

L'opinion face à la crise Google dans les pays sinophones

Li-Chi WU

CLESTHIA / SYLED – Université Sorbonne Nouvelle – lucielichi@gmail.com

Abstract

In the Web 2.0 era, people access information not only through traditional mass media (print media, radio, television) and communication is no longer the exclusive domain of political powers; online communities for information exchange are created ever more frequently and freely on the Internet. Political and economic decisions, whether by governments or commercial enterprises, today can no longer ignore public opinion. However, the communication tool itself is never neutral, it deeply influences the content and the nature of the messages and indeed Internet users posting in community spaces often deviate from standard language found in editorial writing. We analyzed a large corpus of texts from Chinese language networks about the *Google case* occurring in 2010. Our goal is to apply the approach of textual statistics in order to identify an overall typology or a differential system of opinions about this *case* present on the Internet. We distinguish texts of the type *article* and texts of the type *comment* and compare the output in the People's Republic of China and in other Chinese speaking countries. The analysis of these texts, on the basis of the vocabulary used, will be conducted in relation to the place of publishing of the different media concerned. We will see also that the brevity of the Chinese language is very well suited for communication exchange between Internet users, as shown by the strong reactions to the *Google case* on the Internet in mainland China and other Chinese speaking countries.

Résumé

À l'ère du Web 2.0, l'information ne passe plus seulement par les médias traditionnels (presse, radio, télévision) et la communication n'est plus l'apanage du pouvoir ; des communautés virtuelles d'échange d'information se créent de plus en plus librement sur Internet. Qu'il s'agisse de définir des mesures administratives ou des stratégies d'entreprise, les décideurs politiques et économiques ne peuvent plus aujourd'hui ignorer la voix de l'opinion. Or, l'outil de communication n'est jamais neutre, il influence profondément le contenu et la nature même des messages et l'on constate en effet que le style des internautes rompt souvent avec les usages de la langue standard typiques de l'écriture éditoriale. Nous avons procédé à l'analyse d'un vaste corpus de textes émanant des médias chinois sur la toile ; ils concernent tous l'*affaire Google*, survenue en 2010. Notre objectif est d'expérimenter, sur ce corpus, l'approche statistique textuelle pour définir une typologie générale ou un système différentiel des opinions concernant cette *affaire* sur le net. Nous y distinguerons les textes du type *article* et les textes du type *commentaire* et nous comparerons les productions de RPC et celles des autres pays sinophones. L'analyse de ces textes, réalisée sur la base du vocabulaire employé, sera rapportée aux informations relatives au lieu d'édition par les différents médias concernés. Nous verrons également que la concision de la langue chinoise se prête admirablement au mode de communication des internautes et à l'expression des fortes réactions qu'a suscitées l'*affaire Google* sur la toile sinophone.

Mots-clés : statistique textuelle, textométrie, fouille des données d'opinion, analyse du discours médiatique, fouille textuelle, monosyllabisme

1. Introduction

Grâce au plein essor du monde virtuel au cours des dernières années, l'information et les connaissances se propagent rapidement et abondamment sous forme de *communication*

*médiée par ordinateur*¹ (Panckhurst, 1997). L'élaboration du message par les rédacteurs au service des médias – les émetteurs –, est donc de la plus haute importance pour obtenir l'adhésion et la compréhension des différents publics visés et éviter les erreurs d'interprétation. Le sens du message ne se réalise en effet pleinement que dans la rencontre des schémas d'encodage du destinataire et des schémas de décryptage du destinataire, c'est pourquoi les stratégies de communication des médias audiovisuels ou de la presse écrite prennent une place prépondérante dans la transmission de l'information. La ligne éditoriale des médias de masse est généralement fondée sur un ensemble de prises de position, de repères socio-culturels, d'*a priori* idéologiques ou politiques liés au contexte socio-historique dans lequel ils sont nés. Avec le développement accéléré de l'Internet, la presse en ligne est devenue l'espace idéal de la propagande politique dans le monde sinophone comme ailleurs. Mais, à la différence des médias occidentaux, qui, en général, ne sont pas directement sous le contrôle de l'État ou d'un parti politique (même s'ils assument souvent une orientation politique), beaucoup de médias chinois puisent directement leurs informations aux sources gouvernementales (Cao, 2005 ; Clérot, 2010) et sont parfois même organiquement liés avec l'État et le Parti au pouvoir. C'est pourquoi nous avons décidé d'utiliser une méthode textométrique, en l'occurrence, le *calcul des spécificités* (Lafon, 1980, 1984 ; Lebart et Salem, 1994) pour étudier la polémique survenue entre *Google* et l'État chinois. Le lieu géographique est un des critères majeurs permettant de distinguer les différentes voix qui s'expriment dans la presse écrite diffusée sur l'Internet. Nous voulions également savoir si les commentaires auxquels donnent lieu les articles de la presse en ligne se conformaient aux orientations respectives des médias qui les publient. Les textes publiés dans les espaces de discussion en ligne sont en effet rédigés dans un langage moins formel que celui des médias institutionnels – ce qui permet aux utilisateurs d'exprimer leurs pensées et leurs sentiments individuels, éventuellement en rupture avec la ligne officielle – tout en participant d'une même conscience nationale.

2. Le contexte du conflit

La bataille entre le géant des moteurs de recherche et le plus grand marché du monde ayant fait couler beaucoup d'encre sur la toile chinoise, nous avons voulu savoir quels étaient les discours diffusés par les médias sinophones sur le web et comment ils se propageaient.

