

Lire le chinois - analyse quantitative

philippe.coueignoux@eprivacy.com

Ph Coueignoux 16/07/20

-0- Table des Matières

-1- Introduction:

-2- Liste des marques

nombre des marques

essai de classification graphique

-3- Histogrammes concernant la phonétique

représentation des sons par les marques

dispersion des marques sur les sons

-4- Quantification de la dispersion phonétique

calcul de la mesure de dispersion

un dernier histogramme basé sur la mesure de dispersion

-5- Examen des différences entre phonologie et décodage

le caractère apico-alvéolaire de 'sh'

le caractère prédorso-palatal de 'l'

la cohérence de la famille 'm'-'h'-'r'-'n'-'l'

-6- Conclusion

-7-Annexes

histogramme de la distribution des marques sur les sons

histogramme de la distribution des sons sur les marques

histogramme de la distribution des 'coudes phonétiques' sur les marques

exemples de calcul de dispersion phonétique

histogramme de la mesure de dispersion phonétique

-1- Introduction

Basé sur une approche originale de détermination de la structure des caractères chinois, l'auteur a compilé une liste de marques [Dictionnaire des marques du chinois écrit, Ph Coueignoux, février 2020, disponible sur demande]. Ces marques ou 部件 (bùjiàn), sont les éléments constitutifs des caractères qui influent normalement sur le sens ou le son des caractères où elles apparaissent.

Le concept de 部件 est bien connu même si leur liste peut varier légèrement d'un compilateur à l'autre. On peut comparer par exemple avec [GF 0014-2009 现代常用字部件及部件名称规范. pdf disponible sur le site <http://www.moe.gov.cn>]. Le fait qu'un 部件 peut 'marquer' le sens d'un caractère a servi depuis des temps immémoriaux à différencier des caractères homophones [L'emploi des phonétiques dans l'écriture chinoise, Laurent Sagart 2006]. Quant à leur influence phonétique, elle a été analysée dès la fin du XIXème siècle par des pionniers comme Léon Wieger, SJ [Caracteres chinois Sienhsien, Leon Wieger, imprimerie de la mission catholique, 3e edition, 1916] et de Bernhard Karlgren [Analytic Dictionary of Chinese, Bernhard Karlgren,

Dover Publications, 1974 (from the 1923 original)].

L'originalité de la démarche de l'auteur est donc à chercher dans sa méthode plus que dans ses résultats. Cette méthode est basée sur le fait que l'influence phonétique des marques sur les caractères peut être assimilée à un code que l'on peut retrouver selon les techniques traditionnelles du déchiffrement [Décodage des caractères chinois, Ph Coueignoux 23/05/20, non publié]. Indépendante des efforts de ses prédécesseurs, cette démarche renouvelle les dictionnaires compilés par Wieger et Karlgren. Une raison immédiate en est l'utilisation de l'écriture chinoise simplifiée des caractères cités dans le corpus choisi pour l'étude, le 'Pocket Oxford Chinese Dictionary, troisième édition, 2003, Oxford University Press. Une raison plus intéressante vient de son caractère systématique et exhaustif indépendant des apports de la phonologie.

Ceci a conduit à la fois à redécouvrir les résultats déjà obtenus par la phonologie mais aussi à quelques différences significatives qui sont détaillées dans [Lire le chinois - l'apport de l'analyse, Ph Coueignoux 15/04/19, non publié]. Ce dernier article ne proposait encore que des arguments qualitatifs. Depuis l'achèvement de son dictionnaire de marques, l'auteur peut dorénavant donner des données quantitatives, objet du présent article.

On commencera par organiser la liste des 1254 marques compilées dans le dictionnaire de l'auteur, liste qui lui sert d'index au même titre que l'index des clés constitue la première partie d'un dictionnaire de caractères.

On poursuivra en donnant trois histogrammes. Le premier donnera une idée de la distribution du nombre de marques par son, c'est à dire pour chaque son pinyin, combien de marques sont susceptibles de le représenter dans les caractères du corpus choisi. On complétera cet histogramme en considérant qu'un son peut aussi provenir d'un caractère sans qu'il soit possible de l'attribuer à l'une de ses marques constitutives. Réciproque du premier, le deuxième histogramme modélisera la distribution du nombre de sons par marque, c'est à dire pour chaque marque, combien de sons elle influence dans les caractères du corpus choisi. Dans les deux cas, il apparaîtra que ces liens d'influence sont, sinon prédictifs (cas non réalisé où chaque son est représenté par une marque distincte et unique), du moins très sélectifs (le plus souvent, peu de marques pour chaque son, peu de sons pour chaque marque). Le troisième histogramme complètera la modélisation de la dispersion des sons par marque en considérant la forme de cette dispersion sur le fond de carte phonétique présenté dans [Décodage des caractères chinois, déjà cité].

Ce fond de carte phonétique universel et optimal est utilisé pour reporter les sons représentés par chaque marque et former ce que l'on appelle sa carte phonétique. On donnera alors en détail la façon de mesurer de

dispersion d'une marque donnée, selon une jauge de l'étendue occupée par les sons de son spectre phonétique sur le fond de carte phonétique, depuis la valeur zéro pour les marques qui n'admettent qu'un seul son pinyin (sans tenir compte des tons), jusqu'aux valeurs supérieures à dix des marques d'influence aussi multiforme que diluée. Cette mesure fera naturellement l'objet d'un nouvel et dernier histogramme.

Enfin on donnera un tour plus quantitatif aux anomalies que l'on constate en comparant les résultats de l'analyse par décodage à ceux de la phonologie, notamment sur le rôle de la consonne initiale 'sh' comme quasi apico-alvéolaire (dentale), celui du 'l' comme quasi pré-dorso-palatale (palatale) et sur la famille constituée par les initiales 'm', 'h', 'r', 'n' et 'l'.

-2- Liste des marques

nombre des marques

La liste des marques compilée par l'auteur comporte 1254 entrées. En lui-même, ce nombre est arbitraire et n'a qu'une valeur indicative que l'on pourra comparer au nombre des entrées du dictionnaire de Karlgren déjà cité, 1350, sachant que Karlgren travaillait sur les caractères traditionnels et non pas simplifiés.

