rubriques similaires partagées dans le gov et le people avec l'état de la répartition des deux mots-clés

Vu que les deux mots-clés sont tous bien fréquents dans la rubrique « lingdao » (les dirigeants chinois) partagée du *gov* et du *people*, nous focalisons l'analyse sur les termes les plus fréquents dans cette rubrique.

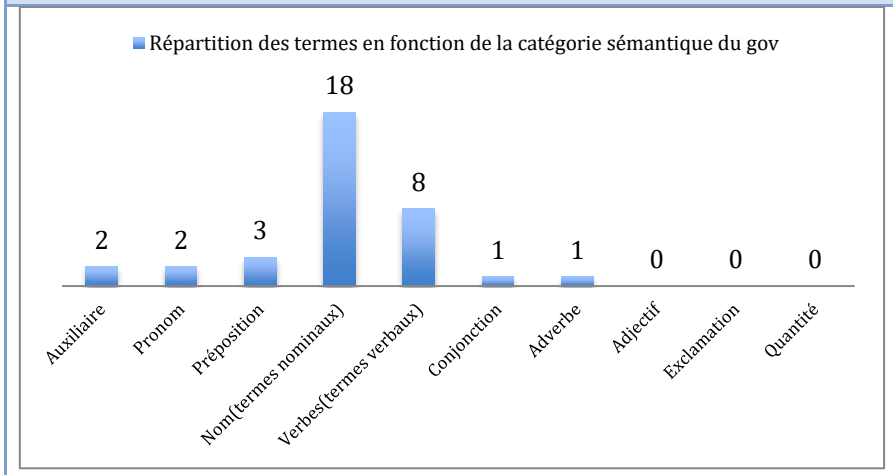
Les mots les plus fréquents dans Lingdao du *gov*				**VS**	**Les mots les plus fréquents dans Lingdao du *people***		
Numéro	token	totfreq	traduction		token	totfreq	traduction
1	的/u	829	auxiliaire		/w	10288	espace
2	和/c	259	avec		的/u	2188	auxiliaire
3	要/vu	123	vouloir		了/u	396	auxiliaire
4	在/p	187	dans		在/p	454	dans
5	工作/n	56	le travail		北京/ns	270	Pékin
6	治理/v	102	régulariser		霾/x	1040	le smog
7	国务院/ni	60	le Conseil d'État		雾/n	826	le brouillard
8	大气污染/n	97	la pollution atmosphérique		馒头/n	19	pain chinois cuit à la vapeur
9	会议/n	51	la conférence		领导/n	39	les dirigeants
10	日/nt	253	la jounée		也/d	183	aussi
11	霾/x	339	smog		办/v	18	s'occuper
12	李克强/nh	44	le nom personnel		副/a	44	vice
13	月/nt	126	le mois		是/u	387	être
14	等/d	229	etc.		空气/n	245	l'air
15	防治/v	67	prévenir		日/nt	361	la journée
16	雾/n	271	le brouillard		大/a	203	grand
17	了/u	109	marqueur du passé composé		有/v	394	avoir
18	马凯/nh	12	le nom personnel		月/nt	193	le mois
19	以/p	45	prépisition		一/m	241	un
20	政府/n	55	le gouvernement		任/v	20	assumer la charge de
21	措施/n	53	les mesures		两会/j	25	les Deux Sessions

22	春运/n	13	la période de pic de la fête du printemps		天/nt	170	le jour
23	运输/v	13	le transport		10/m	134	numéro
24	安全/n	22	la sécurité		网友/n	40	l'internaut
25	相关/v	22	relativiser		南京/ns	41	Nankin
26	总理/n	24	le premier ministre		过/vd	53	passer
27	加大/v	26	élargir		不/d	179	non
28	落实/v	28	exécuter		被/p	166	préposition
29	他/r	43	il		就/d	129	adverbe
30	重点/n	38	le majeur		这/r	122	pronom
31	发展/v	57	développer		我/r	119	je
32	环境/n	62	l'environnement		省长/n	11	gouverneur de la province
33	能源/n	62	l'énergie		推荐/v	47	recommander
34	对/p	82	envers		咋/r	41	quoi

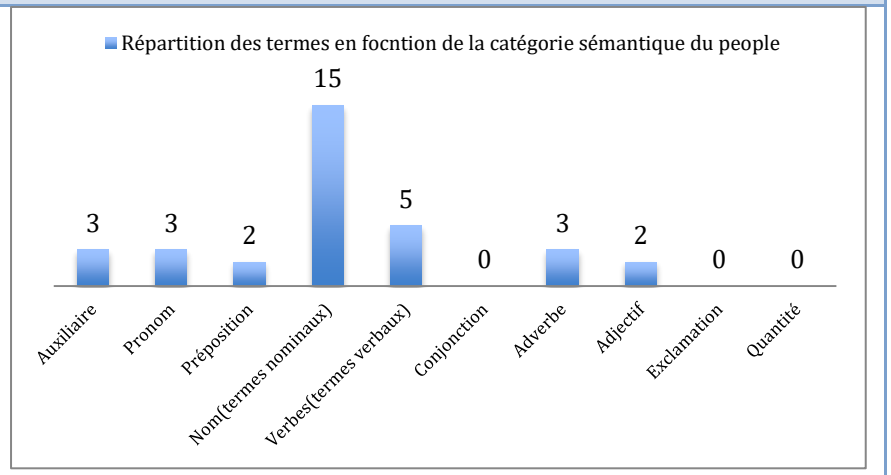
Tableau des mots les plus fréquents dans la rubrique « Lingdao » du gov et ceux du people

Les termes dans la rubrique « Business » du *sohu* :	Les termes dans la rubrique « Business » du *people* :
Auxiliaire : 2	Auxiliaire : 3
Pronom : 2	Pronom : 3
Préposition : 3	Préposition : 2
Nom (termes nominaux): 18	Nom (termes nominaux): 15
Verbes (termes verbaux) : 8	Verbes (termes verbaux) : 5

- **Conjonction : 1**
- **Adverbe : 1**
- **Adjectif : 0**
- **Exclamation : 0**
- **Quantificateur : 0**



- Conjonction : 0
- Adverbe : 3
- Adjectif : 2
- Exclamation : 0
- Quantificateur : 0



⇒ ***Analyse de la rubrique « Lingdao » partagée du people et du gov (les dirigeants chinois) sur la répartition des termes en fonction du trait lexicale***

Il y a peu de différence de répartition au niveau lexicale. Les termes nominaux représentent presque la moitié dans l'ensemble des mots. *Gov* possède trois mots en plus que people en matière des noms et des verbes.