2.1. Étapes de la crise entre la Chine et Google

La bataille entre le géant des moteurs de recherche, Google, et le plus grand marché du monde, la Chine, a eu lieu entre janvier et juillet 2010 : l'*affaire Google*, à laquelle ont pris part les dirigeants du pays, les gérants des entreprises, les politiciens, a suscité de nombreux commentaires sur la toile chinoise. L'*affaire* a débuté par un billet publié sur le blog officiel de l'entreprise *Google* annonçant son retrait éventuel du marché chinois en raison d'une série de cyber-attaques chinoises contre les comptes Gmail de militants des droits de l'Homme. Elle s'est terminée par une mise hors-circuit de la version chinoise du site dès lors que la

¹ Le terme anglais *computer-mediated communication* (CMC) désigne une forme de communication entre deux ou trois personnes par ordinateurs interposés. On traduit souvent ce terme en français par « *communication assistée par ordinateur* » (CAO) ou par « *communication médiatisée par ordinateur* » (CMO). Selon Panckhurst (1997) qui a proposé de traduire *computer-mediated* par « *communication médiée par ordinateur* », l'ordinateur serait le médiateur qui modifierait indirectement le discours ; il induirait la création d'autres formes, d'autres « genres » de discours.

société américaine ne se conformait plus aux accords concernant le contrôle des flux d'information conclus lors de son implantation sur le marché chinois en 2006. Pour mémoire, suivent ci-dessous les étapes-clefs qui ont marqué le contentieux entre la Chine et *Google*.

- Janv. 2006 ouverture du site google.cn en Chine à la suite de négociations avec les autorités chinoises ; Google accepte un contrôle des flux d'information ;
- Nov. 2009 conflit entre *Google* et l'association des écrivains chinois sur les droits d'auteur de leurs livres numérisés mis en ligne ;
- Janv. 2010 protestation de *Google* auprès du gouvernement chinois à la suite de cyberattaques contre des comptes Gmail appartenant à des militants chinois des droits de l'Homme ;
- Mars 2010 annonce par *Google* de la mise hors-circuit de google.cn qui entreprend de rediriger les internautes vers un site miroir domicilié à Hong Kong et ne respectant plus les règles négociées en 2006 ;
- Juin 2010 menace du gouvernement chinois de ne pas renouveler la licence de l'entreprise *Google* si elle ne se conforme pas à la législation chinoise. Fermeture du site google.cn.

2.2. Du langage standard au langage des supports numériques

Ces nouvelles *manières de dire* ne se conforment pas aux règles et aux normes de la langue réputée standard. Elles revêtent une forme particulière adaptée à la rapidité du monde virtuel. Ce nouveau langage se caractérise par un pouvoir d'expression plus fort, par un sens du lexique plus direct et plus audacieux. Les internautes n'hésitent pas à exprimer leurs opinions et leurs émotions dans un langage familier, voire argotique, émaillé de mots vulgaires et grossiers, bannis des journaux et des magazines en ligne (Li, 2006).

3. Le corpus *GoogleCh*

Les textes chinois réunis dans le corpus *GoogleCh*² ont été recueillis de janvier à septembre 2010. Ils ont été prélevés sur les sites de trente-quatre médias distincts du web chinois. Sept de ces médias (4 en Chine populaire, 3 à l'extérieur) autorisent les internautes à réagir aux *articles* de la rédaction par le biais de *commentaires* qui sont ensuite postés sur le site : le corpus regroupe ainsi 227 articles et 8 634 commentaires. L'écriture chinoise étant une *écriture continue* sans blancs typographiques entre les caractères, nous avons utilisé un outil de segmentation pour procéder aux découpages en occurrences de formes lexicales³. On dénombre sur l'ensemble du corpus 343 919 segments utilisant 15 341 formes lexicales différentes ; 5 608 de ces formes sont des hapax ; la particule 的 *de*⁴, forme la plus fréquente, compte 19 800 occurrences. Nous détaillons les données chiffrées selon le type de textes et le

² Afin de faciliter la manipulation des textes, les ponctuations chinoises, les nombres numéraux et les lettres latines sous forme de différents formats typographiques ont été transcrits dans les formes que nous utilisons habituellement dans les langues alphabétiques. Les textes sont sauvegardés au format texte brut avec le jeu de caractères GB2312.

³ Nous avons utilisé l'outil de segmentation ICTCLAS élaboré pour le traitement des textes d'articles de presse (Wu, 2010). Le segmenteur d'ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) a été mis au point par Huaping Zhang à l'Institut de technologie informatique de l'Académie des Sciences de Chine ; il est adapté à la recherche. Nous avons utilisé la version 2008, disponible en téléchargement sur le site officiel <http://ictclas.nlpir.org/> (page consultée le 20 déc. 2013).

⁴ La particule structurale 的 *de*, la plus usitée en chinois moderne, relie le déterminant au déterminé. Par exemple, 我的书 *wo de shu* (mon livre / moi *de* livre).

lieu d'édition des textes dans le tableau 1. On observe quelques traits caractéristiques sur l'ensemble du corpus :

- les textes des commentaires sont en moyenne 1,7 fois plus longs que ceux des articles proprement dits (217 800 segments contre 126 119).
- quant au nombre de formes différentes, le type *commentaire* en possède 1,2 fois (11 191/9 587) plus élevé que le type *article*.
- la longueur des textes publiés varie également selon le lieu d'édition : elle est 4 fois plus élevée en République populaire de Chine (RPC) que dans les textes publiés à l'extérieur du territoire (278 365/65 554) ; le nombre de formes est 1,8 fois plus élevé dans les textes publiés en RPC que dans ceux de l'extérieur (17 732/9 636).
- le renouvellement du vocabulaire est moins rapide en RPC qu'à l'extérieur de la Chine.

| Lieu | Type de textes | Nb d'occ | Nb de formes | Nb d'hapax | Fréquence maximale |
|---------------|----------------|----------|--------------|------------|--------------------|
| Chine | Articles | 82 342 | 7 609 | 3 236 | 5 168 的 |
| | Commentaires | 196 023 | 10 123 | 3 708 | 10 850 的 |
| | Total | 278 365 | 17 732 | 6 044 | 16 018 |
| Hors de Chine | Articles | 43 777 | 5 635 | 2 586 | 2 338 的 |
| | Commentaires | 21 777 | 4 001 | 2 062 | 1 444 的 |
| | Total | 65 554 | 9 636 | 4 648 | 3 782 |

Tableau 1. Principales caractéristiques quantitatives du corpus *GoogleCh*

En ce qui concerne les textes du type *commentaire*, on observe les caractères suivants :

- la longueur des commentaires publiés en Chine est neuf fois supérieure à celle des textes provenant de l'extérieur du territoire chinois (196 023/21 777).
- le nombre de formes recensées sur les sites chinois est 2,5 fois supérieur à celui qu'on observe sur les sites extérieurs (10 123/4 001).
- la longueur de chaque commentaire publié hors de RPC est presque trois fois supérieure à celle d'un commentaire publié en RPC : 64,6 occurrences (196 023/8 297) contre 23,6 (21 777/337).
- les internautes chinois de l'extérieur font preuve de plus d'originalité, produisent des discours moins stéréotypés que leurs homologues de RPC.