Il est clair en effet que le nombre de marques dépend du corpus de caractères choisi et que rien n'interdit dans le futur de créer de nouvelles marques, soit pour étendre le vocabulaire, soit pour simplifier les caractères.

Mais l'incertitude sur le nombre des marques provient aussi d'un conflit inhérent aux deux buts qui justifient leur importance. Le premier, théorique, est de respecter la caractéristique des marques, à savoir le fait d'être un élément constitutif répété dans plusieurs caractères, la base du code graphique qui donne sa cohérence à l'écriture du chinois. Le deuxième, pratique, est de pouvoir couvrir par les dérivés de ces marques l'ensemble des caractères du corpus. Cet impératif permet en effet à la liste des marques de servir d'index au dictionnaire des caractères du corpus, comme la liste des clés indexe les dictionnaires classiques et facilite ainsi leur usage. Mais cela conduit à violer le but théorique car comment indexer les caractères singuliers, par exemple 年 ou 凸, qui ne peuvent être décomposés qu'en traits et qui ne sont pas repris dans la composition d'autres caractères? Soit on les prend comme des marques en tant que telles bien que l'association de traits qui les caractérise n'apparaisse nulle part ailleurs, soit l'index envisagé rend infructueuse la recherche des caractères concernés. Aussi bien Karlgren que l'auteur ont préféré la pratique à la théorie en incluant à leur compilation les quelques caractères que l'on ne peut indexer que par eux-mêmes dont les deux exemples ci-dessus.

essai de classification graphique

Quelle qu'en soit la valeur exacte, le nombre de marques dépasse le millier et il se pose à leur égard le même problème que pour les clés des dictionnaires classiques, qui sont quelques centaines. Pour avoir accès à ces listes qui servent d'index aux caractères qui en dérivent, il faut concevoir un deuxième niveau d'index.

Pour les clés, on utilise traditionnellement le nombre de traits. Ceci n'est pas sans présenter deux difficultés. La première est que ce nombre n'est pas défini avec toute la rigueur souhaitable en ce sens qu'il ne peut pas être déterminé par l'analyse visuelle mais en se rappelant leur synthèse gestuelle, qui ressort d'une autre logique. Ainsi 一 et | ont visuellement un seul trait et 二 a deux traits mais 匚 et 凵 n'ont eux aussi que deux traits et non pas trois comme on pourrait s'y attendre. Et si l'on cherchait une nouvelle règle pour harmoniser ce qui précède, elle pourrait expliquer le cas de 母, qui a cinq traits, mais serait immédiatement violée par 冂 qui n'a ni deux ni quatre traits mais trois ! La deuxième difficulté est que l'ordre dans lequel sont énumérées les clés ayant le même nombre de traits est, sinon arbitraire, du moins variable selon le dictionnaire concerné.

Pour les marques, la première méthode est de les indexer selon leur son représenté en pinyin. Karlgren avait déjà fait un choix similaire selon une transcription phonétique propre à l'auteur, sachant qu'il écrivait plus de cinquante ans avant la mise au point du pinyin. Une telle approche est très pratique si l'emploi du dictionnaire se limite à retrouver le sens d'un caractère ou d'un mot dont on a retenu la prononciation. Car, comme le détaillera la section suivante sur la représentation des sons par les marques, le nombre de marques pour un son donné est en général très limité. Ainsi la recherche d'un caractère prononcé 'jia', quel qu'en soit le ton, donnera la liste suivante des marques responsables du son 'jia': 家 圭 加 介 夹 段 甲. Retrouver par exemple le caractère 价 en sélectionnant sa marque phonétique 介 dans cette liste est trivial. On peut d'ailleurs automatiser cette étape et présenter d'emblée tous les caractères de son 'jia' dérivés de ces marques, comme si l'on feuilletait aux pages consacrées au son 'jia' un dictionnaire moderne utilisant le pinyin. L'avantage est de pouvoir parcourir la liste des caractères d'un coup d'œil au lieu de devoir feuilleter des pages entières pour y repérer les caractères au milieu des articles de longueur variable qui leur sont consacrés.

Mais que faire si l'on est en présence d'un caractère inconnu ou dont on a simplement oublié la prononciation? L'auteur propose de classer les marques selon leur apparence graphique. Cette approche est décrite dans [Lire le chinois - l'apport de l'analyse, déjà cité]. Elle utilise trois dimensions. Les deux premières disposent les marques dites principales dans le tableau ci-dessous.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	P	Q	R	S	T	U	W	X	Y	Z		
a	一	丨	十	丁	冂	乙	乚	厂	冂	飞	人	大	木	小	艹	×	△	匚	口	日	五	田			a	
b	二	亻	土	ノ	刁	之	辶	讠	斤	冂	戈	入	犬	术	小	井	义	合	巨	口	白	互	由		b	
c	三	彳	艸	リ	勹	了	辶	彳	丘	冂	戈	欠	太	木	尔	井	义	谷	臣	回	旦	丑	甲		c	
d	丶	川	士	リ	勹	子	圣	也	巾	辰	巳	戈	久	李	本	少	井	父	舍	臣	只	旧	尹	申	d	
e	宀	冫	工	丁	勹	子	亥	世	长	反	巴	戈	人	尤	米	小	甘	交	金	亡	兄	亘	争	电	e	
f	冂	卜	干	于	勹	亨	己	心	厶	后	冂	戒	人	尤	采	冰	艹	丈	立	匚	台	百	月	龟	f	
g	宀	上	平	寸	包	享	己	必	么	辰	冂	戎	八	尤	采	水	共	又	辛	冂	占	早	月	曳	g	
h	宀	卜	千	可	司	予	弓	屯	公	皮	民	戌	穴	无	禾	永	共	又	亲	冂	古	卓	丹	更	h	
i	ノ	下	午	乎	刀	矛	弟	七	宀	虎	艮	戌	天	无	未	永	孔	又	幸	冂	周	卓	舟	更	i	
j	彳	卜	牛	笋	刃	正	弟	七	云	广	艮	戌	天	未	求	氏	文	幸	冂	舌	曰	母	冉		j	
k	彳	卡	牛	才	乃	正	夷	匕	会	疒	良	戌	儿	天	朱	示	氏	支	音	冂	言	甘	每	禺	k	
l	宀	赤	牛	牙	及	象	弗	比	去	疒	良	成	九	关	余	余	代	支	音	冂	高	白	月	单	l	
m	宀	止	半	半	分		丐	毛	至	疒	食	咸	几	矢	乐	京	升	豆	向	高	目	肖	里		m	
n	宀	正	半	扌	力		丐	毛	彡	尸	目	或	几	夫	东	巾	开	夕	壹	尚	高	且	青	甫	n	
p	宀	步	生	彳	万		专	瓦	彡	户	自	戈	几	夫	东	市	并	夕	壹	冂	自	脊	用		p	
q	宀	足	王	手	方		丐	止	糸	尸	阜	气	兀	央	束	巾	亦	夕	史	录	冂	身	骨	角	q	
r	宀	乍	壬	豸	广		与	疒	率	不			免	奂	束	巾	亦	名	吏	冂	凹	耳	骨	角	r	
s	彳	斗	羊	豸	羽		写	足	参	石			先	失	来	雨	赤	夕	中	冂	凸	其	毋	鱼	s	
t	彳	斗	羊	豸	勿		号	走	彡	ナ			光	夹	来	雨	业	各	虫	山	四	具	册	卑	t	
u	宀	宀	丰	丰	勿		马	走	彡	ナ			兆	头	林	丽	亚	女	虫	缶	西	直	朋	鬼	u	
w		宀	韦	片			鸟	表	彡	发			火	彡		两	並	夕	禹	巾	西	垂			w	
x			巾	竹			鸟	衣		发			贝	彡		而	兼	文	串	出	酉	皿			x	
y			车	川			鸟	衣		龙			页	美		爪	非	多	革	齿	鹵	血			y	
z			佳	州			彳			农			见			瓜			董	鹵		面			無	z
	.A.	.B.	.C.	.D.	.E.	.F.	.G.	.H.	.I.	.J.	.K.	.L.	.M.	.N.	.P.	.Q.	.R.	.S.	.T.	.U.	.W.	.X.	.Y.	.Z.		

