

# *Des clés pour comprendre ...*

Analyses textométriques de textes chinois  
à partir de décomptes portant sur les *clés* des sinogrammes

André Salem<sup>1</sup>, Jun Miao<sup>2</sup>

<sup>1</sup> Université de la Sorbonne nouvelle - Paris 3, France – salem@msh-paris.fr

<sup>2</sup> Université Sun Yat-sen, Chine – miaoj5@mail.sysu.edu.cn

## **Abstract**

The methods of chronological textometrics, applied to a corpus of Chinese political texts (congresses of the Communist Party of China CPC, 1982-2017), highlight a strong renewal of vocabulary used over the years. The method presented here improves the readability of the results by relying on a peculiarity of the Chinese writing system: each character of the text can be attached to a *key* carrying a more general meaning.

The use of *semio-chronological cohorts*, lists of forms based on the double criterion of a) a homogeneous statistical distribution over time, b) the connection of the selected forms to the same key, provides powerful syntheses on the evolution of vocabulary over time. The proposed method could be applied without difficulty to any corpus of Chinese texts.

**Keywords:** chronological textual series, analysis of Chinese political discourse, semio-chronological cohorts.

## **Résumé**

Les méthodes de la textométrie chronologique appliquées à un corpus de textes politiques chinois (congrès du Parti communiste chinois PCC, 1982-2017) permettent de mettre en évidence un fort renouvellement du vocabulaire au fil des années. La méthode présentée ici permet d'améliorer la lisibilité des résultats en s'appuyant sur une particularité du système d'écriture chinois : chaque caractère du texte peut être rattaché à une *clé* porteuse d'un sens plus général.

L'utilisation *des cohortes sémio-chronologiques*, listes de formes constituées sur le double critère a) d'une répartition statistique homogène au regard du temps, b) du rattachement des formes à une même clé, fournit des résultats propices à l'élaboration de synthèses sur l'évolution du vocabulaire dans le temps. La méthode proposée devrait s'appliquer sans difficulté à tout corpus de textes chinois.

**Mots clés :** séries textuelles chronologiques, analyse du discours politique chinois, cohortes sémio-chronologiques.

## **1. Introduction**

Dans la plupart des langues occidentales, le système d'écriture employé pour la conservation des textes repose à la fois sur l'existence d'un alphabet restreint et sur la possibilité d'un découpage de la séquence textuelle en une suite de mots graphiques, séparés par des caractères "espace". Ces caractéristiques constituent un appui naturel pour la définition d'un système d'unités de comptage que l'on soumettra ensuite aux analyses textométriques dans le but d'explorer les corpus de textes de manière automatisée.

Le système d'écriture utilisé dans le monde sinophone possède des caractéristiques différentes. Pour soumettre un corpus de textes chinois à des comparaisons textométriques, il est nécessaire de soumettre les textes qu'il contient à une segmentation préalable en mots-graphiques. Par ailleurs, les propriétés graphiques des sinogrammes permettent d'envisager des procédures d'interprétation qui s'appuient directement sur des éléments de sens portés par chaque caractère. C'est à l'utilisation de cette seconde particularité du système d'écriture chinois pour les études textométriques que sera consacrée notre étude.

La section §1 présente quelques rappels sur les particularités du système d'écriture chinois et sur les conséquences qu'elles entraînent lors des traitements textométriques. On commente ensuite, section §2, des résultats obtenus précédemment sur une série chronologique de textes socio-politiques chinois qui constitueront un point de référence pour juger des améliorations méthodologiques proposées dans la présente étude. La section §3 est consacrée à la définition du concept de *cohorte sémio-chronologique* qui doit permettre, selon nous, une exploration plus efficace des corpus de textes chinois. Enfin, la section §4 décrit deux exemples d'application concrète de cette approche au corpus qui sert de support à l'étude.

## 2. L'approche textométrique des textes chinois

Les chercheurs qui entreprennent d'utiliser les méthodes de la textométrie pour explorer des corpus de textes produits dans le monde sinophone se trouvent d'emblée confrontés à deux particularités du système d'écriture chinois :

- la première de ces particularités réside dans le fait que les textes se présentent sous forme de longues suites de caractères chinois (ou sinogrammes) dans lesquelles les "mots" ne sont pas systématiquement séparés par des espaces, comme c'est le cas dans d'autres systèmes d'écriture;
- la seconde tient au fait que ces caractères sont souvent intrinsèquement porteurs de traits sémantiques qu'il est possible de repérer de manière semi-automatisée.

Depuis les années 90, des segmenteurs automatiques de texte, élaborés par des équipes de lexicologues chinois, peuvent être utilisés pour découper la séquence textuelle en suites de mots graphiques, à partir de dictionnaires préétablis<sup>1</sup>. De manière similaire à ce qui se fait couramment pour le traitement textométrique des textes occidentaux, ces procédures permettent de construire des décomptes statistiques croisant formes lexicales et parties du corpus qui serviront lors des analyses textométriques. L'ensemble de ces possibilités se trouvent désormais largement utilisées dans les études textométriques consacrées aux corpus de textes chinois.

La possibilité d'établir un lien entre la forme graphique des caractères utilisés dans un corpus de textes chinois et des éléments de sens plus généraux peut être considérée comme un avantage important du système d'écriture chinois pour l'étude textométrique des textes, notamment lors de la phase d'interprétation des résultats issus des procédures statistiques. C'est à l'utilisation de cette seconde particularité du système d'écriture chinois que nous proposons de consacrer la présente étude.