Ces données quantitatives montrent que le lieu d'édition influe grandement sur la longueur moyenne des textes produits et sur la diversité relative du lexique utilisé. Elles nous permettent de tirer deux conclusions : elles témoignent d'abord d'une différence de nature entre les articles-sources et leurs commentaires. Globalement, les articles sont plus originaux que les commentaires mais les textes publiés en Chine populaire, qu'il s'agisse des articles-sources ou des commentaires, sont plus stéréotypés que les textes publiés dans les autres pays sinophones.

4. Variations lexicales et prises de position dans le débat

L'analyse présentée sur la figure 1 montre la ventilation du vocabulaire spécifique à l'intérieur de chaque publication.

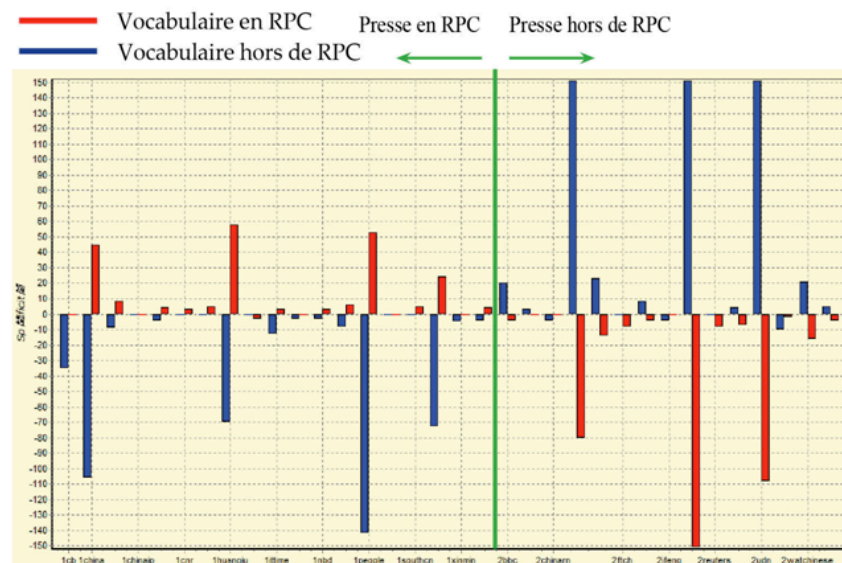


Figure 1. Ventilation du vocabulaire spécifique des médias chinois

On voit que certains médias utilisent de préférence le vocabulaire caractéristique du groupe auquel ils appartiennent. Par exemple, le vocabulaire spécifique des médias publiés à l'extérieur de la RPC est sur-employé dans trois journaux taiwanais (*China Times*, *UDN*, *Liberty Times*). De même, le vocabulaire spécifique des médias de la RPC est sur-employé dans quatre journaux, *China*, *Huanqiu*, *People*, *Xinhua*.

4.1. Les oppositions lexicales

Deux listes de vocabulaire, correspondant au vocabulaire spécifique⁵ pour chacun des lieux d'édition, sont présentées sur la figure 2.

| Médias en Chine | | | Médias hors de Chine | | |
|-----------------|-------|------------|----------------------|--------|-----------------|
| Formes/SR | chine | hors_chine | Formes/SR | chine | hors_chine |
| 互联网 | 483 | +34 65 | -34 | google | 230 *** 734 *** |
| 网络 | 324 | +30 32 | -30 | 网路 | 3 *** 283 *** |
| 了 | 779 | +17 225 | -17 | 大陆 | 36 *** 283 *** |
| 黑客 | 140 | +16 10 | -16 | 报导 | 8 *** 194 *** |
| 美国 | 522 | +15 136 | -15 | 报导 | 0 90 +43 |
| 信息 | 213 | +13 37 | -13 | 中共 | 0 70 +33 |
| 的 | 5223 | +12 2338 | -12 | 骇客 | 0 62 +30 |
| 记者 | 188 | +12 31 | -12 | 资讯 | 7 -20 57 +20 |
| 报道 | 143 | +11 20 | -11 | 谷歌 | 0 41 +20 |
| 内地 | 145 | +11 21 | -11 | 封锁 | 3 -19 48 +19 |
| 搜索引擎 | 195 | +11 37 | -11 | 当局 | 7 -17 51 +17 |
| 国家 | 221 | +10 48 | -10 | 法轮功 | 1 -15 33 +15 |
| 文章 | 84 | +10 6 | -10 | 官方 | 47 -15 93 +15 |
| 谷歌 | 2747 | +10 1179 | -10 | 网页 | 15 -14 55 +14 |

Figure 2. Spécificité du vocabulaire des médias chinois

⁵ L'analyse des spécificités, méthode proposée par Pierre Lafon (1980, 1984), permet de mesurer la probabilité de l'apparition d'une unité textuelle dans une partie de corpus donnée. L'analyse indique un sur-emploi ou un sous-emploi des unités spécifiques en fonction de leur fréquence dans l'ensemble du corpus.

