

Y.-L. CHEUNG

S.-C. LEUNG

J.-P. BENZÉCRI

Essai de typologie des écritures manuscrites chinoises

Les cahiers de l'analyse des données, tome 13, n° 1 (1988),
p. 81-92

http://www.numdam.org/item?id=CAD_1988__13_1_81_0

© Les cahiers de l'analyse des données, Dunod, 1988, tous droits réservés.
L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ESSAI DE TYPOLOGIE DES ÉCRITURES MANUSCRITES CHINOISES

[TYP. CHIN.]

*Y.-L. CHEUNG**

*S.-C. LEUNG***

*J.-P. BENZÉCRI****

1. Le problème et les données

L'étude des écritures manuscrites est utile notamment dans les expertises judiciaires. Une grande expérience a été accumulée dans l'étude des écritures alphabétiques telles que l'écriture latine; mais pour une écriture idéogrammatique telle que celle du chinois, le problème est beaucoup plus complexe. Pourtant les caractères manuscrits se prêtent à des mesures précises et l'analyse statistique de ces mesures devrait aboutir à une typologie rigoureusement définie. Telle est la voie où nous nous engageons dans le présent travail.

1.1. Le matériel de base

Sans énumérer toutes les recherches en cours, nous nous bornerons à décrire les données utilisées ici. On a retenu un ensemble C de 25 caractères d'usage très courant et sur chacun desquels on peut effectuer une ou plusieurs mesures qui seront précisées ci-dessous. Un ensemble S de 21 sujets ont fourni un ensemble I de réalisations manuscrites de l'ensemble C. Afin que le tracé soit fait de façon naturelle, on demande aux sujets d'écrire un petit texte où figure une fois chacun des caractères de C. Chaque sujet ayant écrit ce texte (à différents jours) de 11 à 25 fois (en général 20), I compte au total 376

(*) Lecturer, City Polytechnic of Hong Kong.

(**) Senior Chemist, Forensic Science Division, Hong Kong Government Laboratory.

(***) Professeur, Université Pierre et Marie Curie.

réalisations de C. Il sera commode de désigner par $s(i)$ le sujet ayant produit la réalisation i ; et de noter I_s l'ensemble des réalisations dues au sujet s .

1.2. Les mesures effectuées

Il n'existe pas aujourd'hui de lecteur français qui n'ait vu des caractères chinois: caractères imprimés carrés; caractères réguliers mais plus libres, présents sur toute peinture chinoise; caractères enfin des chefs d'œuvre de la calligraphie, où l'Occidental peut apprécier une harmonie de formes mais ne distingue, à proprement parler, aucun trait. Les réalisations qui font l'objet de la présente étude sont assez soignées pour que l'on puisse effectuer des mesures par référence au modèle carré de l'écriture imprimée. Ces mesures portent soit sur l'ensemble du caractère soit sur un de ses traits soit sur deux traits. La mesure W est un rapport largeur/hauteur; la mesure T porte sur la pente d'un trait qui, en imprimerie serait horizontal; la mesure V exprime de même l'inclinaison d'un trait vertical. Enfin la mesure F est un facteur de symétrie calculé comme le rapport (a/b) des angles que font avec la verticale deux traits inclinés t_a et t_b : t_a montant de la gauche vers le haut et t_b descendant du haut vers la droite; les deux traits devant être symétriques ($a/b=1$) selon le modèle de l'écriture carrée. Du fait des habitudes de l'écriture manuscrite (a/b) est généralement inférieur à la valeur prescrite de 1. Nous donnons d'une part, pour chacune des quatre mesures W, F, T, R l'ensemble des caractères sur lesquels elle a été effectuée, en précisant s'il y a lieu par des flèches les traits concernés; et d'autre part un tableau récapitulatif, par caractère, des mesures effectuées.

	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
W	+	+	+		+				+		+				+	+	+	+	+	+		+	+	+	
F											+	+			+							+	+		
T +				+	+	+	+	+	+	+		+	+	+		+									
V +				+	+	+	+	+	+	2		+	2	+		+									

Tableau récapitulatif des mesures effectuées sur les caractères:

W: rapport largeur/hauteur (Width/height);

F: facteur de symétrie (entre deux traits obliques);

T: (pente (Tilt) d'un trait quasi horizontal;

V: inclinaison (slant) droite (right) ou gauche (left) d'un trait vertical;

On lit, e.g., que pour le caractère 9 on a mesuré W, T (sur 1 trait), V (sur un trait); sur le caractère 11 on a mesuré T (sur un trait) et V (sur 2 traits); etc...