### *Clés des caractères*

L'ordre retenu pour classer les mots de la langue dans les dictionnaires chinois résulte d'une méthode de classement complexe qui permet des regroupements de mots que l'on peut rattacher à une même famille sémantique. Cette classification s'appuie sur le repérage d'une *clé* qui possède un lien avec la signification du premier caractère du mot (Cui, 1997, 2008 ; Tang, 2005). Les manuels d'enseignement du chinois expliquent, de manière simplifiée, que la plupart des caractères chinois peuvent être décomposés en deux parties<sup>2</sup> : a) un premier élément qui constitue la *clé* du caractère et qui concerne prioritairement son *sens*; b) un

---

<sup>1</sup> Pour la présente étude, nous avons utilisé le logiciel NLPir-ICTCLAS (<http://ictclas.nlpir.org/>), conçu par Huaping Zhang.

<sup>2</sup> Pour Ramsey (1987 :136-137) Wieger (e.g., p.14-15), le *radical* est un élément graphique d'un sinogramme porteur de sens. D'autres chercheurs, préfèrent utiliser le concept de *clé*.

second élément qui renseigne le lecteur sur la manière dont le caractère doit être prononcé. Un exemple simple est celui du caractère 抓 (zhuā, *saisir*). Dans ce caractère, la *clé*, à gauche (扌) constitue une indication sémantique (l'idée de *main*); le composant de droite : 爪 (zhuǎ, *patte*), fournit une indication qui concerne plutôt la prononciation du caractère.

Dans la pratique, il arrive que cette décomposition ne soit pas aussi simple à réaliser<sup>3</sup>. Cependant, le fait important pour notre étude est que des dictionnaires, largement reçus dans la communauté sinophone, permettent de rattacher chaque caractère chinois à une *clé* ou *radical*, choisi parmi un ensemble qui compte *canoniquement* 214 éléments. Ces clés jouent un rôle important dans la lexicographie chinoise ; elles permettent de classer les mots au sein des dictionnaires de langue. Notons que la recherche de l'étymologie exacte de chaque caractère constitue un travail d'érudit (cf. Gu, 2003 ; Yan, 2010 ; Yu, 2013) qui sort largement du cadre de notre étude. Dans ce qui suit, nous utiliserons le système des radicaux parce qu'il possède une propriété essentielle pour la mise en œuvre des méthodes d'exploration textométrique. La présente étude constitue une tentative de mobiliser les ressources offertes par la prise en compte des traits sémantiques que l'on peut rattacher aux caractères-clés afin de mieux caractériser les évolutions lexicales au sein d'un corpus de textes politiques.

### 3. Analyses sur la série chronologique PCC1982-2017

Nous illustrerons la démarche que nous proposons de suivre en empruntant nos exemples à un corpus de textes qui a déjà fait l'objet de recherches sur l'évolution des formes du vocabulaire découpées par les segmenteurs au sein d'un corpus de textes (Miao et Salem, 2018 et 2019). Ce corpus, que nous appelons PCC1982-2017, est constitué par la réunion des rapports d'ouverture présentés durant les 8 congrès du Parti du communiste chinois (PCC) qui se sont tenus tous les cinq ans, entre 1982 et 2017<sup>4</sup>.

#### 3.1 Typologie sur les congrès

Comme c'est souvent le cas lorsqu'on analyse une série textuelle chronologique<sup>5</sup>, la typologie que l'on obtient, en soumettant à l'analyse des correspondances (AFC) le tableau qui rassemble les décomptes des occurrences de chacune des formes graphiques dans chacune des 8 parties du corpus montre que les congrès se placent sur le premier facteur en fonction de la date à laquelle ils se sont tenus. On en conclut que, dans le cas qui nous occupe, l'évolution du vocabulaire employé au fil des congrès constitue, au plan quantitatif, la caractéristique fondamentale de la série étudiée.<sup>6</sup>

Le schéma intuitif qui préside à ce type d'exploration peut être formulé comme suit : au fil des congrès, les modifications qui surviennent tout à la fois dans le contexte socio-économique, dans sa perception par les rédacteurs des textes de congrès et dans la formulation des objectifs que ces derniers proposent pour la période à venir, entraînent des modifications constantes dans l'utilisation des formes linguistiques utilisées par les rédacteurs des textes. Ces variations

---

<sup>3</sup> La place de la clé au sein du caractère peut varier (droite, gauche, haut, bas, etc.). Certains caractères, sont constitués de leur seule clé (*monogrammes*).

<sup>4</sup> Les congrès du PCC se sont tenus en : 1982, 1987, 1992, 1999, 2002, 2007, 2012, 2017.

<sup>5</sup> (Salem et Miao, HAL 2019), article à paraître dans un numéro spécial de la Revue *Histoire & mesure*, consacré aux séries textuelles chronologiques, version provisoire : <https://hal.archives-ouvertes.fr/hal-02119927/>

<sup>6</sup> Sur les séries textuelles chronologiques, on peut également consulter (Salem 1991).

des conditions de production du texte influent fortement sur les fréquences avec lesquelles sont utilisées les différentes formes de vocabulaire.<sup>7</sup>

Les méthodes de la textométrie chronologique permettent de mettre en évidence celles des unités de dépouillement utilisées (formes graphiques, par exemple) qui sont les plus fortement responsables de la variation chronologique d'ensemble. Cependant, l'analyse des variations de fréquences, réalisée à partir du dépouillement en formes graphiques, montre que les changements que l'on peut observer ne se réduisent pas à de simples substitutions progressives de lexèmes dotés chacun d'un fort contenu sémantique. Ces modifications concernent, au contraire, tous les niveaux de la langue qui retiennent traditionnellement l'attention des analystes. Ainsi, par exemple, la diminution progressive des occurrences de la forme 我们 (*wǒmen*, *nous*) doit être mise en rapport avec l'augmentation parallèle des occurrences du nom propre 中国 (*zhōngguó*, *Chine*) (Salem et Miao, 2019). Cette transformation du texte peut être analysée en termes d'énonciation : au fur et à mesure que l'on progresse dans le temps, le *pays* tout entier tend à remplacer, dans le rôle d'énonciateur principal, un *nous* moins bien défini, plus fréquemment employé lors des premières périodes.