Deux mots-clés de cette étude, le nom officiel en chinois de la société *Google*, 谷歌 *guge*⁶, et le nom original, *Google*, sont spécifiques, respectivement aux médias de la Chine populaire (2 747/3 926 occurrences, spécificité +10) et aux médias sinophones extérieurs à la RPC (734/964, spécificité > +43). Les médias publiés en Chine ont une prédilection pour l'usage du nom chinois, alors que les médias publiés à l'extérieur préfèrent garder le nom original. Dans les médias publiés en RPC, on observe d'autres termes spécifiques comme 美国 *meiguo* (USA), 国家 *guojia* (pays), 法律 *falü* (loi), 标准 *biaozhun* (règle), 我们 *women* (nous), 遵守 *zunshou* (respecter) ; les termes sur-employés dans les médias édités à l'extérieur sont 大陆 *dalu* (continent), 中共 *zhonggong* (la Chine communiste)⁷, 封锁 *fengsuo* (bloquer), 法轮功 *falungong* (Falun Gong), 天安门 *tian'anmen* (Tian'anmen)⁸, 配合 *peihe* (coordonner). Le vocabulaire de la première liste, qui correspond aux formes spécifiquement utilisées en RPC, fait apparaître des formes qui appartiennent au domaine juridique et à celui des relations du pays avec le monde extérieur. Dans le vocabulaire de la seconde liste, formes utilisées spécifiquement à l'extérieur de la Chine, on remarque la présence d'un lexique critique, employé pour dénoncer la politique autoritaire du gouvernement chinois. La confrontation des deux listes permet également de mettre en évidence le contraste lexical entre communautés sinophones, particulièrement entre la Chine et Taïwan.

4.1.1. Le thème "Google et les lois chinoises"

Intéressons-nous, maintenant, aux cooccurrences du terme, 谷歌 *guge*, et aux relations sémantiques que ces co-occurents entretiennent avec la forme-pôle⁹. L'unité lexicale 法律 *falü* (loi) apparaît essentiellement dans les médias de RPC (262/333 occ.). Sa fréquence y est trois fois plus élevée que dans les textes publiés à l'extérieur de la RPC (71/333 occ.). Les segments répétés, suites de formes graphiques répétées à l'identique, permettent de clarifier les réseaux sémantiques au sein du vocabulaire co-occurent¹⁰. Les segments 遵守中国 *zunshou zhongguo* (respecter la Chine ; 81 occ. +3), 遵守 *zunshou* (respecter ; 136 occ. ; spécificité +3), 遵守中国法律 *zunshou zhongguo falü* (respecter les lois chinoises ; 51 occ. +3), 按照 *anzhao* (selon ; 49 occ. +2), 按照中国法律 *anzhao zhongguo falü* (selon les lois chinoises ; 16 occ. +2), sont mis en valeur autour de la forme 谷歌 *guge*.

⁶ Les caractères chinois présents dans cette étude sont accompagnés de leur transcription en *pinyin*, le système de transcription romanisée des caractères chinois le plus répandu de nos jours.

⁷ 中共 *zhonggong* est le sigle de 中国共产党 *zhongguo gongchandang*, le Parti Communiste Chinois.

⁸ Le Falun Gong, un art de vivre inspiré des traditions bouddhiste et taoïste, est interdit par les autorités chinoises : le nombre de ses adeptes, sans doute supérieur à celui des partisans du Parti communiste chinois, représente une menace inacceptable pour le pouvoir. Des manifestations étudiantes revendiquaient des réformes politiques et démocratiques sur la place Tian'anmen à Pékin en avril 1989. Le mouvement a fini par une violente répression de la part du gouvernement chinois.

⁹ Une cooccurrence désigne l'apparition simultanée de deux formes données dans un même fragment de texte. La méthode des *cooccurrences* permet de mettre en évidence les relations d'attraction lexicale des unités linguistiques (Lafon, 1984).

¹⁰ La méthode de l'*Inventaire des Segments Répétés* (ISR) permet de mettre en évidence les associations de formes employées dans un corpus comme des expressions figées, répétées, des couples de formes présentes (Lafon & Salem, 1983 ; Salem, 1986, 1987). Les segments répétés désignent dans un texte une suite de formes dont la fréquence est égale ou supérieure à deux.

| Forme pôle : 按照 (selon) | | | Forme pôle : 遵守 (respecter) | | |
|-------------------------|-----------------------------|------|-----------------------------|---------------------------------------|------|
| Segments répétés | Équivalent traductionnel | Fréq | Segments répétés | Équivalent traductionnel | Fréq |
| 按照法律 | selon les lois | 6 | 遵守当地... | respecter ... du pays hôte | 7 |
| 按照法律法令来办 | entreprendre selon les lois | 5 | 遵守当地的法律 | respecter les lois du pays hôte | 3 |
| 按照中国法律 | selon les lois chinoises | 16 | 遵守东道国的 | respecter ... du pays hôte | 2 |
| | | | 遵守所在国... | respecter ... du pays où nous | 9 |
| | | | 遵守所在国的法律 | respecter les lois du pays hôte | 8 |
| | | | 遵守中国... | respecter ... chinois | 46 |
| | | | 遵守中国的 | respecter ... chinois | 35 |
| | | | 遵守中国的法律 | respecter les lois chinoises | 22 |
| | | | 遵守中国法律 | respecter les lois chinoises | 43 |
| | | | 遵守中国政府 | respecter ... du gouvernement chinois | 5 |

Guide de lecture :
 Cette présentation permet de visualiser aisément et de hiérarchiser les segments en fonction de leur longueur. Par exemple, parmi les 46 occurrences du segment 遵守中国... *zunshou zhongguo* (respecter ... chinois), 35 occurrences se prolongent en 遵守中国的 *zunshou zhongguo de* (respecter le/la/les ... chinois/chinoise/chinois). Sur ces 35 occurrences, 22 correspondent au segment 遵守中国的法律 *zunshou zhongguo de falü* (respecter les lois chinoises).