⇒ ***Analyse des termes dans la rubrique« Lingdao » partagée du people et du gov (les dirigeants chinois) au niveau de la signification***

- ***Mots croisés de la rubrique « Lingdao » (les dirigeants chinois) dans people et gov***: 雾 / n(le brouillard)，霾 / n(le smog).

Il n'y a que les deux mots-clés qui sont croisés dans ces deux sous-corpus même si leur nature est tous « institutionnelle ». Est-ce que nous pouvions ainsi estimer que peu de similarité est partagée sur le contenu des textes en « lingdao » entre ces deux sous-corpus ? Avec cette question, nous allons effectuer les analyses de manière plus profondes.

- ***Mots exclusifs dans la rubrique « Lingdao » (les dirigeants chinois) du gov :*** 治理 / v(régulariser)，国务院 / ni(Conseil d'État)，大气污染 / n(la pollution atmosphérique)，防治 / v(prévenir)，政府 / n(le gouvernement chinois)，措施 / n(les mesures)，运输 / v(le transport)，安全 / n(la sécurité)，加大 / v(renforcer)，落实 / v(exécuter)，重点 / n(le majeur)，发展 / v(developper)，环境 / n(l'environnement)，能源 / n(l'énergie).

À travers ces mots montrés dans la rubrique « lingdao » du *gov*, nous avons déjà rencontré *治理 / v(régulariser)，大气污染 / n(la pollution atmosphérique)，防治 / v(prévenir)，加大 / v(renforcer)* dans la partie d'analyse du *gov*. En observant les mots spécifiques dans cette rubrique du *gov*, trois groups principaux peuvent être extraits :

- organisme gouvernemental : 国务院 / ni(Conseil d'État), 政府 / n(le gouvernement chinois) ;
- les réactions et les mesures prises : 措施 / n(les mesures), 防治 / v(prévenir), 治理 / v(régulariser), 加大 / v(renforcer), 落实 / v(exécuter), 发展 / v(développer), 能源 / n(l'énergie)
- objet de contrôle : 大气污染 / n(la pollution atmosphérique) et 环境 / n(l'environnement).

En tant que site officiel du gouvernement de la Chine, le GOV aborde la question du smog épais à l'échelle de l'État, ce qui se traduit surtout par le reportage des actions et des meures prises par le gouvernement chinois. Le résultat d'analyse de cette rubrique est en cohérence avec celui du tout le sous-corpus *gov*.

- ***Mots exclusifs dans la rubrique « Lingdao » (les dirigeants chinois) du people :*** 北京 / ns(Pékin)，领导 / n(les dirigeants chinois)，空气 / n(l'air)，两会 / n(Deux Sessions)，网友 / n(l'internaute)，南京 / ns(Nankin)，我/j(je).

Par rapport au résultat représenté du *gov*, deux mots attirent notre attention : un pronom personnel « je » et un nom « internaute ». Normalement, le « je » s'emploie rarement dans les discours institutionnels par exemple dans la rubrique « lingdao », c'est souvent le « on » ou le « nous » qui prévaut. Dans l'intention de trouver la réponse pour expliquer cette occurrence irrégulière, nous avons consulté la concordance du « je » du texte. En effet, on organise le contenu du texte bien différemment que d'habitude. Vu qu'il était la période des « Deux Sessions », le texte est édité sous forme de Q&A, composé des questions posées par les internautes, et des réponses proposées par les représentants populaires ou les dirigeants chinois sur les sujets de focus de la société chinoise, dont le « smog épais en Chine » fait partie. Nous pouvons ainsi en déduire que dans cette rubrique « lingdao » partagée dans *gov* et *people*, en comparaison avec « smog épais en Chine », se concentre sur tous les aspects du gouvernement chinois, alors que le deuxième fait participer au peuple chinois à la bataille contre le smog épai, et passer

les messages des grandes masses aux dirigeants chinois. Voici la différence principale des deux sous-corpus dans cette rubrique « lingdao ».

⇒ ***Petit résumé des résultats d'analyse dans cette partie de comparaison***

Avec les termes répétitifs dans la rubrique « lingdao » du *gov*, cette rubrique met en avant dans différents aspects le rôle principal du gouvernement chinois : participant dans la bataille contre le smog épais en Chine, on met en relief cette participation active à travers les verbes d'actions ainsi que l'évocation du nom des institutions gouvernementales de manière répétitive. Quant au *people*, la diversité du style d'organisation du texte montre la souplesse de ce sous-corpus institutionnel, le format Q&A s'emploie par le *people* prouve suffisamment qu'il fait participer le grand public contre le problème atmosphérique concernant tout le monde en Chine, le gouvernement a la volonté d'écouter la grande masse et veut aussi transférer au *people* les précisions et les messages du milieu gouvernemental. L'interaction mutuelle entre les Chinois et le gouvernement s'est bien manifestée dans « lingdao » du *people*.

4.8.4. Analyses et comparaison de la rubrique « News » partagée du *gov*, du *people* et du *sohu*

Après avoir fini les analyses et les comparaisons sur les rubriques similaires partagées du groupe ''*people et sohu*'' et du ''*people et gov*'', il reste la dernière comparaison sur la rubrique « News » partagée par les trois sous-corpus, nous allons regarder ensemble les résultats obtenus à l'issue d'un travail analytique et contrastif.