On voit que les mesures T et V sont effectuées sur les mêmes caractères; par ailleurs, deux mesures ne se rencontrent qu'occasionnellement sur le même caractère; la caractère 10 n'est pas utilisé ici.

NB: les traits verticaux des caractères chinois étant toujours tracés en partant du haut, il faut entendre par inclinaison droite R (Right slant) décalage

vers la droite de l'extrémité inférieure relativement à l'extrémité supérieure; et le contraire pour L.

v2 = 月	v3 = 日	v4 = 用	v6 = 有	v9 = 把
v12 = 命	v16 = 香	v17 = 星	v18 = 家	v19 = 們
v20 = 報	v21 = 道	v23 = 察	v24 = 劉	v25 = 鄭
W: mesure du rapport largeur/hauteur (Width/height)				
v12 = 命	v13 = 林	v16 = 香	v22 = 會	v23 = 察
F: mesure du facteur de symétrie				
v1 = 千	v5 = 生	v6 = 有	v7 = 年	
v8 = 地	v9 = 把	v11 = 何	v13 = 林	
v14 = 姐	v15 = 到	v17 = 星		
T: mesure de la pente (Tilt)				
v1 = 千	v5 = 生	v6 = 有	v7 = 年	
v8 = 地	v9 = 把	v11(a) = 何	v11(b) = 何	
v13 = 林	v14(a) = 姐	v14(b) = 姐	v15 = 到	
			v17 = 星	
V: mesure de l'inclinaison sur la Verticale (Slant)				

§1.2: Tableau des caractères sur lesquels sont effectuées les mesures

1.3. Codage des données et notations

Comme on le voit sur le tableau récapitulatif, on effectue sur chaque réalisation i (de l'ensemble C des 25 caractères retenus) un ensemble Q de 44 mesures: 15W, 5F, 11T, 13V. Il sera commode de noter $Q = QW \cup QF \cup QT \cup QV$; dans la suite, au vu des résultats de classifications, on subdivisera QF , QT , QV respectivement en $FX \cup FY$, $TX \cup TY$, $VX \cup VY$. Les 44 variables de Q sont découpées en un ensemble J de 119 modalités: on notera $J = JW \cup JF \cup JT \cup JV$. Voici comment le découpage a été effectué. Pour chacune des variables q de QW , QF ou QT on a construit l'histogramme des 376 valeurs qu'elle a prises dans les diverses réalisations de C ; puis on a défini 3 modalités q_1, q_2, q_3 à peu près égales en fréquence. Pour chacune des variables de QR (inclinaisons) on a seulement distingué deux modalités: R (right = droite) et L (left = gauche). Finalement on notera:

$$JW = W1 \cup W2 \cup W3; JF = F1 \cup F2 \cup F3; JT = J1 \cup J2 \cup J3; JR = R \cup L;$$

avec éventuellement des subdivisions en deux groupes X et Y : $F1 = FX1 \cup FY1$; etc. . Quant aux modalités individuelles on notera par exemple 13T1 la modalité 1 (i.e. faible) de la pente du trait quasi horizontal considéré sur le caractère 13; pour les caractères 11 et 14 sur lesquels on a mesuré l'inclinaison de 2 traits verticaux, on précise par une lettre A ou B duquel de ces traits il s'agit: e.g. 14AR pour l'inclinaison à droite du trait A du caractère 14; 11BL pour l'inclinaison à gauche du trait B du caractère 11; etc... Ceci fait, on a, pour décrire l'ensemble I des 376 réalisations un tableau $I \times J$ (376×119) sous forme disjonctive complète avec e.g., $k(i, 13T1) = 1$ si dans la réalisation i de C le caractère C se présente avec un trait de pente faible; et sinon $k(i, 13T1) = 0$.

1.4. L'enchaînement des analyses

Notre but est double: d'une part reconnaître des sous-ensembles qui définissent des styles cohérents d'écriture ou, comme nous le dirons, des syndromes; d'autre part, observer dans quelle mesure le style d'un même individu qui écrit sans contrainte peut varier d'une réalisation à l'autre. Ainsi, on pourra, dans la suite, déceler une écriture déguisée d'après l'incohérence de son style associant des caractères disparates; ou, peut-être, reconnaître l'auteur qui n'a pu, malgré ses efforts, s'affranchir de certaines de ses habitudes.