Ces exemples soulignent la nécessité de dépasser le cadre de l'étude purement lexicale des variations qui surviennent au fil des textes pour tenter d'élargir nos systèmes d'observations à des unités textuelles plus complexes.

### 3.2 Différentes unités de dépouillement

A partir du corpus qui sert de base à nos expériences, nous avons effectué des décomptes qui portent sur différents types d'unités que l'on peut recenser dans la séquence textuelle. Ces décomptes ont été rassemblés au tableau 1. La colonne *formes* rappelle les caractéristiques des décomptes en formes graphiques que nous avons analysés dans nos études précédentes ; la colonne *caractères* présente les mêmes décomptes réalisés à partir des sinogrammes.

**Tableau 1** : Dépouillements textométriques en *formes graphiques*, en *caractères* et en *clés*, à partir du corpus PCC1982-2017

	<i>formes</i>	<i>caractères</i>	<i>clés</i>
<i>nombre d'occurrences</i>	115 138	216 175	216 175
<i>nombre de formes</i>	7 365	1 938	261 <sup>8</sup>
<i>nombre d'hapax</i>	2 776	321	13
<i>fréquence maximale</i>	7 664	7 838	15 400

On note que les occurrences des *caractères* sont plus nombreuses que celles des *formes* mais que leur variété est plus réduite. On trouve, dans la colonne *clés*, les décomptes qui correspondent à un dépouillement dans lequel chaque caractère du texte a été remplacé par la *clé* à laquelle on peut le rattacher (ex : 执 *exécuter* a été remplacé par sa clé 手 *main*).

A l'instar de ce que nous avons réalisé sur la base des décomptes en formes graphiques, nous avons construit des tableaux qui rassemblent les décomptes de chacune de ces unités dans chacun des congrès :

<sup>7</sup> Il va sans dire que ces variations, constatées lors des analyses quantitatives réalisées *a posteriori*, n'impliquent pas nécessairement une volonté consciente de la part des rédacteurs du texte.

<sup>8</sup> Ce nombre rassemble les caractères clés, rencontrés dans le corpus ainsi qu'un certain nombre de caractères "monogrammes" du corpus (i.e. qui ne relèvent d'aucune clé et que nous avons codés par eux-mêmes).

- le tableau *Congrès X Caractères* dénombre les occurrences de chacun des caractères dont la fréquence dépasse 10 occurrences dans chacune des 8 parties du corpus.
- le tableau *Congrès X Clés* recense les occurrences des clés de fréquence supérieure à 10 dans le même corpus, avec la même partition en 8 congrès.

Les tableaux ainsi constitués ont été soumis aux mêmes procédures statistiques que le tableau des *Congrès X formes graphiques* dont nous avons évoqué l'analyse au paragraphe précédent. Comme on peut le voir sur la figure 1, la typologie obtenue sur les congrès à partir du décompte en *caractères* (en haut à droite), ressemble assez fortement à celle que nous avons obtenue à partir des décomptes en *formes graphiques* (en haut à gauche). Cette circonstance renforce notre conviction que l'évolution des textes réunis dans le corpus ne peut être ramenée à une simple évolution au plan du lexique, même si l'évolution du vocabulaire entraîne des conséquences sur la fréquence de chacun des caractères.

Pour la troisième analyse (*clés*), nous avons remplacé chacun des caractères du corpus par la *clé* à laquelle on peut le rattacher<sup>9</sup>. Le tableau issu du décompte de chacune des *clés* dans les parties du corpus a ensuite été soumis au même type d'analyse (AFC). Comme sur les représentations établies à partir des décomptes de formes graphiques (ou de caractères), les congrès s'alignent toujours, du premier au dernier, sous forme d'une parabole qui résulte de l'évolution progressive des unités de décomptes au fil du temps.

La conclusion qui se dégage de l'ensemble de ces expériences est que l'évolution des discours, au fil du temps, inscrit très profondément sa marque dans l'ensemble des unités qui servent à construire le texte. A mesure que les années passent, l'évolution des contextes d'énonciation propres à chacun des congrès entraîne des changements profonds qui touchent à la fois le lexique et les différentes formes de l'énonciation. Ces changements s'inscrivent profondément dans les textes produits. Au delà du seul plan lexical, ils peuvent être repérés et étudiés à partir des différents types d'unités statistiques qui entrent dans la composition du texte (*formes graphiques, caractères, clés, etc.*).