Les mots les plus fréquents dans « News » du Gov				**VS**	**Les mots les plus fréquents dans « News » du People**			**VS**	**Les mots les plus fréquents dans « News » du Sohu**		
Numéro	Token	Freq	Traduction		Token	Freq	Traduction		Token	Freq	Traduction
1	霾/x	99	le smog		/w	50	espace		/w	575	espace
2	日/nt	78	la journée		的/u	11	auxiliaire		的/u	135	auxiliaire
3	雾/n	69	le brouillard		霾/x	62	le smog		霾/x	44	le smog
4	的/u	60	auxiliaire		雾/n	49	le brouillard		空气/n	36	l'air
5	有/v	58	avoir		在/p	22	dans		污染/v	35	la pollution
6	地/u	56	adverbe		天气/n	20	le temps		在/p	35	dans
7	地区/n	51	la région		月/nt	17	le mois		2/m	33	le numéro
8	等/v	44	attendre		年/nt	16	l'année		5/m	33	le numéro
9	中/nd	43	le centre		中/nd	15	centre		雾/n	33	le brouillard
10	东部/nd	37	l'est		和/c	15	avec		pm/ws	31	particule mètre cube
11	月/nt	35	le mois		了/u	14	auxiliaire		北京/ns	30	Pékin
12	时/nt	31	l'heure		一/m	14	un		大雾/n	21	le brouillard
13	摄/v	30	photographier		大/a	13	grand		人/n	21	l'être humain
14	新华社/ni	30	la Presse XINHUA		等/d	13	etc.		和/c	21	avec
15	北部/nd	29	le nord		中央/n	12	l'autorité centrale		了/u	20	auxiliaire
16	在/p	28	dans		从/p	12	depuis		中/p	20	dans...
17	南部/nd	27	le sud		与/c	12	avec		是/vl	18	être
18	和/c	24	avec		中国/ns	12	la Chine		影响/v	16	influencer
19	笼罩/v	23	couvrir		污染/v	12	polluer		多/a	16	nombreux
20	级/n	23	le degré		是/vl	11	être		天气/n	16	le temps

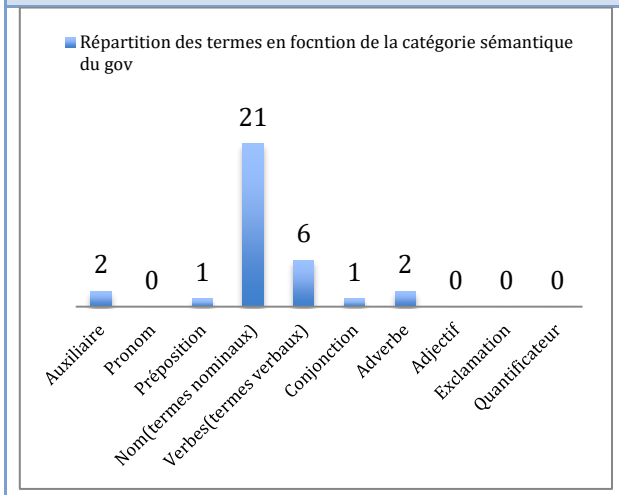
21	08/m	22	le numéro		对/p	10	envers		天/nt	15	la journée
22	10/m	21	le numéro		这/r	10	pronom		等/d	15	etc.
23	西部/nd	19	l'ouest		副/q	9	vice		颗粒/n	14	le particule
24	或/d	19	ou		次/q	9	fois		有/v	14	avoir
25	部分/n	18	la partie		习近平/nh	9	le nom personnel		高速/a	13	la grande vitesse
26	1/m	18	le numéro		要/vu	9	vouloir		我们/r	13	nous
27	到/v	18	arriver		地区/n	9	la région		就/d	13	adverbe
28	11/m	17	le numéro		2/m	9	numéro		一/m	13	un
29	5/m	17	le numéro		严重/a	9	grave		质量/n	12	la qualité
30	发/v	16	distribuer		而/c	8	cependant		都/d	12	tous
31	度/n	16	le degré		图/n	8	le plan		烟花/n	11	le feu d'artifice
32	华北/nl	16	le nord		将/d	8	adverbe		伦敦/ns	11	Londre
33	至/v	16	arriver		被/p	8	préposition		物/n	11	l'objet
34	将/d	16	adverbe		影响/v	8	influencer		但/c	11	mais

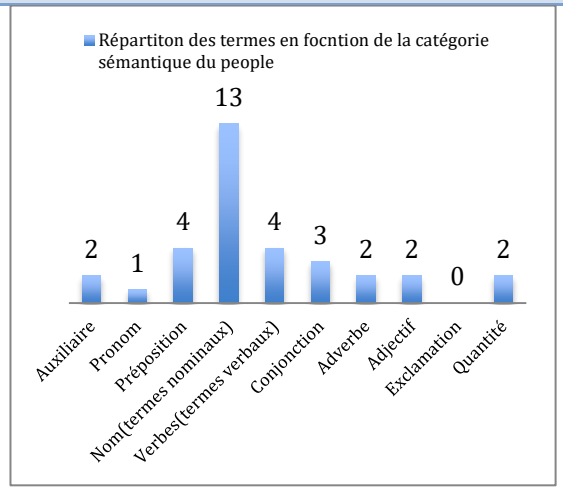
Tableau des mots les plus fréquents dans la rubrique « News » du sohu et ceux du people et du gov

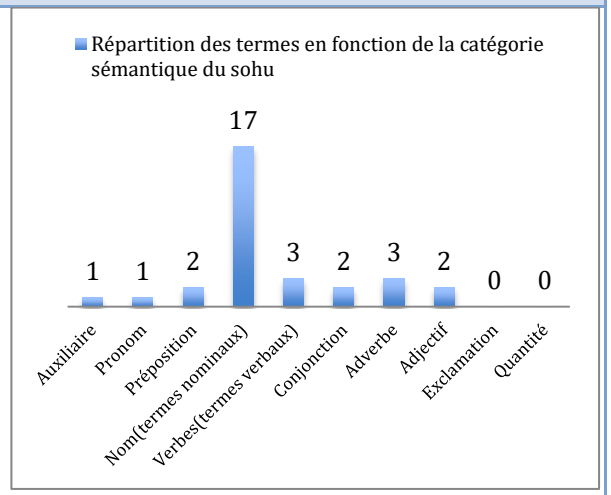
Mettons nous en ordre les termes selon leur catégorie lexicale