tableau des marques principales

Le but de ce tableau est de limiter les énumérations à 24, un nombre proche du nombre de lettres dans les alphabets latins et que l'on est donc capable, au moins en théorie, d'apprendre par cœur. La plus grande complexité de l'écriture chinoise est reflétée par le fait qu'une simple énumération de 26 lettres comme l'alphabet français fait place à 24 énumérations de familles de 12 à 24 marques principales, chacune d'entre elles étant notée par deux lettres, la première, majuscule, désignant une famille caractérisée par une certaine ressemblance visuelle entre ses membres, la deuxième, minuscule, désignant un membre particulier de cette famille. La marque 甲 que l'on a déjà rencontrée pour le son 'jia' est par exemple à la case Zc, la troisième position de la vingt-quatrième famille, caractérisée par une grille sur le modèle de sa tête de liste Za, 田.

A titre de curiosité l'on peut construire l'histogramme de ces familles, c'est à dire de compter combien de familles ont une certaine taille sur la plage des tailles réelles qui va de 12 à 24. On obtient le résultat suivant:

1	0	0	0	2	0	0	0	4	2	0	5	10	24
12	13	14	15	16	17	18	19	20	21	22	23	24	total

histogramme de la taille des familles de marques principales

Ceci permet de retrouver facilement le total des marques principales en calculant la somme

$$1 \times 12 + 2 \times 16 + 4 \times 20 + 2 \times 21 + 5 \times 23 + 10 \times 24 = 521$$

Comme pour tous les chiffres donnés dans cette étude, il est illusoire de voir en 521 une valeur précise, car ce total inclut les variantes graphiques que l'on estime suffisamment différentes visuellement pour les compter séparément mais qui renvoient in fine à la même marque pour ce qui est de la liste des dérivées, comme le triplet 羊(Cs), 𦍋(Ct), 𦍋(Dt) ou la paire 人(Ma)/ 𠤎(Me). En excluant ces variantes, dont la détermination ne peut être totalement objective, le total n'est que de **496**.

Si l'on qualifie de primitive une marque qui ne contient aucune autre marque à part elle-même, comme 羊 ou 人, on notera que le tableau des marques principales ne contient pas uniquement des marques primitives mais aussi des marques dérivées comme 里 (Zm), composée de 田 (Za) et de 土 (Cb), ou 肖 (Ym) composée de 丩 (An) et de 冂 (Yl). Le but est en effet d'inclure autant de marques que possible dans ce tableau, qui peut se consulter d'un coup d'œil, en privilégiant en plus des primitives, celles dont les formes graphiques composées ne sont pas immédiatement reconnaissables et peuvent prêter à confusion, comme 島 (Gy) avec 𪇑 (Gx), ou dont la décomposition en marques primitives est sujette à débat, comme 亨 (Fg).

La troisième dimension complète les marques principales en énumérant pour chacune d'elle les marques secondaires qui en dérivent en la contenant visiblement. Elle donne lieu à une liste indexée par un numéro d'ordre, le numéro '01' correspondant à la marque principale concernée. Voici par exemple les listes associées aux marques principales Ma-Me.

-M-	1	2	3	4	5	6	7	8	9	10	11
a	人	𠤎	𠤎	𠤎							
b	入										
c	欠	次	歛	厥	欣	歛					
d	久										
e	人	坐	从	以	巫	夾	内	丙	𠤎	囚	

tableau illustrant l'énumération des marques secondaires

Sur cette troisième dimension, les répétitions sont admises pour faciliter la recherche des marques secondaires qui contiennent plusieurs marques principales: ainsi la marque 从 est codée respectivement Ma-03 et Me-03 selon ses marques constitutives 人 et 人. Pour revenir aux marques responsables du son 'jia', on pourrait s'exercer à chercher 加 dans les listes En (力) ou Wa (口). On la trouverait soit en En-02, soit en Wa-18. Par convention le code ayant le plus petit numéro d'ordre et, en cas d'égalité, le premier par ordre alphabétique est celui qui est utilisé pour indiquer la marque et le fichier correspondant. Hors répétitions, on compte **758** marques secondaires, soit approximativement 60% du total des marques. Le nombre des 'duplications' est de **764**. Ceci souligne qu'en moyenne une marque secondaire se compose de deux marques principales, les exceptions faites par les marques comprenant plus d'une marque principale (ex.: 磨 est connue à la fois comme Js-03, Pu-05 et Jj-07 car elle dérive de 石 (Js-01) et de 麻 (Pu-04, Jj-06), elle même dérivée de 林 (Pu-01) et 广 (Jj-01)) compensant en gros celles faites d'une seule marque principale associée avec une forme sans existence propre (ex.: 曷 (Xj-02) dérivée uniquement de 日 (Xj-01), 匈 étant toujours la partie inférieure de 曷).