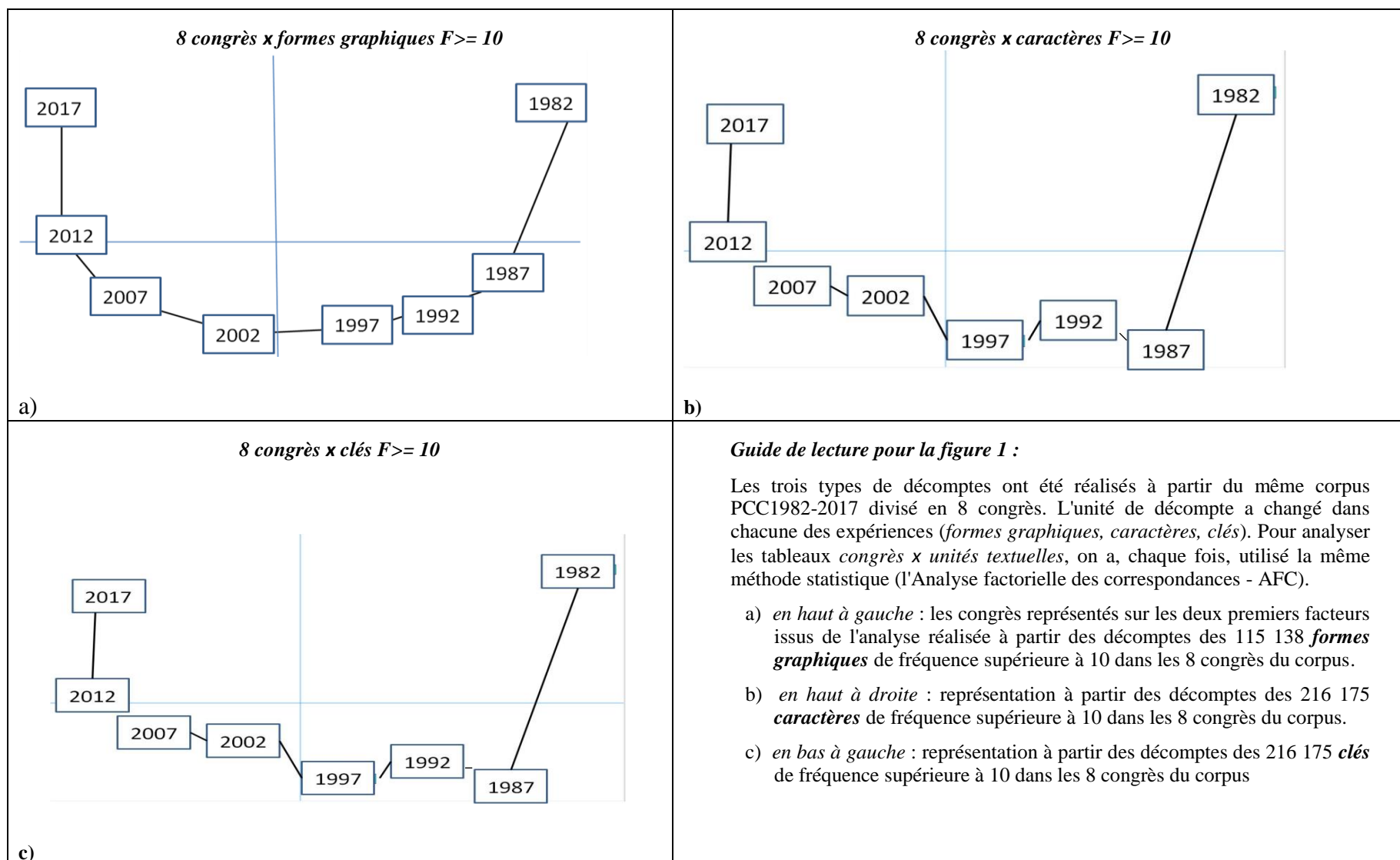
### 3.3 Analyses à partir des clés

Comme nous l'avons expliqué plus haut, une même clé peut concerner des formes et des caractères dont le sens est fortement éloigné. La répartition de certaines clés confirme certains constats que nous avons pu faire à partir des formes graphiques. La forte présence de la clé 我 (*je*), dans la première partie des congrès, s'explique par le nombre élevé des occurrences de la forme 我们 (*nous*) dans cette même partie. Cependant, l'expérience montre que la seule présence d'une clé ne permet pas de constituer des classes sémantiques homogènes dont l'étude des variations de fréquence conduirait à éclairer valablement les changements qui surviennent au fil des textes. C'est pour ces raisons que nous avons choisi d'introduire les *cohortes sémio-chronologiques* qui seront décrites au point suivant.

On trouvera en annexe, les principales spécificités chronologiques calculées à partir des clés sur le corpus PCC1982-2017.

---

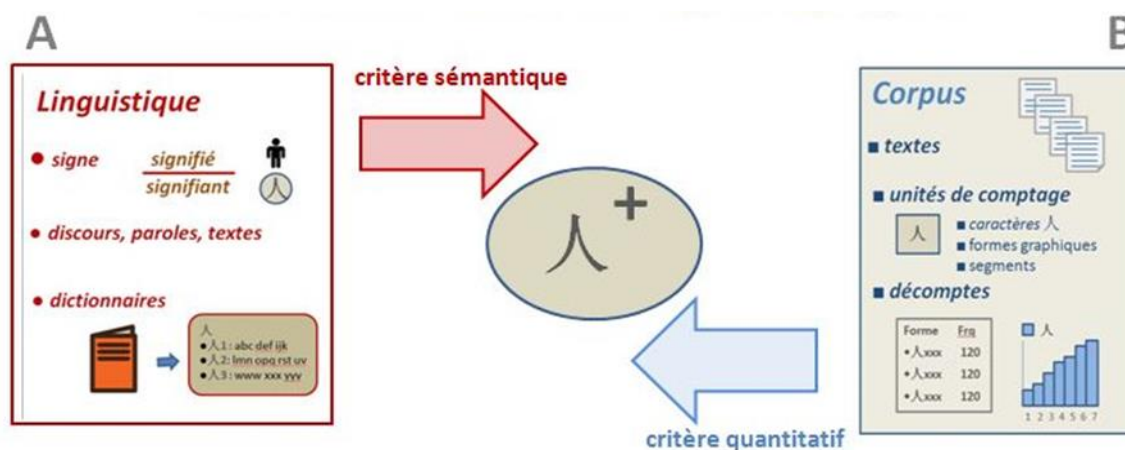
<sup>9</sup> Ce dictionnaire peut être téléchargé sur le site : <http://www.lexi-co.com/ressources/clePCC>. Pour confectionner ce dictionnaire, nous avons commencé par trier l'inventaire typographique du corpus par ordre alphabétique, afin de rapprocher automatiquement les caractères qui se rattachent à une même clé. Nous avons ensuite vérifié les rattachements les plus complexes dans le dictionnaire chinois en ligne Xinhua (<http://xh.5156edu.com/>). A l'issue de ce processus, chacun des caractères du texte initial a ensuite été remplacé par sa clé..