Les termes dans la rubrique « Business » du *gov* :	**Les termes dans la rubrique « Business » du *people* :**	**Les termes dans la rubrique « Business » du *sohu* :**
Auxiliaire : 2	Auxiliaire : 2	Auxiliaire : 1
Pronom : 0	Pronom : 1	Pronom : 1

Préposition : 1	Préposition : 4	Préposition : 2
Nom (termes nominaux): 21	Nom (termes nominaux): 13	Nom (termes nominaux): 17
Verbes (termes verbaux) : 6	Verbes (termes verbaux) : 4	Verbes (termes verbaux) : 3
Conjonction : 1	Conjonction : 3	Conjonction : 2
Adverbe : 2	Adverbe : 2	Adverbe : 3
Adjectif : 0	Adjectif : 2	Adjectif : 2
Exclamation : 0	Exclamation : 0	Exclamation : 0
Quantificateur : 0	Quantificateur : 2	Quantificateur : 0







⇒ ***Analyse de la rubrique « News » partagée dans les trois sous-corpus sur la répartition des termes en fonction de la catégorie lexicale***

Dans l'ensemble, la répartition des mots dans les trois sous-corpus n'est pas bien identique, même si *gov* et *people* sont tous de nature institutionnelle. La différence la plus marquée est celle qu'on a observée entre les termes nominaux. Bien évidemment, *gov* possède plus de termes nominaux parmi les trois, il devance sur ce point respectivement *sohu* puis *people*. Faisons la comparaison en binôme. D'abord dans le groupe institutionnel entre *gov* et *people*. En plus de la différence que nous avons mentionnée sur les termes nominaux, l'« adjectif », le « quantificateur » et le « pronom » sont absents dans *gov*, mais *people* en contient respectivement 2, 2, 1 pour chacun ; en outre, *gov* a juste une seule « conjonction », alors que *people* en a trois. Nous pouvons dire qu'en plus de l'uniformité sur la nature, la composition des catégories lexicales est relativement homogène du *gov* que celle du *people*. En suite, mettons en contraste *people* et *sohu*. Comparé à la divergence évidente sur la répartition entre *gov* et *people*, à part l'absence du « quantificateur » chez *sohu*, le dernier et *people* se ressemblent beaucoup dans cette matière. En tant que « site médiatique », ces deux sites emploient plus de conjonctions, d'adverbes et de prépositions pour organiser leur texte par rapport au *gov*.

⇒ ***Analyse des termes dans la rubrique« News » partagée dans les trois sous-corpus au niveau de la signification***

- ***Mots croisés de la rubrique « News » dans trois sous-corpus***: 雾 / n(le brouillard)，霾 / n(le smog).

- ***Mots croisés de la rubrique « News » du gov et du people :*** 地区 / n(la région).

  Si le mot « la région » est affiché aussi dans *people*, ce quotidient, contraiement au *gov*, n'a pas indiqué les régions concrètes touchées par le smog épais en Chine.

- ***Mots croisés de la rubrique « News » du sohu et du people :*** 天气/n(le

temps),污染/v(polluer), 影响 / v (influencer).

Parmi les trois mots croisés dans *sohu* et *people*, il y a deux tiers de verbes, qui ont un lien très étroit avec le smog épais, pour décrire la nocivité de cet air de pollution et pour ainsi sensibiliser la société chinoise.

- ***Mots exclusifs dans la rubrique « News » du gov*** :中 / nd(le centre)，东部 / nd (l'est)，北部 / nd (le nord)，南部 / nd (le sud)，西部 / nd(l'ouest)，华北 / nd（la région du nord）,笼罩(couvrir).

Ces mots apparus dans la partie d'analyse sur *gov* ne nous paraissent pas étranges. Ils nous appennent l'ampleur répandue du smog épais sur le territoire de la Chine.

- ***Mots exclusifs dans la rubrique « News » du people*** : 中央 / n(l'autorité centrale), 中国 / ns(la Chine), 习近平 / nh(nom personnel), 严重 / a(grave).

Par rapport aux mots indiqués dans « News » du *gov*, nous avons 4/5 mots nouveaux dans *people*. L'« autorité centrale » et le nom du président de la Chine, d'une part, révèlent la nature institutionnelle du *people*, et d'autre part, nous dialoguent l'attention accordée par le gouvernement chinois à ce problème qu'il considère comme très « préoccupant ».

- ***Mots exclusifs dans la rubrique « News » du sohu***: 北京 / ns(Pékin)，空气 / n(l'air)，大雾 / n(le brume)，颗粒 / n(la particule fine)，pm / ws(l'unité de la particule fine par mètre cube)，高速 / a(en grande vitesse)，我们 / r(nous)，烟花 / n(le feu d'artifice)，伦敦 / ns(Londres).

Ces termes apparaissent de nouveau dans cette rubrique. Cela nous amène à croire que ces mots sont tellement fréquents dans le sous-corpus sohu qu'ils peuvent être considérés comme les termes représentants du *sohu*.

⇒ **Petit résumé sur les résultats d'analyse de la comparaison sur la rubrique « News » partagée des trois sous-corpus**

Après avoir réuni les trois sous-corpus pour comparer les termes spécifiques dans la rubrique similaire « News », nous remarquons que la homogénéité a été enfilée et pénétrée dans la rubrique « News » de chaque sous-corpus, ils gardent non seulement leur style d'organisation du texte mais surtout leur point essentiel sur le smog épais en utilisant les termes similaires ou bien répétitifs. Plus explicitement parler, en ce qui concerne les catégories lexicales, le *gov* est plus homogène que les deux autres, l'adjectif, le pronom et le quantificateur y sont tous absents. Alors *people* et *sohu* possèdent plus de catégories, dont la répartition des mots est relativement équilibré entre eux. Quant au sens des termes indiqués, *gov* et *people* sont tous de côté gouvernemental, mais *sohu* est de côté du grand public.