On constate qu'il n'est pas possible de limiter la troisième dimension à 24 membres. Wa (口) compte 36 membres et Xa (日) 34 membres mais toutes les autres listes ont 24 membres au plus. On peut encore construire l'histogramme de ces listes de marques secondaires, c'est à dire de compter combien de listes ont une certaine taille pour les tailles admissibles de 0 à 24, mis à part les deux exceptions de 34 et 36. On obtient le résultat suivant:

184	101	62	48	22	17	17	7	13	9	6	3	3	4	5	6	3	2	0	1	0	2	1	3	519
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	total

histogramme de la taille des familles de marques secondaires (hormis 日 et 口)

Une façon simple d'interpréter ce résultat est de regrouper les tailles de familles en plages parlantes et de présenter des pourcentages, par exemple, en incluant cette fois les deux exceptions:

35%	31%	20%	10%	3%	1%	100%
1	2-3	4-7	8-15	16-23	24-39	total

distribution de la taille des familles de marques secondaires

On voit donc qu'un bon tiers des primitives principales n'ont pas de marques secondaires associées, un autre

petit tiers n'ont en fait qu'une ou deux associées et que, sur le tiers restant, seule une toute petite fraction ont plus de quinze associées. Au regard d'une distribution aussi biaisée, parler de moyenne, en l'occurrence 4 (une principale et trois secondaires associées), n'a pas beaucoup de sens. Il vaut mieux parler de la médiane, qui est à 2+. C'est à dire que grosso modo il y a autant de listes à 1 ou 2 membres que de listes de plus de deux membres. On notera que les marques principales qui ont plus de vingt marques secondaires associées sont toutes des clés d'influence sémantique notable: 口 (Wa-01), 日 (Xa-01), 十 (Ra-01), 木 (Pa-01), 宀 (Ag-01), 田 (Za-01), 亻 (Bb-01) et 又 (Sg-01). Mais la réciproque n'est pas vrai, car la clé 扌 (Dn-01), dont la liste de caractères dérivés est parmi les plus longues, en tant que marque principale a peu de marques secondaires associées, à savoir 6.

Cette approche par impression visuelle est loin d'éliminer toute difficulté. Bien que l'auteur ait adopté une certaine logique dans la composition du tableau des marques principales et des listes de marques secondaires, la part d'arbitraire reste importante. Et s'il semble relativement facile de retrouver une marque secondaire via la ou les marques principales qui la composent, il n'est pas possible de retrouver une marque principale donnée dans le tableau des marques principales sans tâtonner, à moins naturellement de savoir par cœur ses vinq-quatre familles. L'auteur constate seulement que l'habitude permet de s'y orienter de plus en plus vite.

-3- Histogrammes concernant la phonétique

A propos des listes de marques secondaires, un concept arbitraire, la section précédente donne un exemple frappant de distribution biaisée. Le même biais est souvent à l'œuvre dans les phénomènes naturels et les créations humaines et réapparaît à propos de concepts beaucoup plus objectifs centrés sur la prononciation.

représentation des sons par les marques

En règle générale le son d'un caractère chinois est influencé par l'une des marques qui le constituent, y compris le cas où le caractère lui-même est la marque en question, comme par exemple 一 /yī (Aa-01) ou 明 /míng (Xa-03). Seulement **136** caractères du corpus étudié font exception, comme par exemple 好 (hǎo). Mais plusieurs marques peuvent être derrière le même son et l'on peut donc calculer l'histogramme représentant cette distribution, celle des marques sur les sons. Le tableau complet, avec ses notes explicatives, est donné en annexe et ce qui suit en donne un résumé parlant en le regroupant par plages:

11%	20%	24%	24%	18%	3%	100%
0-1	2-3	4-5	6-9	10-19	20-44	total

distribution de la taille des représentations des sons par les marques

Sur les 403 sons pinyin représentés dans le corpus, plus de dix pour cent n'admettent qu'une seule origine phonétique. C'est évidemment toujours le cas des sons rares ne correspondant qu'à un caractère, comme 'fo' qui admet seulement 佛, dérivé phonétique de la marque 弗 (G1-01). C'est aussi souvent vrai pour des sons qui n'admettent pas plus de quelques caractères comme 'hei', deux caractères dérivés de la marque 黑 (Au-09) ou 'hen', quatre caractères dérivés de la marque 艮 (Ki-01). Mais le son 'fang', qui n'est exprimé que par la marque 方 (Eq-01), correspond lui à 11 caractères du corpus dont certains très courants et ne peut pas être considéré comme un son rare.

A l'extrême opposé, trois pour cent des sons peuvent s'exprimer avec plus de vingt marques distinctes sans compter les caractères exceptionnels. Il s'agit, classés par nombre de signes croissants, des douze sons 'li', 'xian', 'yan', 'jian', 'yu', 'zhi', 'xi', 'qi', 'shu', 'shi', 'yi' et 'ji'.

La distribution apparaît encore comme très biaisée, avec une moyenne de 6, peu significative, et une médiane de 4+: il y a environ autant de sons exprimés par quatre marques au plus que de sons exprimés par au moins cinq marques. Dit autrement, la moitié des sons pourraient faire l'objet d'une mémorisation de toutes les marques qui les influencent.

dispersion des marques sur les sons

Réciproquement chaque marque peut prétendre en théorie à avoir une influence phonétique sur un ou plusieurs sons. En incluant la possibilité de l'absence d'influence comme donnant lieu à 0 son, on peut donc encore une fois calculer l'histogramme d'une distribution, celle des sons sur les marques. Le tableau complet, avec ses notes explicatives, est donné en annexe et ce qui suit en donne un résumé parlant par plages:

7,8%	39,8%	25,3%	18,6%	7,9%	0,6%	100%
0	1	2	3-4	5-9	10-14	total

distribution de la taille de dispersion des marques sur les sons

Pour cette distribution encore une fois biaisée, la moyenne, peu significative, est de deux sons par marques et la médiane, de 2-, est elle-même moins parlante que de dire: seulement un peu plus du quart des marques influencent plus de deux sons distincts.