Tableau 2. Inventaire distributionnel des formes 遵守 *zunshou* (respecter) et 按照 *anzhao* (selon)

L'inventaire distributionnel réalisé autour des deux unités lexicales 遵守 *zunshou* (respecter) et 按照 *anzhao* (selon, d'après) dans le tableau 2 permet de hiérarchiser les entités corrélées aux expressions qui relèvent du vocabulaire juridique, et qui désignent les lois ou les règlements chinois¹¹. Le verbe 遵守 *zunshou* (respecter) et la préposition 按照 *anzhao* (selon, d'après) suivis des compléments précisant le sens du verbe, de la préposition se présentent sous des formes variées : 中国 *zhongguo* (la Chine), 当地 *dangdi* (de la localité, du pays), 东道国 *dongdaoguo* (pays hôte), 所在国 *suozaiguo* (pays hôte), lesquels désignent le pays « la Chine ». La mise en avant du pays, la Chine, est particulièrement marquée en RPC. La présence ou l'absence de ces segments, ainsi que leur fréquence, sont des indices qui permettent de cerner les opinions que véhiculent les différents médias chinois : du point de vue de la Chine, le respect des lois est une obligation fondamentale et indiscutable. Le ton affiché dans les énoncés est impératif, voire comminatoire (tableau 3).

| | Contexte | Traduction |
|---|---|---|
| | Contextes pour 遵守 <i>zunshou</i> (respecter) et 按照 <i>anzhao</i> (selon) | |
| 1 | [Source:China, Texte:a_19, Date:0312010] 如果做出不遵守中国法律 & 法令, 甚至做出违背中国法律 & 法令的事情, 就是不友好 & 不负责任的, 也是不应该的. | Si Google ne respecte pas les lois et les réglementations chinoises, voire s'il les viole, il est incorrect, irresponsable et il ne devrait pas l'être. |
| 2 | [Source:China, Texte:a_23, Date:03232010] 如谷歌公司执意将谷歌中国网站的搜索服务撤走, 那是谷歌公司自己的事情, 但必须按照中国法律和国际惯例, 负责地做好有关善后工作. | Si Google souhaitait mettre fin à son service en Chine, cela ne regarderait que lui. Il devrait sérieusement remédier aux suites de l'affaire conformément aux lois chinoises et à l'usage international. |
| | Contextes pour 我们 <i>women</i> (nous) et 自己 <i>ziji</i> (soi-même) | |
| 3 | [Source:Huanqiu, Texte:a_50, Date:01152010] 我们欢迎谷歌留下, 但留下来的谷歌必须要做出调整, 也要适应中国的法律. | Nous (la Chine) étions favorables à Google. S'il souhaite rester, il doit se conformer et s'adapter aux lois chinoises. |
| 4 | [Source:Huanqiu, Texte:a_49, Date:01152010] 在过去几年中, 正是谷歌自己选择配合北京进行自我过滤. | Depuis des années, c'est Google qui a décidé d'avoir des accords avec Pékin. Mais il a choisi de se faire censurer. |

Tableau 3. Contextes pour 遵守 *zunshou* (respecter) et 按照 *anzhao* (selon), 我们 *women* (nous) et 自己 *ziji* (soi-même)

¹¹ La méthode de l'inventaire distributionnel se présente comme une extension de la technique des segments répétés dont elle permet d'affiner les résultats en hiérarchisant les séquences repérées pour montrer celles qui sont les extensions d'autres segments.

4.1.2. Les désignations de la Chine

Différents termes servent à désigner le gouvernement chinois ou la Chine. Dans les médias de RPC, les formes spécifiques sont des pronoms personnels comme 我 *wo* (je) et 我们 *women* (nous) et des noms ou des groupes nominaux comme 内地 *neidi* (le territoire intérieur)¹², 国家 *guojia* (le pays, l'État), 华 *hua* (la Chine), etc. Dans les médias sinophones extérieurs à la RPC, les entités désignant la Chine sont 大陸 *dalu* (le continent), 中共 *zhonggong* (la Chine communiste), 官方 *guanfang* (le gouvernement, les autorités), etc. Le contraste dans l'utilisation du vocabulaire désignant la Chine est donc avant tout lié à la situation géographique respective des médias et à la position idéologique dominante que cela implique. Ces termes, qui nous permettent d'identifier l'émetteur du message et de le situer dans le champ des oppositions géographiques, idéologiques ou politiques déjà repérées, véhiculent des jugements et des points de vue propres à l'énonciateur (l'auteur d'un article) ou aux médias eux-mêmes. Selon que les médias publient en RPC ou à l'extérieur, la société Google et l'État chinois occupent une place différente par rapport à l'affaire Google tout en exprimant leurs avis respectifs au moyen des mêmes pronoms personnels 我 *wo* (je), 我们 *women* (nous), 自己 *ziji* (soi-même/eux-mêmes), qui prennent un sens différent selon le lieu d'où émane le message. Ces pronoms constituent des traces consistantes du contexte qui mettent les énoncés en relation avec leur situation d'énonciation (Kerbrat-Orecchioni, 2002). À l'aide de ces termes-clés, nous avons repéré les contextes dans lesquels opinions et critiques prennent sens (tableau 3). Le ton des énoncés dans les médias de RPC est toujours dur et cassant. S'adapter à la culture chinoise est un impératif pour toute entreprise étrangère installée en Chine, y compris pour la grande société qu'est Google. C'est également le message que le gouvernement chinois veut transmettre à Google.

5. Le retour tendanciel au monosyllabisme chez les internautes

Avant de mettre en évidence les caractéristiques du vocabulaire utilisé dans les commentaires publiés à la suite des articles-sources, il nous faut, tout d'abord, faire retour sur l'évolution de la langue chinoise. Nous pourrions alors mesurer l'originalité relative de ce vocabulaire et en dégager le sens dans une double perspective, diachronique et synchronique. L'évolution de la langue chinoise se caractérise par le passage d'un système monosyllabique à un système polysyllabique. Le lexique de la langue classique s'est enrichi grâce à la culture occidentale et cette évolution s'est poursuivie jusqu'au chinois moderne où la plupart des unités lexicales sont dissyllabiques. Or, par un étrange renversement, on constate aujourd'hui un retour vers le monosyllabisme originel : le rythme trépidant de la vie moderne tend à privilégier la lecture de textes courts au détriment de longs discours, voire, diront certains, à *appauvrir* la langue. Le développement d'Internet accentue encore ce phénomène : les monosyllabes, dans lesquels les internautes trouvent un vocabulaire plus rapide et plus expressif que dans la langue standard, reviennent en force sur la toile (Zheng, 2009).

5.1. La longueur des unités lexicales

L'analyse des textes publiés par les médias de RPC nous conduit à étudier ce qui les oppose aux textes produits par leurs homologues à l'extérieur.