**Figure 1** : Comparaison des analyses réalisées à partir de différentes unités de dépouillement du texte

## 4. Les cohortes sémio-spécifiques

Dans ce qui suit, nous allons décrire des unités textuelles d'un genre nouveau : *les cohortes sémio-spécifiques*. L'apport que nous espérons obtenir de cette approche concerne le langage de description des évolutions textuelles. Dans l'expérience réalisée à partir des décomptes de formes graphiques, nous avons illustré les transformations qui s'opéraient dans les textes, au fil du temps, par des variations de la fréquence de certaines des formes graphiques. Désormais, il va être possible de décrire ces transformations en termes de variation de la fréquence d'ensembles de formes présentant des traits communs, à la fois au plan sémantique et au plan de la variation de leur fréquence au fil des textes.



Une *cohorte sémio-spécifique* est un sous-ensemble de formes du corpus :

- A - contenant un **marqueur sémantique** commun
- B - respectant un **schéma d'évolution chronologique**

Figure 2 : Cohortes sémio-spécifiques

Le schéma présenté sur la figure 2 représente, de manière synthétique, la notion de *clé sémio-chronologique*. Sur la partie gauche on a représenté l'approche linguistique traditionnelle avec ses oppositions fondatrices : *signifié/signifiant, langue/parole*. Dans ce cadre général d'analyse des textes, l'explication du sens des unités linguistiques est recherchée à la fois dans la compétence commune que partagent les locuteurs et dans les dictionnaires de langue qui tentent d'en rendre compte. Dans le cas particulier du système d'écriture chinois, un caractère comme 人 (rén), qui signifie *homme*, véhicule en outre des éléments graphiques dont le sens est lié au signifié (deux jambes ou un homme en marche). Cependant, ce caractère peut véhiculer toute une série de sens distincts lorsqu'il apparaît en combinaison avec d'autres caractères : 人生 (rénsǎng, *vie*), 人民 (rénmín, *peuple*); il peut également entrer dans la composition d'un caractère plus complexe 伴(bàn, *accompagner*), sous forme d'une clé (扌), dans laquelle les lecteurs chinois reconnaissent sans difficulté une altération du caractère 人.

La partie droite de la même figure symbolise les approches quantitatives basées sur la comparaison de plusieurs textes réunis en corpus. Au sein d'un corpus ne sont attestés que certains des sens répertoriés dans les dictionnaires de langue. Par ailleurs, une fois le corpus délimité, chaque forme graphique est affectée d'une fréquence dont on peut observer les variations au fil des parties.

Les *cohortes sémio-chronologiques* sont des ensembles composés de formes qui possèdent en commun deux propriétés :

- ces formes partagent une propriété sémantique, comme celle de contenir une clé identique ;
- elles possèdent également des caractéristiques de répartition communes : formes dont la fréquence tend à augmenter au fil des congrès (ou au contraire : formes dont la fréquence tend à diminuer, voire : formes fréquentes dans un empan du corpus).

## 5. Deux exemples d'application

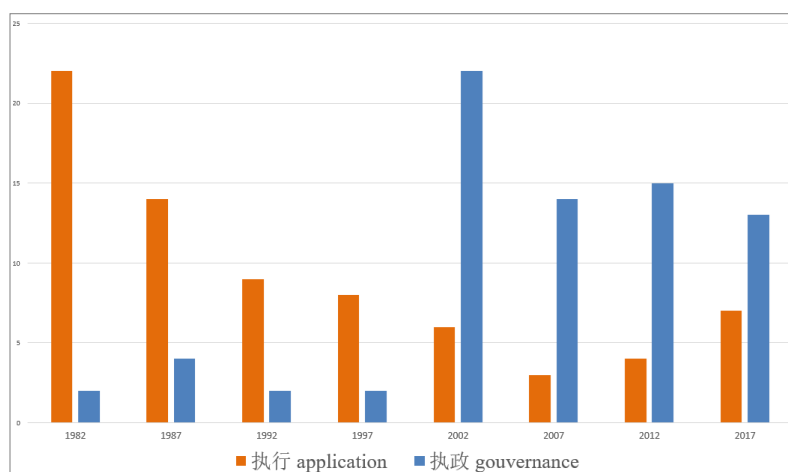
Les deux exemples qui suivent illustrent les possibilités d'exploration des résultats fournis par les méthodes textométriques en utilisant des cohortes sémio-chronologiques.

### 5.1 Le cas de la clé de la main (手, shǒu, main)

Les dictionnaires de langue indiquent que la clé 手/shǒu (shǒu, *main*) renvoie à la notion générale d'*action* (Gu, 2003). Les formes qui viennent naturellement à l'esprit lorsqu'on considère cette *clé-caractère* renvoient à des actions que l'on réalise à l'aide de la main : 推 (tuī, *pousser*), 打(dǎ, *frapper*), 提(tí, *soulever*), 搬(bān, *déplacer*), 投(tóu, *jeter*), etc.

Dans notre corpus PCC1982-2017, la clé de la *main* 手/shǒu compte 4 641 occurrences. Cette clé entre dans la composition de 106 caractères différents : 扑 *jeter*, 打 *frapper*, 托 *soutenir*, 扣 *garder*, 执 *exécuter*, 扩 *élargir*, par exemple. Ces caractères participent eux-mêmes à la formation de 480 formes graphiques dont les plus fréquentes sont : 坚持 (jiānchí, *insister*, 628 occ.), 提高 (tígāo, *améliorer*, 385 occ.), 推进 (tuījìn, *avancer*, 337 occ.), 把 (bǎ, BA<sup>10</sup>, 329 occ.), 发挥 (fāhuī, *déployer*, 178 occ.), etc..