4.8.5. Résumé des résultats d'analyse des rubriques similaires sur les trois sous-corpus

Ayant pour objectif d'intégrer les résultats d'analyse en externe sur les trois sous-corpus ainsi qu'en interne sur les rubriques similaires entre les sous-corpus, nous allons ranger les résultats dans deux tableaux : l'un pour la comparaison en externe parmi les trois sous-corpus, l'autre pour le contraste en interne des rubriques similaires entre les sous-corpus.

4.8.5.1. Résumé des résultats d'analyses en externe parmi les sous-corpus

Rangeons les termes les plus fréquents et les plus co-occurrents avec les deux mots-clés des trois sous-corpus dans le tableau ci-dessous, nous n'avons choisi que les termes qui ont un lien très étroit avec les deux mots-clés et deux catégories lexicales représentatives (les verbaux et les nominaux) :

Nom du sous-corpus	Catégorie lexicale	Les plus fréquents	Co-occurrents du smog	Co-occurrents du brouillard

GOV	Nominaux	雾(le brouillard) 霾(le smog) 地区(la région) 东部(l'est) 天气(le temps) 北部(le nord) 南部(le sud) 大气污染(la pollution atmosphérique)	雾(le brouillard) 天气(le temps) 北京(Pékin) 地区(la région) 大气污染(la pollution atmosphérique)	霾(le smog) 地区(la région) 天气(le temps) 北部(le nord) 内蒙古(Inner Mongolier) 大气污染(la pollution atmosphérique) 北京(Pékin) 我们(nous) 能见度(la portée de vue) 范围(la zone)
	Verbaux	治理(régulariser) 污染(polluer) 防治(prévenir)	治理(régulariser) 笼罩(couvrir) 预警(alerter) 加强(renforcer) 监测(détecter) 出现(apparître) 预报(prévenir) 发布(publier) 形成(former) 治(résoudre) 影响(influencer)	治理(régulariser) 研究(étudier) 出现(apparaître) 加强(renforcer) 监测(détecter)
SOHU	Nominaux	雾(le brouillard) 霾(le smog)	雾(le brouillard) 天气(le temps)	霾(le smog) 天气(le temps)

		空气(l'air) 我(je) 天气(le temps) 北京(Pékin)	北京(Pékin) 中国(la Chine) 灰(le cendre) 原因(la raison) 罚单(l'amande) 口罩(le masque protecteur) 问题(le problème) 城市(la ville) 伏	北京(Pékin) 中国(la Chine) 原因(la raison) 罚单(l'amande) 问题(le problème) 城市(la ville)
	Verbaux	污染(polluer) 要(vouloir) 可以(pouvoir) 能(pouvoir) 会(falloir)	治理(régulariser) 危害(abîmer) 应对(affronter) 防雾(protéger contre le brouillard)	治理(régulariser) 危害(abîmer) 应对(affronter)
PEOPLE	Nominaux	雾(le brouillard) 霾(le smog) 北京(Pékin) 空气(l'air) 天气(le temps) 中国(la Chine) 地区(la région)	雾(le brouillard) 天气(le temps) 北京(Pékin) 中国(la Chine) 口罩(le masque protecteur)	霾(le smog) 天气(letemps) 北京(Pékin) 中国(la Chine) 灯(le feu de circulation) 地区(la région) 问题(le problème 专家(l'expert) 市民(le citoyen)

				口罩(le masque protecteur)
	Verbaux	污染(polluer)	治理(régulariser) 出现(apparaître) 治(résoudre) 预警(alerter) 影响(influencer) 笼罩(couvrir)	治理(régulariser) 影响(influencer) 笼罩(couvrir) 遭遇(subir) 持续(durer)

Comme il y a trois types de mots (les plus fréquents et les plus co-occurrents avec les deux mots-clés) qui sont nombreux mais séparés dans le tableau présent, nous allons les mélanger en éliminant la répétition des termes, puis les synthétiser dans un même tableau selon leur catégorie lexicale et leur sous-corpus.

Dans le cadre de comparaison entre la divergence et la convergence des termes du chaque sous-corpus, voici quelques remmarques ：

- Les termes spécifiques dans chaque sous-corpus vont être soulignés en rouge ;
- Les termes croisés parmi les trois sous-corpus seront en noir ;
- Les termes partagés du *gov* et du *people* seront colorés en bleu ;
- Les partagés entre *sohu* et *people* seront soulignés en vert.

Catégorie lexicale	**GOV**	**PEOPLE**	**SOHU**
Nominaux	地区(la région) 天气(le temps) 北京(Pékin) 东部(l'est) 北部(le nord)	天气(le temps) 北京(Pékin) 口罩(le masque protecteur) 问题(le problème)	天气(le temps) 北京(Pékin) 我(je) 灰(le poudre) 原因(la raison)

	南部(le sud) 大气污染(la pollution atmosphérique) 内蒙古(Inner Mongolier) 能见度(la portée de vue) 范围(l'étendue)	空气(l'air) 地区(la région) 灯(le feu de circulation) 专家(l'expert) 市民(le citoyen) 中国(la Chine)	罚单(l'amande) 城市(la ville) 伏(le volt) 口罩(le masque protecteur) 问题(le problème) 空气(l'air)
Verbaux	治理(régulariser) 污染(polluer) 笼罩(couvrir) 预警(alerter) 影响(influencer) 出现(apparaître) 治(résoudre) 加强(renforcer) 监测(détecter) 预报(prévoir) 发布(publier) 形成(reformer) 防治(prévenir)	治理(régulariser) 污染(polluer) 笼罩(couvrir) 预警(alerter) 影响(influencer) 出现(apparaître) 治(résoudre) 遭遇(subir) 持续(durer)	治理(régulariser) 污染(polluer) 危害(abimer) 应对(affronter) 防雾(se protéger contre le brouillard de pollution) 要(vouloir) 可以(pouvoir) 能(falloir) 会(être capable de)
Total	23 mots	19 mots	20

⇒ Résumé au niveau de la répartition des termes selon la catégorie lexicale

- **<u>La répartition des termes du groupe « termes spécifiques des trois sous-corpus » (en rouge):</u>**

Pour le *gov*, il y a 12 mots divisés en partie égale des nominaux et des verbaux ; le *people* possède 6 mots exclusifs dont 4 nominaux, 2 verbaux ; 6 nominaux et 7

verbaux sont des mots exclusifs du *sohu*.