¹² 内地 *neidi*, terme géographique, est utilisé en Chine, à Hong Kong, à Macao ou dans les régions au bord de mer pour désigner les régions de l'intérieur du pays. Il est synonyme des termes utilisés à Taïwan comme 大陸 *dalu* (continent) ou 中国 *zhongguo* (Chine) pour désigner la Chine continentale.

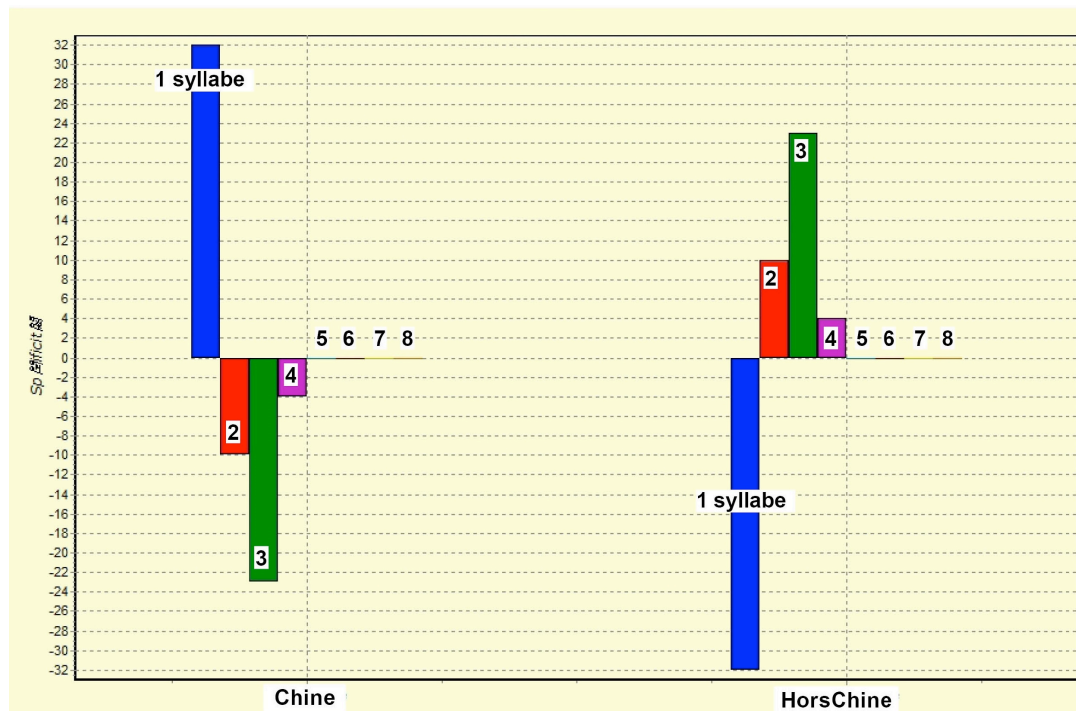


Figure 3. Nombre de syllabes des unités lexicales par lieu d'édition

Les données numériques du tableau 1, qui mettent en évidence un net décalage entre les deux sous-ensembles du type *commentaire*, nous ont amené à prendre en compte la longueur moyenne des unités lexicales propre à chacun d'eux. Le nombre de syllabes¹³ d'une unité s'étend d'une à huit. Or, si les monosyllabes et les dissyllabes occupent une place dominante tant en RPC que dans les autres pays sinophones, on observe une nette différence entre les deux types d'énoncés. Les unités dissyllabiques occupent la première place dans les articles, les monosyllabes dominent dans les commentaires, en particulier dans les publications chinoises. Le calcul des spécificités (cf. *supra*) permet de dégager les traits distinctifs de chaque lieu d'édition : les publications extérieures à la RPC privilégient les unités de deux à quatre syllabes alors que les textes publiés en RPC ont tendance à leur préférer les monosyllabes (figure 3).

5.2. Spécificités du vocabulaire

L'analyse du vocabulaire spécifique employé dans les commentaires publiés en RPC, permet de mettre en évidence ses caractéristiques lexicométriques, notamment, comme nous venons de le voir, la fréquente réduction des unités lexicales à une syllabe. Les termes du vocabulaire spécifique dont un nombre important de monosyllabes, sont classés dans ce qui suit selon la catégorie grammaticale à laquelle ils appartiennent.

1) les verbes

L'apparition d'une grande quantité de verbes monosyllabiques est, entre autres, caractéristique de l'évolution de la langue chinoise (He, 2010), en particulier sur Internet. Cette évolution concerne surtout les verbes exprimant la volonté, l'émotion, le désir, et autres

¹³ Une unité lexicale chinoise est composée de morphèmes ; l'unité est dite « simple » si elle est monosyllabique et s'écrit avec un seul sinogramme, et « complexe » si elle est polysyllabique et s'écrit avec plusieurs sinogrammes.

sentiments des locuteurs. Or, dans la langue courante, il est de bon ton d'utiliser la forme redoublée¹⁴ du verbe de préférence à sa forme simple. La fonction de la reduplication est d'atténuer la force illocutoire, l'intensité de l'action ou de restreindre sa portée dans le temps en termes de fréquence ou de durée (Paris, 2007). Ainsi, par exemple, le verbe 看 *kan*, dans sa forme simple, signifiera « regarder » ; dans sa forme redoublée, 看看 *kankan*, il signifiera « jeter un coup d'œil ». L'emploi des formes simples va de pair avec un ton plus sec, plus péremptoire, plus autoritaire. Les internautes chinois commentant l'affaire Google choisissent précisément la forme simple des verbes pour exprimer leur intransigeance et leur volonté de voir déguerpir sans délai le géant étranger : 要 *yao* (vouloir), 走 *zou* (s'en aller), 想 *xiang* (désirer), 退 *tui* (quitter), 出 *chu* (sortir), 吐 *tu* (vomir) (au sens de « détester »). Quant aux dissyllabes et aux polysyllabes, ce sont, par exemple, 稀罕 *xihan* (s'en fichier), 抵制 *dizhi* (boycotter), 不要脸 *buyaolian* (être éhonté). La détermination des internautes de RPC concernant le retrait du géant moteur de recherche est donc incontestablement véhiculée non seulement par le lexique mais encore par des formes grammaticales préférentielles.