L'éventail des divers sens qu'un locuteur chinois peut envisager à partir de sa compétence de la langue, lorsqu'il considère une forme linguistique hors d'un contexte textuel particulier, se réduit et se précise dès lors qu'il ne considère plus que les seuls emplois effectifs de cette forme au sein du corpus qu'il étudie.



**Figure 3 :** Distribution des formes 执行 (zhíxíng, *application*) et 执政 (zhízhèng, *gouvernance*)

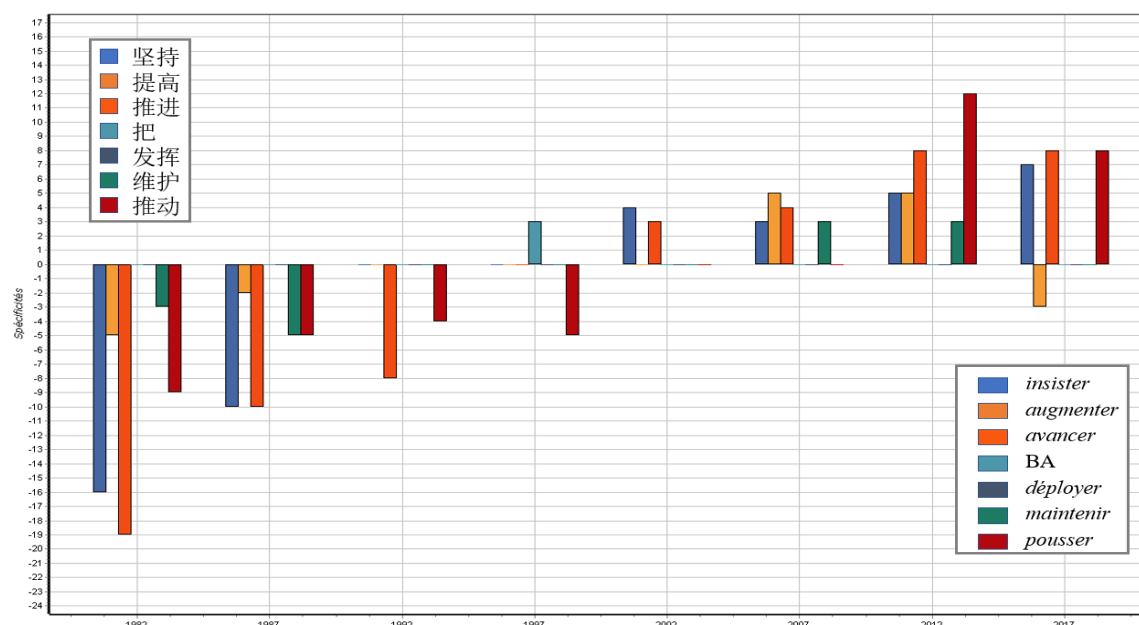
<sup>10</sup> La structure 把 (bǎ) est utilisée en chinois pour accentuer sur le résultat ou l'influence d'une action.



dans le corpus PCC1982-2017

Parmi les formes du corpus PCC1982-2017 qui contiennent la clé de la *main*, certaines voient leur fréquence augmenter fortement entre le début et la fin de la période. Pour d'autres, c'est le contraire qui se produit.

La figure 3 montre, à titre d'exemple, la distribution des formes 执行(zhíxíng, *exécution*, 73 occ.) et 执政(zhízhèng, *gouvernance*, 74 occ). Les emplois de la première diminuent dans la décennie 2002-2012 avant de croître à nouveau en 2017. La deuxième forme, peu employée au début, devient fréquente en 2002 et le restera jusqu'à la fin de la période couverte par le corpus. Comme on le voit sur cet exemple, une même *clé* peut rassembler des unités lexicales dont la variation est opposée. Le regroupement des unités relevant d'une même clé qui possèdent la caractéristique d'augmenter (resp. de diminuer) au fil du temps permet de constituer des cohortes plus aisément descriptibles. Sur la figure 4, nous pouvons voir que la spécificité de certaines formes graphiques qui contiennent la clé *main* augmente à mesure que l'on avance dans le temps, signe manifeste d'un emploi plus important dans les dernières années, en particulier à partir du congrès de 2017.



**Figure 4 :** Spécificités de quelques formes de la cohorte **main +** dans les congrès du corpus PC1982-2017 (dépouillement en formes graphiques)

La cohorte sémio-chronologique 手<sup>+</sup> (*main*<sup>+</sup>) rassemble les formes les plus fréquentes du corpus (fréquence supérieure à 50 occurrences) qui répondent aux deux critères mentionnés plus haut : ces formes contiennent des caractères qui sont régis par la clé *main* ; d'autre part, la fréquence de ces formes, mesurée ici en spécificité, augmente au fur et à mesure que l'on avance dans le temps. Les formes retenues sont : 坚持 *insister*, 推进 *avancer*, 保护 *protéger*, 执政 *gouvernance*, 推动 *pousser*, 维护 *maintenir*, 保持 *garder*, 持续 *continuer*, 科技 *technologie*, 扩大 *développer*, 提高 *améliorer*, 支持 *soutenir*, 把 BA (prep).

Au vu de cette liste, nous avancerons une piste d'interprétation de ce phénomène qu'il faudra ensuite vérifier en retournant soigneusement au contexte : les congrès qui ont lieu avant 2002 insistent sur *l'application* des décisions prises par l'instance dirigeante. A partir de 2002, l'accent est mis, au contraire, sur la nécessité de : *modifier*, *améliorer*, *développer*, *faire*

*avancer, insister sur* (les méthodes de direction de l'Etat et de gestion de l'économie). Le concept de *gouvernance* (治理, zhìlǐ), issu du vocabulaire de l'entreprise, fait irruption dans la sphère des formes qui manifestent l'idée de "direction".