Le nombre des termes spécifiques de chaque sous-corpus	Catégorie lexicale	GOV	PEOPLE	SOHU
	Nominaux	6/23	4/19	6/20
	Verbaux	6/23	2/19	7/20
Total		12/23	6/19	13/20

- **La répartition des termes du groupe « termes croisés dans les trois sous-corpus » (en noir):**

Il y a en total 4 mots croisés dans les trois sous-corpus : 2 mots verbaux 2 mots nominaux.

- **La répartition des termes partagés du *gov* et du *people* (en bleu) :**

Le nombre des termes partagés du *gov* et du *people*	Catégorie lexicale	GOV	PEOPLE
	Nominaux	1	1
	Verbaux	5	5
Total		6	6

- **La répartition des termes partagés du *sohu* et du *people* (en vert):**

Le nombre des termes partagés du *gov* et du *people*	Catégorie lexicale	SOHU	PEOPLE
	Nominaux	3	3
	Verbaux	0	0
Total		3	3

- **Il n'existe pas de termes partagés du *sohu* et du *gov*.**

Comme nous le voyons, dans le groupe des termes exclusifs des trois sous-corpus,

en termes de quantité, *sohu* est placé au premier rang en devançant juste d'un mot le gov, et aucun mot n'est partagé entre les deux sites. Ce qui permet de mieux distinguer ces deux sous-corpus : purement institutionnel pour l'un et purement informel médiatique pour l'autre. Chaque sous-corpus contient des termes qui lui sont propres : *gov* parle des régions touchées par le brouillard et concernées par les mesures prises par le gouvernement chinois ; *sohu* discute des origines, de la nocivité liée au smog épais ainsi que des souhaits prononcés du grand public face à la pollution. Il n'y a pas de chevauchement du sens entre ces deux sous-corpus. Quant au *people*, il est beaucoup moins caractéristique ou moins original par rapport aux *gov* et *sohu* : d'abord, au niveau de la quantité des termes exclusifs, *people*(6) est seulement environ la moitié des deux autres(12 et 13), ensuite, *people* possède toujours des termes croisés soit avec *sohu* soit avec *gov*. Si nous nous appuyons sur la proportion des termes exclusifs, la part du *people* occupe sulement moins d'un tiers dans toute sa collection, qui est loin de celle du *gov*(plus de la moitié) et du *sohu*(plus de la moitié), alors que la proportion des termes croisés du people en occupe plus deux tiers ; troisièmement, en matière du sens des mots : *people* parle aussi de l'origine et de la nocivité du brouillard, d'une part, et des mesures prises par le gouvernement chinois, d'autre part. La seule divergence du *people* se marque par l'objectivité et la scientificité transmises du mot « l'expert ».

4.8.5.2. Résumé des résultats d'analyse en interne sur les rubriques similaires partagées entre les sous-corpus

Deux types de comparaison seront effectués dans les analyses suivantes :

- Comparaison en binôme des rubriques similaires entre *people* et *sohu* ou entre *people* et *gov*;
- Comparaison en trinôme de la rubrique similaire parmi *people*, *sohu* et *gov*.

⇒ Rubriques similaires entre *people* et *sohu* (les termes croisés sont en vert, les termes différents en rouge)

Nom de la	**Nom du sous-**	**Catégorie lexicale**	**Les plus fréquents**

rubrique	**corpus**		
BBS	People	N (7/8)	雾，霾，我，中国，汽车，天气，人人
		V (1/8)	治理
		Total	8
	Sohu	N (4/4)	雾，霾，我，空气
		V (0)	0
		Total	4
Blog	People	N (12/15)	雾，霾，口罩，我，中国，天气，北京，人类， 空气，pm，我们，宠物
		V (3/15)	装修，发展，治理
		Total	15
	Sohu	N (15/19)	雾，霾， 城市，危害，乌鸦，玛雅，俺，财富，自然，稀缺，水资源，城市化，房地产，危机，小镇
		V (4/19)	创造，再造，远离，生存
		Total	19
Business	People	N (8/9)	雾，霾，环保，口罩，天气，空气，市民
		V (1/9)	污染
		Total	9
	Sohu	N (13/16)	雾，霾，环保, 光，伏，能源，问题，我们，西部，发电，政策，我国，经济
		V (3/16)	发电，发展，消耗
		Total	16
Military	People	N (9/13)	雾，霾，问题，能见度，中国，飞行员，飞机，政府，地方
		V (4/13)	影响，起降，引导，着陆
		Total	13
	Sohu	N (8/11)	雾，霾，激光，武器，导弹，目标，天气，

			美军
		V (3/11)	影响，发射，攻击
		Total	11

Tableau des termes nominaux et verbaux les plus fréquents dans quatre rubriques communes du people et du sohu

Nom du sous-corpus	Catégorie lexicale	Quantité des termes exclusifs du BBS	Quantité des termes exclusifs du BLOG	Quantité des termes exclusifs du BUSINESS	Quantité des termes exclusifs du MILITARY	Total
PEOPLE	N	1	10	4	7	22
	V	0	3	1	3	7
SOHU	N	4	13	10	6	33
	V	1	4	3	2	10

Tableau de la quantité des termes exclusifs dans quatre rubriques similaires partagées du people et du sohu

⇒ Rubrique similaire entre *people* et *gov*

Nom de la rubrique	Nom du sous-corpus	Catégorie lexicale	Les plus fréquents
Lingdao	People	N (8/8)	雾，霾，北京，空气，两会，网友，南京，我
		V (0)	
		Total	8
	Gov	N (8/12)	雾，霾，大气污染，政府，措施，安全，环境，能源
		V (4/12)	防治，加大，落实，发展
		Total	12