Au delà de deux sons la meilleure façon de rendre compte de la dispersion des marques sur les sons n'est pas en fait de comptabiliser les sons mais de compter les 'coudes' dans la partie consacrée à la consonne initiale des

cartes phonétiques associées aux marques. Pour la définition de la notion de 'coude', on renvoie à [Décodage des caractères chinois, déjà cité]. Plus une marque a de coudes, plus sa dispersion phonétique est complexe. A moins d'une exception, les marques qui ont deux sons au plus n'ont pas de coude, mais il existe bien des marques qui donnent plus de deux sons sans avoir de coude pour autant, comme par exemple 𠄎 (Dh-01), influençant les sons 'a' (𠄎), 'e' (𠄎), 'ge' (𠄎), 'ke' (𠄎) et 'he' (𠄎), tous sur la même colonne des consonnes initiales.

La prise en considération de ces 'coudes phonétiques' donne lieu à un nouvel histogramme, dont les détails sont donnés en annexe.

4,31%	8,45%	31,18%	37,40%	10,85%	5,98%	1,28%	0,08%	0,48%	100%
aphones	stériles	monosons	simples	1 coude	2 coudes	3 coudes	4 coudes	incohérentes	total

distribution de la taille de dispersion des marques sur les 'coudes phonétiques'

La précision de ces données, avec deux décimales, est évidemment illusoire mais permet de mettre en valeur que les cas limites à quatre coudes (un seul en fait) ainsi que les cas d'incohérence (six en tout) sont suffisamment rares pour ne pas invalider le système présenté dans [Décodage des caractères chinois, déjà cité].

Si l'on admet plus généralement que la présence de deux coudes ou plus est un signe de complexité phonétique, la leçon de cet histogramme est que cela ne concerne que 8% du total des marques, soit 99 sur 1254. Et sur ce nombre, plus des trois quarts, soit 76, se limitent à deux coudes, une complexité toute relative qui laisse encore très lisible la carte phonétique correspondante (voir en annexe le cas de 𠄎 (Ce-01)).

-4- Quantification de la dispersion phonétique

On peut critiquer le développement précédent sur la dispersion phonétique des marques comme insuffisant. Compter le nombre des sons n'est pas assez synthétique car il ne permet pas d'en distinguer la plus ou moins grande proximité phonétique et compter le nombre de coudes est à la fois trop synthétique et limité à l'influence de la consonne initiale. Pour atteindre aussi bien une plus grande généralité et une plus grande précision, il faut s'appuyer sur le concept de mesure de dispersion proposé dans [Décodage des caractères chinois, déjà cité].

calcul de la mesure de dispersion

La mesure de dispersion se détermine selon des règles très précises. Mais avant de les détailler, il convient

encore une fois de souligner que la précision qu'elles sont destinées à produire n'a pas de valeur absolue. Ce que l'on peut appeler 'la mesure de Coueignoux' n'est utile que par le classement qu'elle permet d'effectuer entre les spectres phonétiques de toutes les marques. D'autres mesures peuvent être envisagées dans le même but. On s'attend simplement que toutes donnent des classements voisins qui reflètent l'intuition qualitative.

La première règle s'exprime par la formule suivante, reprise de [Décodage des caractères chinois, déjà cité]:

$$\mathbf{Disp} = \mathbf{DispI} + \mathbf{DispM} + \mathbf{DispF}$$

C'est à dire que la mesure de dispersion est la somme de trois termes, correspondant respectivement aux matrices des initiales, médianes et finales. Des exemples sont donnés en annexe.

Le calcul de la dispersion sur la consonne initiale se fait comme suit, en renvoyant au fond de carte normalisé

					.
b	g	z	zh	d	j
p	k	c	ch	t	q
f	?	s	sh	?	x
m	h	.	r	n	l
w	?	?	?	?	y
- no initial consonant -					

(une case vide susceptible de recevoir une consonne voisine est signalée par un ?)

- si la marque ne met en jeu qu'une seule consonne,

y compris l'absence de consonne initiale ou le son pinyin 'er' équivalent de la consonne initiale 'r',

DispI est égal à zéro

- sinon, on recouvre ces consonnes avec un ensemble connexe de colonnes et de lignes, autrement dit assurant un chemin reliant deux consonnes quelconques du spectre, et pour un tel jeu:

- pour chaque colonne, on ajoute 1, 1,5, 1,75, 1,875, 1,9375 selon que le spectre y a 2, 3, 4, 5, 6 consonnes

- pour chaque ligne, on ajoute 2, 3, 3,5, 3,75, 3,875 selon que le spectre y a 2, 3, 4, 5, 6 consonnes

- pour chaque intersection, on ajoute 8 si cette intersection n'est pas occupée par une consonne du spectre,

c'est à dire qu'il n'y a pas un coude mais une incohérence.

Entre deux jeux possibles, on choisit celui qui minimise la somme précédente. On doit donc:

- éliminer les colonnes et les lignes de recouvrement superflues

- préférer les colonnes aux lignes

- préférer que les intersections entre lignes et colonnes restantes correspondent à un coude du spectre

On notera qu'en cas de coude, la consonne concernée est comptée deux fois, sur sa colonne et sur sa ligne.

Dans le choix des jeux, on a le droit de remplir une case vide '?' par la consonne d'une case voisine:

- le 'w' et le 'y' peuvent s'étendre sur toute la ligne correspondante,
- le 'h', le 'sh' et, dans de rare cas, le 's' peuvent aussi occuper la case vide (?) adjacente hors ligne 'w'-'y'

Ce droit offre deux bénéfices potentiels pour minimiser DispI:

- soit on fait juste migrer la consonne dans la rangée qui minimise sa contribution individuelle à DispI
- soit on duplique la consonne dans deux rangées parallèles pour établir une liaison entre elles en créant deux coudes. Dans ce cas la consonne est comptée trois fois, une fois dans chacune des deux rangées ainsi reliées, et une seule fois dans la rangée de sa duplication sauf à assurer un minimum de 1.