2) les pronoms

Dans le corpus *GoogleCh*, les pronoms désignent généralement : soit l'émetteur du message, les auteurs des articles d'origine ou les internautes qui les commentent, soit leurs destinataires. Le pronom de la deuxième personne du singulier, 你 *ni*, le pronom le plus spécifique de notre corpus, compte 2 623 occurrences (spécificité +22) dans les commentaires publiés en RPC, contre 2 763 occurrences dans l'ensemble des commentaires. Il est principalement employé par les internautes pour interpeller leur destinataire, généralement l'entreprise *Google*, qu'il vise avec des propos exprimant l'ironie, la dérision ou le mépris. La première personne du pluriel, 我们 *women* (nous), spécifique en second lieu des pronoms (1 328 occ./1 386 occ. du total, spécificité +16), renvoie généralement aux énonciateurs. Contrairement au système pronominal français – dans lequel les pronoms personnels, les pronoms et les adjectifs possessifs présentent différentes formes vocales et graphiques –, le système pronominal chinois est remarquablement simple : toutes les marques de la première personne du pluriel – qu'il s'agisse de pronoms personnels ou possessifs – contiennent la forme 我们 *women* à laquelle on ajoute d'autres éléments pour former les différents syntagmes. L'analyse de cette unité lexicale permet donc d'identifier le contexte et l'orientation générale des messages concernés : la science nationale y est exaltée par les internautes de RPC qui se revendiquent comme d'authentiques représentants du peuple chinois. À l'inverse, le pronom de la deuxième personne du pluriel, 你们 *nimen* (439 occ./456 occ. du total, spécificité +7), peut désigner : le gouvernement des USA, les Américains, voire le « parti à un centième de dollar » (美分党 *meifen dang*), soit le « parti à cinq maos » (五毛党 *wumao dang*)¹⁵. Enfin, la forme 老子 *laozi* employée en tant que pronom (moi, homme vénérable) trahit chez les internautes qui l'emploient une certaine condescendance pour leurs adversaires.

¹⁴ On rend souvent le sens en français par « un peu ». De nombreux verbes monosyllabiques sont empruntés au chinois classique, les verbes dissyllabiques sont majoritaires dans le chinois moderne. Tous les verbes dissyllabiques n'ont pas de forme redoublée.

¹⁵ Le terme 五毛党 *wumao dang* ("le parti à cinquante centimes", dixième de Yuan, unité monétaire de la RPC), a été créé par les internautes chinois avec un sens péjoratif de : "parti des gens stipendiés par le pouvoir". De la même manière, le terme 美分党 *meifen dang* ("le parti à un centième de dollar") a été créé par d'autres internautes pour désigner les commentateurs stipendiés qui émettent des messages trop favorables, selon eux, au gouvernement américain.

3) les adverbes

En chinois comme en français, l'adverbe module le sens du verbe, de l'adjectif ou d'un autre adverbe. Mais, à la différence de celles du français, les catégories grammaticales chinoises ne sont pas fixes : un adjectif ou un verbe peuvent fréquemment être employés sous forme adverbiale. Ainsi le mot 老 *lao* qui, employé comme adjectif, signifie « âgé », veut dire « toujours » lorsqu'il est employé comme adverbe ; 快 *kuai*, qui, comme adjectif signifie « rapide », peut également être employé comme adverbe au sens de « vite ». Du côté des verbes, 越 *yue* qui, au sens primitif, signifie « dépasser » ou « franchir », s'emploie comme adverbe pour marquer une progression sous la forme redoublée 越 ... 越 ... *yue ... yue ...* (plus ... plus...) ou 越来越 ... *yue lai yue ...* (de plus en plus) ; le verbe 赶紧 *ganjin*, qui signifie « se dépêcher » ou « se presser », prend le sens d'« en toute hâte » lorsqu'il est employé comme adverbe. On trouve l'adverbe 就 *jiu* dans toutes sortes de propositions exprimant une relation de cause à effet, où il est en relation avec des conjonctions impliquant une condition comme 如果 *ruguo* (si), 假如 *jiaru* (au cas où). On le traduit par « alors », « justement », « en effet ». Mais les propositions contenant le sinogramme 就 *jiu* ont souvent un sens dépréciatif, on le traduit alors par « seulement », « ne ... que », « sauf que », « quand bien même ». Ces emplois spécifiques des adverbes par les internautes soulignent la rapidité et l'intensité de l'action, ils dénoncent les excès ou les défauts chez l'adversaire et constituent de puissants marqueurs de subjectivité : les internautes qui les utilisent expriment clairement leur volonté polémique dans des formules qu'ils veulent saisissantes.

6. L'étranger ennemi, l'étranger ami

Le conflit entre la Chine et *Google* oppose deux mondes, économiquement et culturellement différents. Il est tout à la fois : un conflit commercial, idéologique, stratégique et géopolitique. Il prend parfois même une dimension ethnique, comme en témoigne le champ lexical de *l'étranger*, représenté dans notre corpus par de nombreux termes aux connotations péjoratives ou xénophobes. On note qu'à l'exception de 外国人 *waiguoren* et de 洋人 *yangren*, qui constituent des formes neutres pour désigner les étrangers, la majorité des formes spécifiques des médias de la RPC ont une signification péjorative. Les mots 鬼子 *guizi* (démon), 洋鬼子 *yang guizi* (diable étranger), que l'on ne trouve que dans les médias publiés en RPC, sont employés de façon méprisante. De même que le suffixe 子 *zi* dans 鬼子 *guizi* (démon), le suffixe 佬 *lao* sert à former des unités lexicales péjoratives, telles que 美国佬 *meiguolao* (les sales Américains), 外国佬 *waiguolao* (une espèce d'étranger). Le mot 老外 *laowai*, qui avait lui aussi une connotation péjorative, s'est progressivement banalisé au fil du temps et tend aujourd'hui vers la neutralité (Yuan, 1998)¹⁶. Il garde cependant une acception méprisante dans le contexte conflictuel qui caractérise notre corpus. L'expression 大鼻子 *da bizi* (grand nez) employée par les internautes qui stigmatisent la physionomie des Occidentaux, est notoirement péjorative.