Afin de définir des syndromes, on a construit et analysé le tableau de Burt $I \times I$ (119×119), ou tableau de cooccurrence des modalités associé au tableau de description $I \times J$ (avec, e.g., $k(13T1, 14AR)$ = nombre des réalisations rentrant à la fois dans les modalités 13T1 et 14AR); puis, d'après les résultats de l'analyse factorielle du tableau $J \times J$, on a effectué une classification ascendante hiérarchique (CAH) sur l'ensemble J des modalités descriptives. Afin d'étudier la stabilité du style des individus, on a effectué diverses classifications sur l'ensemble I des 376 réalisations de C ; classifications d'après des tableaux $I \times Z$,

croisant l'ensemble I avec une partition Z de J; partition choisie d'après la CAH sur J.

Nous présenterons successivement les résultats de l'analyse factorielle du tableau de Burt (§ 2); la classification automatique de l'ensemble j, interprétée en termes de styles d'écriture (§ 3).

2. Analyse factorielle du tableau de Burt

2.0. Valeurs propres et taux

rang	1	2	3	4	5	6	7
lambda	0.0289	0.0204	0.0061	0.0045	0.0032	0.0023	0.0021
taux	31.56%	22.24%	6.70%	4.98%	3.50%	2.51%	2.26%
cumul	31.56%	53.81%	60.51%	65.49%	68.99%	71.51%	73.77%

Le tableau ci-dessus montre l'importance primordiale des axes 1 et 2; nous interpréterons également l'axe 3 et effectuerons la CAH dans l'espace engendré par les 7 premiers axes. Le tableau ci-dessous nous guidera dans l'interprétation des facteurs 1 à 3: il donne pour chacun de ceux-ci le pourcentage total des contributions reçues de chacun des quatre groupes de modalités, W (largeur/hauteur), F (facteur de symétrie), T (pente des horizontales) et V (inclinaison des verticales).

rang	1	2	3
W	20%	17%	48%
F	15%	10%	19%
T	14%	58%	26%
V	50%	15%	7%

On voit que W, T, V dominent respectivement les axes 1, 2 et 3; tandis que le facteur de symétrie F, qui n'a fait l'objet que de 5 mesures contribuent modérément aux trois axes; (mais avec, pour les modalités extrêmes F1 et F3, des qualités de représentation de 400 au moins (et généralement 700) dans l'espace engendré par les axes 1, 2, 3).

2.1. Le premier facteur

Voici les modalités qui contribuent le plus à l'axe 1:

du côté $F1 > 0$:

(12F1,22F1,23F1)(6R,9R,17R,13R,14BR,7R,5R,1R,15R,11BR,8R,14AR);

du côté $F2 < 0$: (12F3,22F3)(6L,9L,11AL).

Plus précisément on a du côté $F1 > 0$) toutes les modalités R (inclinaison à droite) avec toutes les modalités F1 (faible valeur a/b: i.e., cf. § 1.2, angle du trait oblique de gauche avec la verticale, inférieur à l'angle du trait oblique de droite); et du côté ($F1 < 0$) toutes les modalités L (inclinaison à gauche des verticales) et toutes les modalités F3 (a/b fort, i.e. voisin de 1, ce qui correspond à une disposition symétrique des deux traits considérés).

2.2. Le deuxième facteur

Voici les modalités qui contribuent le plus à l'axe 2:

du côté $F2 > 0$: 1OF3 (6T1,9T1,17T1,13T1,14T1,7T1,1T1,11T1);

du côté $F2 < 0$: 1OF1 (6T3,9T3,13T3,7T3,1T3,11T3).

Plus précisément, on a du côté $F2 > 0$) toutes les modalités T1 (pente faible, voire descendante, des traits horizontaux); et du côté ($F2 < 0$) toutes les modalités T3 (forte pente ascendante vers la droite des traits horizontaux. Les modalités intermédiaires xT2 se projettent sur l'axe 2 entre xT1 et xT3, excepté pour le caractère 8, la modalité 8T2 étant du côté ($F2 < 0$) au-delà de 8T3; mais ce cas est à part, car comme on peut le voir sur le tableau des caractères (§ 1.2) le trait considéré sur le caractère 8 est incliné même sur le modèle d'imprimerie.

2.3. Le troisième facteur

Voici les modalités qui contribuent le plus à l'axe 3:

du côté $F3 > 0$:

(12F1,22F1)(17T1,5T1,15T1,8T1,11T3)(2W3,25W3,16W1,9W1,12W1);

du côté $F3 < 0$: (5T3,8T3,11T1)(2W1,17W1,25W1,16W3,12W3).