Le calcul de la dispersion sur la partie médiane se fait comme pour la consonne initiale, en renvoyant au fond de carte normalisé

.
a	.	ai	ao	.	.
.	e	ei	.	.	.
.	ia	ie	i	i.o	iu.
.	.	.	o	ou.	.
ua	ue	u.i	uo	u	.
.	ue	.	.	.	u
-- pas de voyelle -- 'er' --					

avec les trois différences suivantes:

- aux lignes et aux colonnes servant à couvrir le spectre, on ajoute la diagonale descendante ('a'...'ü')
- pour les lignes, les colonnes et cette diagonale, on utilise l'échelle 1, 1,5, 1,75, 1,875, 1,9375
- au lieu d'une pénalité de 8 pour une intersection en l'absence d'un coude, on ajoute:
 - 2 en présence d'une consonne finale 'n' ou 'ng' commune (par exemple pour 'iong' and 'uang'),
 - 4 autrement (par exemple en présence de 'uo' et 'a')

Le calcul de la dispersion sur la consonne finale se fait simplement en comptant 0 si le spectre occupe une seule case, 1 pour deux cases et 1,5 pour les trois cases.

On notera que cette mesure privilégie la proximité entre consonnes initiales par les colonnes par rapport à celle par les lignes, reflétant une fréquence plus grande de la transmission de l'influence phonétique par les colonnes. De même les médianes isolées étant moins rares que les consonnes initiales, la pénalité qui leur est attachée est moindre, surtout si la présence effective d'une consonne finale en atténue l'incohérence.

un dernier histogramme basé sur la mesure de dispersion

La prise en considération de la mesure de dispersion donne lieu à un dernier histogramme, dont les détails sont

donnés en annexe.

488	203	108	70	85	80	48	24	34	19	11	9	7	5	1191
0	1	2	3	4	5	6	7	8	9	10	11	12	13	total

distribution de la mesure de dispersion phonétique des marques

Dans le résumé ci-dessus, chaque nombre n de 0 à 13 représente en fait l'intervalle de mesure [n, n+1[.

Tout en reconnaissant que ces données ne sont pas exemptes de toute erreur, elles dessinent toutefois une forme très nette de décroissance, qui ressemble à une distribution naturelle selon une loi de puissance. En l'occurrence la dispersion de 67% des marques ne dépasse pas 21% de la dispersion maximum observée, la distribution moyenne étant de 2,25.

Les marques dont la dispersion est la plus grande sont souvent des marques qui ont agrandi leur domaine d'influence phonétique par suite de la simplification [Lire le chinois - l'apport de l'analyse, déjà cité], comme 𠂇 (Dg-01), dont la dispersion est égale à 13,5, ou bien 𠂇 (Sg-01), avec sa variante 𠂇 (Si-01), dont la dispersion est égale à 11. D'un autre point de vue, ce sont des marques primitives très anciennes, déjà fortement stylisées, dont l'ancienneté, la simplicité et la fréquence d'emploi accroît à la fois l'attraction graphique et l'expansion phonétique. La réciproque toutefois n'est pas vraie car d'autres marques primitives, comme 𠂇 /rén (Bb-01), de dispersion nulle, ou 𠂇 / yú (Zs-01), de dispersion égale à 1, ont été pratiquement cantonnées à leur rôle sémantique, comme l'ont été par construction les marques primitives 'aphones', comme 𠂇 (As-01) et 𠂇 (Ra-01), pour lesquelles la notion de dispersion phonétique n'a pas de sens.

-5- Examen des différences entre phonologie et décodage

L'on trouvera plus une comparaison plus détaillée entre les leçons de la phonologie et les résultats de l'approche par décodage dans [La structure phonétique du chinois - quelques résultats originaux, Ph Coueignoux 23/04/19, non publié]. La terminologie concernant la phonologie est tirée de Lee Wai-Sum [Standard Chinese (Beijing), Lee Wai-Sum et Zee Eric, Journal of the International Phonetic Association. 33 (1): 109–112, 2003], traduite selon la présentation donnée en [Parlons Chinois, Zhitang Yang-Drocourt, 2007 L'Harmattan, pp 168]. Ce qui suit est destiné à donner des bases quantitatives aux résultats les plus remarquables de l'approche par décodage.

le caractère apico-alvéolaire de 'sh'

Pour la phonologie la consonne fricative initiale 'sh' est évidemment apico-palatale (rétroflexe), comme 'zh', 'ch' et 'r'. Par contre l'approche par décodage lui permet de jouer aussi le rôle d'une apico-alvéolaire (dentale), comme 'd' et 't' et 'n', sachant qu'il n'existe pas a priori de apico-alvéolaire fricative à lui opposer. On présente la série suivante, statistiquement significative, pour valider cette conclusion:

- 它 (Ag-09) avec ses dérivés duò (舵), tuó (驼), tuó (駝), tuó (跫), tā (它), shé (蛇)
- 首 (Ai-11) avec ses dérivés dào (道), shǒu (首)
- 采 (Am-02) avec ses dérivés tàn (深), shēn (探)
- 叔 (Bl-02) avec ses dérivés dū (督), shū (叔)
- 逆 (Dj-01) avec ses dérivés shuò (朔), nì (逆)
- 是 (Hr-03) avec ses dérivés dī (堤), dī (提), tí (提), tí (题), shì (是)
- 庠 (Jl-01)) avec ses dérivés dù (度), shù (庶)
- 石 (Js-01) avec ses dérivés tuò (拓), shuò (硕), shí (石), niǎn (碾)
- 弋 (Lb-01) avec ses dérivés dài (代), shì (式)
- 头 (Nu-01) avec ses dérivés dú (读), dú (渎), dú (犊), dú (黠), dú (牖), tóu (头), shí (实), shù (赎)
en appliquant la règle du grand-père (cf annexe sur l'histogramme de la distribution des coudes)
- 水 (Qg-01) avec ses dérivés dá (沓), tà (沓), shuǐ (水), niào (尿)
- 对 (Si-03) avec ses dérivés duì (对), shù (树)
- 聂 (Si-06) avec ses dérivés shè (摄), shè (慑), niè (蹶)
- 俞 (Ta-04) avec ses dérivés tōu (偷), shū (输)
- 爻 (Sm-01) avec ses dérivés tóu (投), shè (设)
- 兑 (We-02) avec ses dérivés duì (兑), tuì (蜕), tuō (脱), shuì (税), shuì (说), shuō (说)

série à laquelle on doit rattacher de nombreux autres cas où le rattachement de la consonne mobile 'sh' aux dentales ne s'impose pas de façon aussi évidente mais se justifie parce qu'il permet de minimiser la dispersion phonétique, soit par migration, soit par duplication:

- 上 (Bg-01) avec ses dérivés tǎn (忐), shàng (上), shang (上), ràng (让)
- 寺 (Cb-08) avec ses dérivés shī (诗), shī (诗), shī (诗), à côté
des dérivés zhì (峙), zhì (痔), chí (持) d'une part,
des dérivés dài (待), dēng (等), tè (特) de l'autre

- 召 (Ei-02) avec son dérivé shào (绍), à côté
 - des dérivés zhāo (招), zhāo (昭), zhǎo (沼), zhào (召), zhào (诏), chāo (超) d'une part,
 - des dérivés diào (貂), tiào (迢), tiào (筲) de l'autre
- 寿 (Du-03) avec son dérivé shòu (寿), à côté
 - des dérivés zhù (铸), chóu (畴), chóu (踬), chóu (筹) d'une part,
 - des dérivés dǎo (祷), tāo (涛) de l'autre
- 也 (Hd-01) avec ses dérivés de (地), dì (地), tā (他), tā (她), tā (祂), tā (牠), tuō (拖), shī (施)
- 以 (Ia-02) avec ses dérivés sì (似), shì (似), nǐ (拟)
- 失 (Ns-01) avec ses dérivés shī (失), shū (殊), à côté
 - du dérivé zhì (秩) d'une part,
 - des dérivés diē (跌), dié (迭), tiě (铁) de l'autre
- 合 (Tb-01) avec ses dérivés dā (荅), dā (答), dá (答), shí (拾), ná (拿)
- 舌 (Wj-01) avec ses dérivés tián (甜), tián (恬), shé (舌), shì (适)
- 真 (Xu-02) avec son dérivé shèn (慎), à côté
 - des dérivés zhēn (真), zhěn (缜), zhèn (镇), chēn (嗔), chēn (瞋) d'une part,
 - des dérivés diān (颠), tián (填) de l'autre
- 单 (Zl-01) avec son dérivé shàn (禅), à côté
 - des dérivés dān (单), dān (殫), dǎn (掸), dàn (憚), dàn (殫), tán (弹) d'une part,
 - des dérivés chán (禅), chán (婵), chán (蝉), chǎn (阐) de l'autre

On notera que Karlgren [opus cité] accepte cette analyse pour les caractères (repris sous leur forme simplifiée) 它 (K1011), 叔 (K908), 弋 (K178), 聂 (K667), 兑 (K1138). Par contre l'accord en cas de duplication ne peut être significatif puisque Karlgren se contente de donner les dérivés phonétiques sans justifier la nature exacte de leur filiation. On rapprochera enfin les cas de 首 (K898) et de 寿 (K899). Si pour le premier Karlgren mentionne le son 'shou' et omet le son 'dao', contrairement à l'approche par décodage, il inclut bien ces deux sons dans l'influence phonétique plus étendue du deuxième.

le caractère prédorso-palatal de 'l'

La phonologie exclut de rapprocher la consonne initiale 'l', considérée comme apico-alvéolaire, des prédorso-palatales 'j', 'q', 'x' comme le demande l'approche par décodage. On présente la série suivante, statistiquement significative, pour valider cette conclusion:

- 见 (Mz-01) avec ses dérivés jiàn (见), xiàn (现), liàng (靛)
- 东 (Pp-01) avec ses dérivés jiǎn (拣), liàn (炼), liàn (练)
- 柬 (Pr-02) avec ses dérivés jiǎn (柬), jiàn (谏), lán (阑)
- 林 (Pu-01) avec ses dérivés jīn (禁), jìn (禁), lín (林), lín (淋), lín (霖), lín (琳)
- 京 (Qm-01) avec ses dérivés jīng (京), jīng (惊), jīng (鲸), jǐng (景), qióng (琼),
liáng (凉), liàng (凉), liàng (谅), liàng (晾)
- 昔 (Rf-03) avec ses dérivés jiè (借), què (鹊), xī (昔), xī (惜), là (蜡), là (腊), liè (猎)
- 兼 (Rx-01) avec ses dérivés jiān (兼), qiān (谦), qiàn (歉), xián (嫌), lián (廉)
- 各 (St-01) avec ses dérivés jiù (咎), lào (烙), lào (酪),
lù (路), lù (赂), lù (辂), lüè (略), luò (络), luò (骆), luò (洛)
- 亼 (Ta-01) et ses variantes [△] et [△] (Mf-02) avec ses dérivés jīn (今), jīn (金), qín (禽),
lǐng (令), ling (令)
- 𠂔 (Ta-02) avec ses dérivés jiǎn (捡), jiǎn (俭), jiǎn (脸), jiǎn (检), jiǎn (硷), jiǎn (剑),
qiān (签), qiān (筌), xiǎn (醜), liǎn (脸), liǎn (敛), liǎn (硷), yàn (险)
- 立 (Tf-01) avec ses dérivés qì (泣), lā (拉), lā (垃), lá (拉), lì (立), lì (粒), lì (笠)
- 肋 (Yg-02) avec ses dérivés jīn (筋), lèi (肋)
- 监 (Xx-03) avec ses dérivés jiān (监), lán (篮), lán (蓝), lán (褴), làn (滥)

On notera que Karlgren [opus cité] accepte cette analyse pour les caractères (repris sous leur forme simplifiée) 柬 (K372), 林 (K555), 京 (K392), 兼 (K374), 各 (K411), 立 (K524), 监 (K376), en faisant l'hypothèse que le voisinage entre 'l' et les prédorso-palatales reconnues est la trace d'un son ancien qui jouerait le rôle d'un ancêtre commun. Autrement dit 'j' ne serait pas frère avec 'l' comme avec 'q' et 'x', mais simplement cousin.