Les brefs exemples contextuels qui suivent permettent d'illustrer cette évolution qui se développe dans le temps.

**Congrès 1987** #<para=13-147>党的建设问题，从来是同党的政治路线密切地联系在一起的。¶<phra=13-0601>新时期党的一切工作，都必须保证党的基本路线的贯彻**执行**。党的自身建设也必须进行改革，以适应改革开放的新形势。党的思想建设\*组织建设\*作风建设，都应当体现这个**指导**思想。

*La consolidation du Parti est, depuis toujours, étroitement liée à sa ligne politique. Durant cette nouvelle période, toutes les activités du Parti devront impérativement garantir l'application effective de sa ligne fondamentale. En outre, le Parti doit mener une réforme concernant sa propre édification pour s'adapter à la nouvelle situation caractérisée par la réforme et l'ouverture sur l'extérieur ; il lui faut appliquer ce principe **directeur** dans son édification sur les plans idéologique et organisationnel ainsi que dans l'amélioration de son style de travail.*

**Congrès 2017** #<para=19-083>实现"两个一百年"奋斗目标\*实现中华民族伟大复兴的中国梦，不断提高人民生活水平，必须坚定不移**把**发展作为党**执政**兴国的第一要务，**坚持**解放和发展社会生产力，**坚持**社会主义市场经济改革方向，**推动**经济**持续**健康发展。

*Pour atteindre les objectifs des " deux centennaires ", réaliser le rêve du grand renouveau national, et **élever** sans cesse le niveau de vie de la population, il faut considérer, sans équivoque, le développement comme la tâche primordiale du Parti dans **la gouvernance** de la nation en vue de son renouveau, **continuer à libérer** et à développer les forces productives sociales, **persévérer** dans la réforme visant le perfectionnement de l' économie de marché socialiste, et promouvoir un développement sain et durable de l' économie.*

## 5.2 Le cas de la clé de l'eau (水, shuǐ, eau)

Dans les textes chinois, l'idéogramme 水 (shuǐ, eau) renvoie principalement à l'idée de *liquide* (mer, rivières, écoulement, etc.). Par extension métaphorique, cet idéogramme est également utilisé pour former des unités de sens qui signifient: *le niveau, la régulation, la réglementation, les lois*. Les expressions chinoises qui correspondent à ces dernières notions comportent souvent la *clé* 氵 qui correspond au sinogramme 水 (shuǐ, eau). On note qu'un grand nombre des formes qui relèvent de cette clé voient leurs occurrences augmenter dans les cinq dernières périodes du corpus (1997-2017). La cohorte sémio-spécifique 水<sup>+</sup> (水 shuǐ, eau) rassemble les formes qui possèdent à la fois la propriété de contenir la *clé de l'eau* et celle de voir leurs occurrences augmenter au fil du temps. Si la forme 经济 (jīngjì, économie), qui contient également la clé de l'eau, est plus fréquente avant le congrès 2002, les termes qui véhiculent les sens de *loi* et de *réglementation*, sont davantage utilisés dans les derniers congrès : 法律 (fǎlǜ, loi), 依法 (yīfǎ, par la loi), 政治 (zhèngzhì, politique), 治理 (zhìlǐ, gouvernance).

Ici encore, nous avancerons des éléments d'interprétation : au fur et à mesure du développement économique du pays, les textes de congrès mettent l'accent sur des *méthodes de gouvernance* qui s'appuient de plus en plus sur la *législation*. A côté de ces formes qui ont trait à la régulation, les formes qui concernent la vie : 生活 (shēnghuó, vie), 活力 (huólì, vitalité), 活动 (huódòng, activité), 水平 (shuǐpíng, niveau) et 资源 (zīyuán, ressources), formes qui contiennent également la *clé de l'eau*, voient également leurs emplois croître dans les derniers congrès. Le niveau de vie, l'utilisation des ressources, l'environnement, deviennent, à l'évidence, des préoccupations centrales dans les dernières années couvertes par notre corpus.

L'extrait qui suit, emprunté au congrès de 2017, nous permet d'illustrer notre propos.

# 第一个阶段，从二〇二〇年到二〇三五年，在全面建成小康社会的基础上，再奋斗十五年，基本实现社会主义现代化。¶ 到那时，我国经济实力\*科技实力将大幅跃升，跻身创新型国家前列；¶ 人民平等参与\*平等发展权利得到充分保障，**法治**国家\***法治**政府\***法治**社会基本建成，各方面制度更加完善，国家**治理**体系和**治理**能力现代化基本实现；¶ 社会文明程度达到新的高度，国家文化软实力显著增强，中华文化影响更加**广泛深入**；¶ **人民生活**更为宽裕，中等收入群体比例明显提高，城乡区域发展差距和居民**生活水平**差距显著缩小，基本公共服务均等化基本实现，全体人民共同富裕迈出坚实步伐；¶ 现代社会**治理**格局基本形成，社会充满**活力**又和谐有序；¶ 生态环境根本好转，美丽中国目标基本实现。