¹⁶ Au terme d'une longue évolution sémantique, le terme 老外 *laowai* n'a plus de connotation péjorative. Il désigne, au sens neutre, toute personne venant de l'extérieur, mais il a pu prendre, dans des contextes culturels et psychologiques variant au gré des circonstances, des acceptions plus ou moins méprisantes, ce qui peut encore entraîner des malentendus entre le locuteur et son allocutaire (Liu et Zhang, 2010 ; Zhang et Chen, 2008).

7. Conclusions et perspectives

Dans ce qui précède, nous avons tenté d'effectuer une exploration automatisée des opinions véhiculées par notre corpus à l'aide de l'approche textométrique. Notre étude nous a permis de distinguer les positionnements des médias chinois dans le monde sinophone, d'une part, selon les différences entre deux types discursifs – les articles de presse et leurs commentaires –, d'autre part, selon les lieux d'édition. Nous avons pu constater une nette opposition des discours produits en RPC avec ceux produits dans les autres pays sinophones. Notre analyse montre également que la puissance de communication des médias de RPC repose, de façon moins évidente mais non négligeable, sur la proximité plus ou moins grande de l'éditeur du message avec les groupes dominants dans tel ou tel champ politique. L'analyse du vocabulaire spécifique nous a permis de localiser, dans le type *article*, des indices d'énonciation qui permettent de mettre les énoncés en relation avec leur situation d'énonciation, d'actualiser leurs contextes et de dégager les opinions véhiculées par les textes. De même, dans le type *commentaire*, nous avons pu dégager, au plan lexical, des utilisations que l'on peut rapprocher des sentiments, des émotions, voire des orientations idéologiques des auteurs de RPC à leurs traits distinctifs formels. Le retour tendanciel au monosyllabisme est l'une des caractéristiques majeures de la langue utilisée par les internautes dans le cadre du conflit entre la Chine et la société *Google*. L'abandon des formes euphémiques redoublées, plus fréquentes dans la langue ordinaire, trahit une attitude plus intransigeante, voire ouvertement hostile, à l'égard de l'entreprise américaine sommée de déguerpir au plus vite. L'emploi adverbial d'un certain nombre d'adjectifs et de verbes va dans le même sens. Un nationalisme exacerbé – sans doute mâtiné d'un sincère patriotisme – se manifeste même à l'égard des Américains et, plus généralement, des Occidentaux, qu'on dénonce comme *étrangers* dans un vocabulaire dénué de toute aménité.

Après cette analyse du vocabulaire spécifique des discours et des médias du monde sinophone concernant l'*affaire Google*, nous envisageons d'approfondir notre étude en comparant les discours des internautes chinois de RPC et ceux des internautes des autres pays sinophones. Les données chiffrées précédentes faisant ressortir un décalage non négligeable entre les uns et les autres, nous pourrions alors envisager une étude interne des commentaires publiés par les internautes de la RPC. Cela nous permettra de mettre en évidence non seulement les singularités géographiques, mais également les spécificités observables dans les productions textuelles des communautés virtuelles en RPC.

Bibliographie

- Cao X. 曹小为. (2005). 中西媒体与政府的关系简述 (Précis des relations entre les médias chinois et occidentaux et leurs gouvernements). *新闻前哨 News Frontier*, pp. 33-34.
- Clérot F. (2010). Une ambition mondiale. *Monde chinois*, (24).
- He Y. 何永春. (2010). 试析网络词语的单音节趋向 (Analyse de la tendance au monosyllabisme sur Internet). *Master 大家*, (3), pp. 21-23.
- Kerbrat-Orecchioni C. (2002). *L'énonciation: de la subjectivité dans le langage* (4ème édition.). Paris: Armand Colin.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, (1), pp. 127-165.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève: Slatkine-Champion.
- Lafon P. et Salem A. (1983). L'inventaire des segments répétés d'un texte. *Mots*, 6(1), pp. 161-177.
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris: Duond.

- Li S. 李少丹 (2006). 谈网络语言的变异现象 (On the variation of the on-line language). *四川理工学院学报 (社会科学版) Journal of Sichuan University of Science & Engineering (Social Sciences Edition)*, (4), pp. 102-103 et 121.
- Liu W. 刘伟 et Zhang W. 张文华 (2010). 老外称呼语的语义色彩研究 (Étude sémantique des termes d'adresse à destination des étrangers). *湖北第二师范学院学报 Journal of Hubei university of education*, 27(1), pp. 25-28.
- Panckhurst R. (1997). La communication médiatisée par ordinateur ou la communication médiée par ordinateur ? *Terminologies nouvelles*, (17), 56-58.
- Paris M.-C. (2007). Un aperçu de la reduplication nominale et verbale en mandarin. *Faits de langues*, (29), pp. 63-76.
- Salem A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, 1(2), pp. 5-28.
- Salem A. (1987). *Pratique des segments répétés - Essai de statistique textuelle* (Vol. 17). Klincksieck.
- Wu L.-C. (2010). Outils de segmentation du chinois et textométrie. In *Actes de TALN 2010 (Traitement automatique des langues naturelles)*. Montréal, Canada.
- Yuan H. 元华 (1998). 对一个称呼语的跨文化剖析一释 « 老外 » (Analyse interculturelle d'un terme d'adresse - Laowai). *北京师范大学学报(社会科学版) Journal of Beijing Normal University (Social Science Edition)*, (6), pp. 87-93.
- Zhang J. 张积家 et Chen X. 陈秀芹 (2008). « 老外 » 究竟意味着什么 ? (What does Laowai really mean ?). *语言文字应用 Applied Linguistics*, (1), pp. 58-66.
- Zheng Y. 郑颖琦 (2009). 当代流行的单音节词语初探 (Étude des termes monosyllabiques en vogue à notre époque). *康定民族师范高等专科学校学报 Journal of Kangding Nationality Teachers College*, (4), pp. 35-38.