Tableau des termes nominaux et verbaux les plus fréquents dans la rubrique similaire partagé du people et du gov

⇒ Rubrique similaire parmi *people* et *sohu et gov*

Nom de la rubrique	Nom du sous-corpus	Catégorie lexicale	Les plus fréquents
News	Gov	N (8/9)	雾，霾，地区，东部，北部，南部，西部，华北
		V (1/9)	笼罩
		Total	9
	People	N (7/9)	雾，霾，天气，中央，中国，污染，地区
		V (2/9)	影响，污染
		Total	9
	Sohu	N (12/14)	雾，霾，空气，污染，北京，大雾，天气，颗粒，高速，我们，烟花，伦敦
		V (2/14)	污染，影响
		Total	14

Tableau des termes nominaux et verbaux les plus fréquents dans la rubrique similaire parmi les trois sous-corpus

En tant que mots-clés, il n'est pas étrange que « le smog » et « le brouillard » soient croisés et répétés dans toutes les quatre rubriques similaires entre ‘‘le *people* et le *sohu*’’, entre ‘‘le *people* et le *gov*’’ ou ‘‘parmi les trois sous-corpus’’. Les mots exclusifs, quant à eux, permettent de mettre en évidence les caractéristiques de chaque rubrique, d'une part, par exemple le mot « je » dans la rubrique « BBS » du *people* confirme les caractéristiques privée et personnelle de cette rubrique par rapport aux autres ; et de différer le style l'un de l'autre, d'autre part, par exemple dans la rubrique « Blog » du *sohu*, on se sert d'un style littéraire et métaphore pour exprimer un véritable sentiment émotionnel et personnel. A partir de ces tableaux ci-dessus, comparons dans un second

temps les trois sous-corpus au niveau de la quantité et du sens des mots en groupe[24].

- **Le résumé du premier groupe entre *people* et *sohu*:**

Que ce soit au niveau de la quantité totale des mots indiqués dans quatre rubriques ou de la quantité des termes exclusifs, *sohu* dépasse toujours *people*. Cela montre que par rapport au *people*, la particularité du *sohu* est plus marquante ; en plus, au niveau du sens des mots, sauf les mots apparus dans la rubrique « Military », la plupart des mots du *people* sont déjà apparus dans d'autres parties d'analyses. Dans le sens strict, ils ne sont pas forcément des termes proprement dit « exclusifs », et nous pouvons même les considérer comme des mots servant seulement à différencier les rubriques communes qu'on peut trouver à la fois chez *sohu* et chez *people*.

- **Le résumé du deuxième groupe entre *people* et *gov* :**

Gov est supérieur au *people* tant au nivau du nombre total des mots indiqués dans le tableau que du nombre total des termes exclusifs colorés en rouge. Malgré cela, tous les termes en rouge du *gov*, en matière du sens des mots, sont déjà apparus dans les analyses précédentes, alors que la moitié des ceux en rouge du *people* ne sont pas dans le même cas. En tant que sous-corpus purement institutionnel, *gov* garde son homogénéité en parlant des actions prises par le gouvernement chinois tout au long de notre analyse, ce caractère pénètre même dans la comparaison dans la rubrique « Lingdao » partagée avec *people*. Ce qui se contraste avec *people* où on a constaté un changement de style d'organisation et de format Q&A pour inciter le grand public à à participer à la discussion du smog épais.

- **Le résumé du troisième groupe parmi *gov*, *people* et *sohu*:**

À part la supériorité de la quantité totale des mots du *sohu*, les deux autres sous-corpus sont coude à coude en cette matière. Quant au sens du mot, les termes indiqués

[24] Premier groupe : la comparaison entre people et sohu ; deuxième groupe : la comparaison entre people et gov ; troisième groupe : la comparaison parmi les trois sous-corpus.

dans les trois sous-corpus sont tous des termes répétitifs dans les parties précédentes, reste à indiquer que tous les trois gardent leur homogénéité dans la rubrique « News » qu'il s'agisse de l'organisation du texte ou de position prise à propos du smog épais de la Chine.

5. Conclusion

Ayant pour objectif de départ de comparer les textes dans trois types de sites afin d'en relever les attitudes et les réactions différentes entre le gouvernement et le grand public chinois, une série d'analyses ont été effectuées : en externe parmi les trois sous-corpus sur les mots les plus fréquents et les mots les plus co-occurrents des mots-clés « le smog » et « le brouillard » ; en interne à l'égard des rubriques similaires partagées des sous-corpus sur les mots les plus fréquents. Nous allons maintenant traiter les résultats d'analyses sous trois angles :

- Organisation des textes ;
- Contenu et l'opinion principaux sur « smog épais en Chine » ;
- Répartititon des mots selon la catégorie lexicale.

En tant que première cible de recherche, *gov* est un sous-corpus purement institutionnel. La caractéristique principale est l'homogénéité, qui enfile tout au long de l'analyse, à la fois au niveau du style et du point de vue essentiel sur le smog épais en Chine. Nous parlons d'abord du style d'organisation des textes du *gov*. À travers les études, les textes du *gov* sont tellement structurés et organisés qu'ils gardent un style uniforme dans toutes les rubriques, cela devient plus évident lors de la mise en propre du corpus, au cours de laquelle nous pouvons établir une simple règle pour enlever les bruits des textes que ce soit sa rubrique d'appartenance. Ensuite, quant au point de vue principal reflété par le contenu au sujet du « smog épais en Chine », *gov* indique toutes les régions qui pâtissent du brouillard de pollution en Chine avec les termes nominaux, il centralise tous les aspects sur les actions et les mesures prises par le gouvernement chinois, reflétéés par les termes verbaux apparus de manière répétitive. En ce qui concerne la répartition des mots selon la catégorie lexicale, les verbes sont plus nombreux que les noms, ce qui s'explique par la volonté du site de propager les actions et les mesures prises par le gouvernement chinois face à la pollution atmosphérique.