Près de la moitié de ces contributions viennent des variables W (largeur/hauteur); toutefois on trouve simultanément des modalités W1 et W3 de chacun des deux côtés ($F3 > 0$) et $F3 < 0$); ce qui rend difficile l'interprétation de l'axe 3.

2.4. Explication des graphiques plans

Plutôt que de figurer par leurs sigles alphanumériques les modalités de l'ensemble j, on a représenté chacune d'elles par un caractère où elle est réalisée: par exemple, à la place de la modalité 2W1 (très faible valeur du rapport largeur/hauteur pour le caractère 2) on a tracé un spécimen très étroit du caractère 2 (lune); etc. .

3. Classification de l'ensemble des modalités descriptives

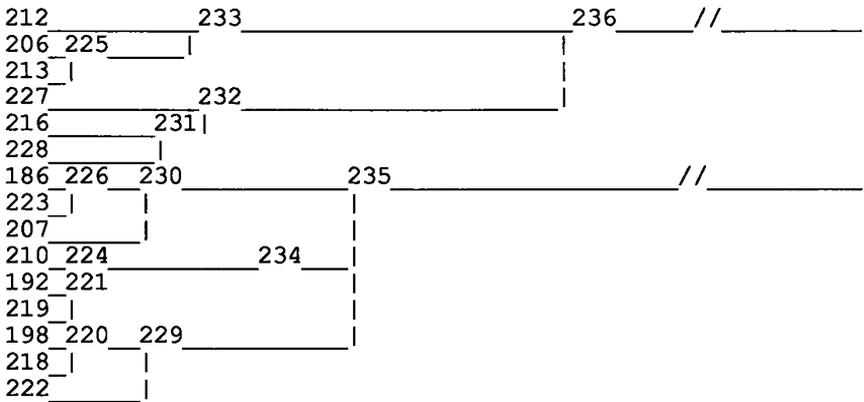
Nous présenterons successivement le schéma de la partition retenue (§ 3.1), le contenu des classes (§ 3.2), leur interprétation en termes de syndromes stylistiques (§ 3.3) et une subdivision de l'ensemble des variables mesurées (§ 3.4).

3.1. La partition retenue

En coupant l'arbre de la CAH au-dessous des quatre nœuds les plus hauts, on obtient une partition en 5 classes dont chacune pourra s'interpréter naturellement comme définissant un style d'écriture.

En retenant également les 5 nœuds suivants et 5 autres nœuds, on partage chacune des 5 classes en 3 subdivisions qui, étant relativement homogènes quant à leur contenu, (e.g. composées exclusivement de modalités R, ou de modalité T1, etc.), aideront à l'interprétation et serviront dans la suite (§ 4) à interpréter les classifications de l'ensemble des réalisations et donc à caractériser les styles des sujets et à en apprécier la stabilité.

Voici donc le schéma de la partition retenue.



3.2. Le contenu des classes

Nous considérerons successivement les 5 classes principales avec leurs subdivisions;

classe 233 = (212,225) = (212,(206,213)):

212 = (6,11,9,13,1,7)T1;

206 = (11B,14B,15,11A,6,9)R + 4W1;

213 = (5,17,7,13,14A,8,1)R;

on voit que la subdivision 212 comprend 6 modalités T1 (pente faible) associées aux six caractères (6,11,9,13,1,7); tandis que 206 contient, en plus de 6 modalités R, (inclinaison à droite), la modalité 4W1; etc...

classe 232 = (227,231) = (227,(216,228)):

227 = (1,7,13,9,6,11)T3 + (16,13)F1 + 24W1;

216 = (12,16)W1 + (12,22,23)F1;

228 = (18,9,20,23,3,2,17,25)W1 + (14,8,17,5,15)T3;

classe 230 = (226,207) = ((186,223),207):

186 = (5,14,17)T1 + 17W3;

223 = (4,2,3,19,25,18,6,24)W3 + (8,15)T1 + (16,13) F3;

207 = (12,9,20)W3 + (12,22,23) F3;

classe 224 = (210,221) = (210,(192,219)):

210 = (21,23)W3 + 23F2 + 11T2;

192 = (17,13,5,7,1,14A,8)L + 6T2;

219 == (11B,14B,11A,15,6,9)L + (9,14,1,13)T2;

classe 229 = (220,222) = ((198,218),222):

198 = (4,24,12,16)W2 + (16,13)F2;

218 = (23,17,19,6,25,18)W2 + (15,5,7,17)T2 + 21W1;

222 = (9,20,3,21,2)W2 + (19,6)W1 + (12,22)F2 + 8T2 + 16W3.