*La première phase va de 2020 à 2035 ; phase pendant laquelle, partant de l'établissement de la société de moyenne aisance, nous poursuivrons nos efforts pour réaliser l'essentiel de la modernisation socialiste. La Chine verra alors sa puissance économique, scientifique et technologique accroître considérablement, et se hissera au premier rang des pays novateurs. Le droit du peuple à la participation et au développement sur un pied d'égalité se verra pleinement garanti ; l'édification d'un État, d'un gouvernement et d'une société qui **respectent la loi** sera parachevée pour l'essentiel ; les systèmes institutionnels seront améliorés dans divers domaines ; et la modernisation du système et de la capacité de **gouvernance** de l'État sera réalisée dans son ensemble. Le degré de civilité de notre société atteindra un niveau plus élevé, le "soft power" culturel s'accroîtra considérablement, et le rayonnement de la culture chinoise deviendra plus large et plus **profond**. La **vie** de la population sera plus aisée ; la proportion des personnes à revenu moyen accroîtra sensiblement ; les écarts de développement entre les régions, et entre les villes et campagnes, ainsi que les écarts du **niveau de vie** entre les habitants seront notablement réduits ; l'homogénéisation des services publics fondamentaux sera pratiquement réalisée ; notre marche vers l'enrichissement commun accomplira ainsi un pas substantiel. Une structure de **gouvernance** sociale moderne prendra forme, créant une société où règnent le **dynamisme**, l'harmonie et l'ordre. L'écosystème connaîtra une amélioration fondamentale, et la belle Chine deviendra réalité.*

## 6. Conclusion

A partir d'une série textuelle chronologique de textes chinois (PCC1982-2017) qui a déjà fait l'objet d'analyses effectuées sur la base de dépouillements en formes graphiques, nous avons tenté de mettre en place une méthodologie de recherche appuyée sur le dénombrement des *clés* auxquelles peuvent être rattachés chacun des caractères qui composent le texte.

Les typologies obtenues à partir des différents types de décomptes sur l'ensemble des textes se révèlent profondément cohérentes entre elles. Nous en concluons que l'évolution générale, au fil du temps, des textes que nous avons comparés peut être observée à des niveaux de dépouillement en unités statistiques très différentes (*formes graphiques, caractères, clés*).

Les *cohortes sémio-chronologiques* possèdent, par construction, des propriétés statistiques qui sont en liaison étroite avec l'évolution du discours au plan textométrique. Elles possèdent en outre des propriétés sémantiques, du fait de leur liaison avec une clé sémantique particulière. La prise en compte de ces cohortes facilite la production d'un commentaire qui doit cependant tenir compte des sens contextuels que les formes revêtent nécessairement dans chaque corpus particulier. De manière plus efficace que sur la base de textes découpés en formes graphiques, les analyses réalisées directement à partir des clés conduisent à des constats qui portent sur des unités auxquelles on peut rattacher des éléments de "sens". L'interprétation des résultats s'en trouve nettement facilitée.

Les analyses effectuées à partir des décomptes de clés ouvrent des perspectives pour l'interprétation des changements qui surviennent au fil des textes rassemblés dans une série textuelle chronologique. Au lieu des longues listes de formes dont il peut être parfois difficile de dégager les traits communs et les raisons qui expliquent leur accroissement simultané au fil des textes (ou leur diminution), les constats effectués à partir des *cohortes sémio-*

*chronologiques* nous renseignent de manière plus directe sur l'évolution d'unités statistico-linguistiques auxquelles il devient plus facile d'associer un sens.

La méthode proposée devrait s'appliquer, sans difficultés, à tout type de corpus rédigé en langue chinoise.

## Références

- Cui Y.H. (崔永华). (2008). 对外汉语教学设计导论 (*Introduction à la conception de l'enseignement du chinois comme langue étrangère*). Beijing: 北京语言大学出版社.
- Cui Y.H. (崔永华) ed. (1997). 汉字部件和汉字教学词汇文字研究与对外汉语教学 (*Etudes des clés et des caractères chinois dans l'enseignement du chinois et l'enseignement du chinois comme langue étrangère*). Beijing: 北京语言大学出版社.
- Gu Y.K. (谷衍奎). (2003). 《汉字源流字典》 (*Dictionnaire étymologique des caractères chinois*). Beijing: 华夏出版社.
- Miao J. et Salem A. (2019). Séries textuelles homogènes. In JADT'18 - *Proceedings of the 14th international conference on statistical analysis of textual data*, UniversItalia, Roma, 2018.
- Ramsey S.R. (1989). *The Languages of China*. Woodstock : Princeton University Press
- Salem A. (1991), Les séries textuelles chronologiques. *Histoire & mesure*, vol. (6) : 149-175.
- Salem A. et Miao J. (2019). *Le texte se transforme...* article à paraître dans un numéro spécial de la Revue *Histoire & mesure*, consacré aux séries textuelles chronologique, version provisoire (HAL 2019) : <https://hal.archives-ouvertes.fr/hal-02119927/>
- Tang L. (唐兰). (2005). 《中国文字学》 (*Écriture chinoise*). Shanghai: 上海古籍出版社
- Yan Y. (严焯). (2010). 浅析对外汉字教学之部件教学 (*Enseignement du chinois comme langue étrangère à l'aide des clés*). in 海外华文教育 (*Pédagogie du chinois à l'étranger*), vol. (4-57): 69-73
- Yu X.G. (裘锡圭). (2013). 《文字学概要》 (*Synthèse des études du caractère chinois*). Beijing: 华夏出版社.

## Annexe

Spécificités chronologiques des différentes *clés* dans le corpus PCC1982-2017

Clé	équiv. fr	F	f	S	Spéc. chronologique
我	<i>je</i>	1229	930	24	—
走	<i>marcher</i>	4618	1534	20	—
了	<i>LE</i>	613	467	39	—
白	<i>blanc</i>	7841	6042	51	—
开	<i>ouvrir</i>	779	672	20	—
才	<i>main</i>	5940	4698	31	—
水	<i>eau</i>	5715	1874	20	—
斤	<i>livre</i>	1243	446	26	—
弓	<i>arc</i>	1324	465	30	—
入	<i>entrer</i>	1854	647	42	—
民	<i>peuple</i>	2255	911	23	—
尸	<i>chair</i>	2648	1629	36	—
力	<i>force</i>	4641	2646	26	—
人	<i>homme</i>	16967	8967	21	—
色	<i>couleur</i>	401	264	31	—