Sohu est un sous-corpus purement informel et médiatique. De nature complètement opposée au *gov*, il a été sélectionné par exprès. Revêtu d'un style bien varié et divers, ses styles d'organisation et de rédaction des textes varient selon les rubriques. Chaque rubrique propose ses propres mots-clés (un seul mot ou des mots combinés), dans la mesure où les mots les plus fréquents ou les termes les plus co-occurrents des deux mots-clés ne sont pas les mêmes dans différentes rubriques. Malgré cela, *sohu* est à même de garder son « opinion » verticalement identique : on introduit les origines, la nocivité du smog épais en Chine, en même temps, les mesures de protection prises par les Chinois sont aussi présentées à travers les termes nominaux du *sohu*. Quant aux termes verbaux, on peut dire que deux fonctions leur ont été confiées, étant donné que la plupart des verbes du *sohu* sont des verbes de modalité : premièrement, on exprime leur souhaits et espérances sur la situation de pollution dans leur propre pays à travers ces verbes ; deuxièmement, les verbes de modalité mettent en relief la nature privée de ce sous-corpus que les gens peuvent s'y exprimer librement. Concernant la répartition des mots en fonction de la catégorie lexicale, on a plus de nominaux que verbaux. Nous pouvons en déduire que *sohu* insiste sur les origines, les nocivités et les mesures de protections prises par les Chinois face au smog épais.

Dans l'intention de choisir un site qui combine les caractéristiques du *gov* et du *sohu*, *people* est choisi au moment voulu, il partage le côté institutionnel comme *gov* et la côté médiatique que *sohu*. Les résultats de travail comparatif confirment sa caractéristique de généralité. *People* neutralise les styles vifs et distincts du *gov* et du *sohu*, en synthétisant les « points essentiels » transmis par ces deux derniers. Présentons d'abord le style d'organisation des textes du *people*. Par rapport au *gov* qui est rigide et stable, le contenu *people* est beaucoup plus souple et varié qui permet de changer son style en fonction des rubriques, par exemple, l'emploi du format Q&A du *people* dans la rubrique similaire « Lingdao » partagée du *people* et du *gov*, alors que par rapport au *sohu*, *people* est relativement plus strict, la répétition des mots dans les quatre rubriques partagées du *people* et du *sohu* en est le meilleur témoin. Puis en ce qui concerne le point central du *people*, il parle aussi deux côtés de sens en synthétisant les idées du *gov* et du *sohu*: les termes verbaux mettent en évidence les mesures prises par le

gouvernement chinois, et les nominaux présentent les nocivités et les problèmes produits par le smog épais. La seule particularité du *people* est marquée par deux mots : « l'expert » et « la Chine ». Le premier est en fait à l'opposé du « je », au lieu d'exprimer des sentiments personnels et émotionnels, « l'expert » transmet plutôt des messages objectifs et scientifiques au sujet de l'air de pollution ; « la Chine » est utilisé pour mettre en contraste la Chine avec d'autres pays qui subissent aussi le smog épais, le but final est d'en tirer des expériences et trouver les solutions les plus pertinentes et efficaces pour régulariser les problèmes en Chine. Finalement, il n'y a pas de borne nette entre la quantité des termes verbaux et nominaux du people, cette dernière spécificité lui attribue de nouveau le rôle neutre par rapport au *sohu* et au *gov*.

Les caractéristiques manifestées par les trois sous-corpus peuvent être résumées par la divergence des attitudes et des réactions du milieu gouvernemental et du grand public. Il n'est pas difficile de voir que chaque partie s'intéresse au smog épais, le gouvernement présente ses actions immédiates pour étudier la situation de pollution sur le territoire chinois et les mesures prises par la suite ; alors que le grand public s'intéresse plutôt à des questions telles que « D'où vient le smog épais ? » « Quelles sont les nocivités que le smog épais produit dans la vie courante ? » « Comment se protéger contre le brouillard quotidiennement ? », ils cherchent les réponses sur internet en « exprimant leur souhaits » et par eux-mêmes.

Bibliographie et sitographie

Lebart, L. & Salem, A. (1994). - *Statistique Textuelle*, DUNOD, Paris.

René L'Écuyer, Ph.D., (1990). - *Méthodologie de l'analyse développementale de contenu*. *Méthode GPS et Concept de Soi*, Presses de l'Université du Québec, Québec.

Kim Gerdes, Corpus collection and analysis for the linguistic layman: - *The Gromoteur, Proceedings of the JADT 2014*, Paris.

Luc ALBARELLO, Étienne BOURGEOIS, Jean-Luc GUYOT (2007), - *Statistique Descriptive*. *Un outil pour les praticiens – chercheurs*, De Boeck&Larcier, Bruxelles.

Achard P. (1993) - *La sociologie du langage*, *Que-sais-je* ? PUF, Paris.

Bardin L. (1989) - *L'analyse de contenu*, PUF, Paris.

Francine M. (2010) - *L'analyse du discours*, PUF, Paris.

Benveniste E. (1996) – *Problèmes de linguistique générale*, Gallimard, Paris.

Jean de BONVILLE. (2006) – *L'analyse de contenu des médias : De la problématique au traitement statistique*, De Boeck Supérieur, Bruxelles.

Fleury, S., and A. Salem (2002), Lexico 3. Outil de statistique textuelle. Paris: Université Sorbonne Nouvelle-Paris 3.

Guiraud P. (1960) – *Problèmes et méthodes de la statistique linguistique*, P.U.F., Paris.

MUCCHIELLI, R. (1974) – *L'analyse de contenu des documents des communications : connaissance du problème*, Paris, Les Éditions ESF.

Les sources du sous-corpus gov : www.gov.cn

Les sources du sous-corpus sohu : www.sohu.com

Les sources du sous-corpus people : www.people.com.cn

L'adresse du site web de l'outil gromoteur : http://gromoteur.ilpga.fr/

L'adresse du site web de l'outil développé par le *Institute of Applied Linguistice Ministry of Education in China* : http://www.cncorpus.org/

Analyse du discours, Frank COBBY, 2009, disponible sur site : http://www.analyse-du-discours.com/l-analyse-de-contenu-du-discours