3.3 Syndromes stylistiques

Pour suivre l'interprétation des 5 classes principales, nous recommandons au lecteur de considérer le tableau des caractères chinois figurant les contenus de ces classes avec leurs subdivisions: dans ce tableau, comme sur les graphiques plans (§ 2.4) une modalité est représentée par un caractère où elle est réalisée, en traçant de plus en double les traits sur lesquels des mesures ont été effectuées.

classe 233: syndrome (R,(T1)) de l'écriture penchée en arrière avec pente généralement faible des traits horizontaux;

classe 232: syndrome (T3,F1,W1) de la calligraphie verticale traditionnelle: pente des traits horizontaux nettement positive, trait oblique gauche nettement moins écarté de la verticale que son homologue droit, caractères étroits;

classe 230: syndrome (F3,W3,(T1)) de l'imprimerie carrée symétrique: caractères bien symétriques, larges, avec pente faible de traits horizontaux;

classe 224: syndrome (L,T2): écriture penchée en avant avec pente modérée des traits horizontaux;

classe 229: W2 tendance centrale particulièrement quant à la largeur.

212:	{有把林年十何}
206:	{有把姐到何何} + {用}
213:	{星林年生十地姐}
227:	{有把林年十何} + {香林} + {劉}
216:	{命察會} + {月命}
228:	{星姐生到地} + {家把報察日月星鄭}
186:	{星姐生} + {星}
223:	{到地} + {月有日劉鄭們家用}
207:	{把命報} + {香林} + {命察會}
210:	{察} + {何} + {道察}
192:	{有} + {星林年生十地姐}
219:	{把姐林十} + {有把姐到何何}
198:	{香林} + {香命劉用}
218:	{道} + {有星察鄭們家} + {星年生到}
222:	{香} + {們有} + {命會} + {月日把道報} + {地}

Tableau des modalités descriptives rangées en 5 classes subdivisées chacune en 3; sur les caractères représentant une modalité F, T ou V les traits, objet de mesure, sont tracés en double

3.4. Subdivision de l'ensemble Q des variables mesurées

Au § 3.1 on a noté: $Q = QW \cup QF \cup QT \cup QV$; et de même pour l'ensemble des modalités $J = JW \cup JF \cup JT \cup JV$. Ces ensembles sont diversement morcelés dans la partition retenue, selon les modalités dont il s'agit:

1) rien de net n'apparaît quant au partage de QW et de W1, W2, W3;

2) pour le facteur de symétrie il convient de poser:

$$QF = (16,13) \cup (12,22,23) = FX \cup FY;$$

en effet les sous-ensembles FX1, FX2, FX3, FY1, FY2, FY3 se retrouvent comme constituants des classes inférieures de la partition retenue;

3) pour la pente (Tilt) on posera de même:

$$QT = (1,6,7,9,11,13) \cup (5,8,14,15,17) = TX \cup TY;$$

en effet TX1, TX3, TY1, TY3 se retrouvent exactement; et TX2, TY2 approximativement;

4) pour l'inclinaison (droite ou gauche) on posera

$$QV = (6,9,11A,11B,14B,15) \cup (1,5,7,8,13,14A,17) = VC \cup VY;$$

car les ensembles de modalités RX, RY, LX, LY se retrouvent respectivement dans 206, 213, 219 et 192.

Si l'on cherche à interpréter les subdivisions introduites ici, on note d'abord une similitude dans la disposition des paires de traits obliques symétriques des caractères 16 et 13 de FX. Quant aux mesures de pente (Tilt) des traits horizontaux, on a dans TY les mesures portant sur des traits situés à la base du caractère notamment pour les caractères 5 et 17; dont le second s'écrit en superposant au premier le caractère³); tandis que dans TX il s'agit de traits situés à un niveau supérieur ou intermédiaire. Enfin pour les mesures d'inclinaison (slant) droite ou gauche de traits verticaux, on a dans VX des mesures effectuées sur un trait vertical au bord droit ou gauche du caractère (le cas 9 étant seul douteux); et dans VY les autres mesures. Cette interprétation explique pourquoi, bien que les mesures QT et QV portent sur les mêmes caractères, les subdivisions en X et Y ne coïncident pas.

Remerciements:

One of us, (S.-C. L.), would like to record here his gratitude to Mr. N.-S. Lee, the Government Chemist of Hong Kong, for his interest and encouragement in this work.