On notera que malgré tout Karlgren qualifie la dispersion phonétique de 𠂔 sur toutes les consonnes de la colonne 'j'-'q'-'x'-'l'-'y' comme "très irrégulière" (K1078), alors que l'approche par décodage la trouve tout à fait normale sinon par le nombre maximum de consonnes concernées.

la cohérence de la famille 'm'-'h'-'r'-'n'-'l'

La phonologie rapproche bien les consonnes 'n' et 'l' comme apico-alvéolaires mais isole sur quatre modes d'articulation différents la paire de nasales 'm'-'n' d'une part, la fricative 'h' et deux autres consonnes initiales 'r' et 'l' de l'autre, alors que l'approche par décodage propose d'en faire une seule famille.

Ce dernier point n'est pas en soi un obstacle car le décodage ne distingue pas nécessairement les modes

d'articulation. C'est ainsi qu'il place sur la même ligne les consonnes 'b', 'g', 'z', 'zh', 'd' et 'j', en confondant les occlusives et les affriquées que distingue pourtant la phonologie, confusion rendue possible car elle n'entraîne pas de collisions, chaque consonne se distinguant par un lieu d'articulation qui lui est propre.

Il n'en est pas de même avec la famille 'm'-h'-r'-n'-l'. Si l'on cherche à confondre les quatre modes d'articulation concernés, on butte sur le fait que le 'l' et le 'n' se disputent alors la même case apico-alvéolaire et que le 'h' est déjà pris par la famille des fricatives non voisées (fricatives).

Ces obstacles de principe sont levés si l'on accepte que le 'h' puisse recevoir un double caractère, à la fois fricative non voisée et membre de la famille composite 'm'-h'-r'-n'-l', comme le 'sh' s'est vu attribué un double rôle comme apico-alvéolaire et apico-palatale, et si l'on se rappelle que 'l' s'est vu rapproché de la famille des prédorso-palatales.

A cela on rattachera la réinterprétation de la syllabe pinyin 'er' comme représentant la consonne initiale isolée 'r'.

On sait que les Chinois utilisent la consonne initiale 'l' pour transcrire les noms propres étrangers qui comportent un 'r', utilisant par exemple 玛丽 / mǎlì pour 'Marie'. Attention, ce rapprochement entre le 'r' et le 'l' n'est pas une preuve de cohérence de la famille phonétique chinoise 'm'-h'-r'-n'-l'. C'est simplement la conséquence du fait que, les sons 'r' et 'l' étant rapprochés par les langues occidentales comme étant toutes deux des liquides [Précis de grammaire grecque, Ch. Maquet et F. Flutre, Hachette, 1925], on peut légitimement approcher le son occidental 'r', absent des prononciations chinoises, par le son 'l' commun aux deux registres. Mais ceci ne concerne en rien le son du Mandarin tel que le transcrit le pinyin à l'aide de 'r' ou 'er' et que les occidentaux ont d'ailleurs du mal à prononcer.

On présente la série suivante, statistiquement significative, pour valider cette conclusion:

- 貳 (Ab-02) avec ses dérivés èr (貳), nì (膩),
- 任 (Bb-02) avec ses dérivés rěn (荏), rèn (任), lìn (赁)
- 牛 (Cj-01) avec ses dérivés móu (牟), mǔ (牡), mù (牧), niú (牛), láo (牢)
- 乃 (Ek-01) avec ses dérivés nǎi (奶), réng (仍), nǎi (乃), nǎi (奶)
- 万 (Ep-01) avec ses dérivés wàn (迈), lì (厉),
- 蓼 (Es-02) avec ses dérivés liǎo (蓼), liáo (寥), lù (戮)
- 矛 (Fi-02) avec ses dérivés máo (矛), máo (茅), róu (柔)
- 弱 (Gh-03) avec ses dérivés ruò (弱), nì (溺)
- 乚 (Ha-01) avec ses dérivés rǔ (乳), lǐ (礼), luàn (乱)
- 爰 (Iu-07) avec ses dérivés huǎn (缓), nuǎn (暖)

- 卯 (Ka-05) avec ses dérivés mǎo (卯), liáo (聊), liǔ (柳)
- 令 (Kb-02) associant, via la marque 令 (Mf-05) dont elle est une variante, mìng (命) avec lìng (令)
- 尧 (Lc-02) avec ses dérivés ráo (饶), rào (绕), náo (挠)
- 户 (Jp-01) avec ses dérivés hù (户), hù (护), hù (扈), lì (戾), lú (庐), lú (炉), lú (芦), lú (驴)
- 儿 (Mk-01) associant, via la marque 兒 (Xl-03) dont elle est la simplification, ér (儿) avec ní (霓)
- 米 (Pe-01) avec ses dérivés mǐ (咪), mī (眯), mí (迷), mí (眯), mǐ (米), lèi (类), liào (料)
- 尔 (Qc-01) avec ses dérivés mí (弥), ěr (尔), nǐ (你)

auxquels on a adjoint luó (覩) pour appliquer la règle du chemin submergé

(cf annexe sur l'histogramme de la distribution des coudes)

- 少 (Qd-01) avec ses dérivés miǎo (秒), miào (妙), liè (劣)
- 而 (Qx-01) avec ses dérivés ér (而), rú (儒), rú (蠕), nài (耐), nuò (糯), nuò (懦)
- en appliquant la règle du grand-père (cf annexe sur l'histogramme de la distribution des coudes)
- 若 (Ra-09) avec ses dérivés rě (惹), ruò (若), nì (匿), nuò (诺)
- 女 (Su-01) avec ses dérivés rú (如), rǔ (汝), nèn (嫩), nú (奴), nǚ (女), lóu (娄)
- 文 (Sx-01) avec ses dérivés mǐn (闵), lìn (吝), liú (刘)
- 亨 (Wn-01) avec ses dérivés háo (豪), háo (毫), liàng (亮)
- 目 (Xm-01) avec ses dérivés méi (眉), mù (目), mù (睦), lèi (泪)
- 耳 (Xr-01) avec ses dérivés mǐ (弭), ěr (耳), ěr (饵), róng (茸), lián (联)
- 每 (Yk-01) avec ses dérivés méi (霉), méi (酶), méi (梅), méi (莓), méi (镁), měi (每), mǐn (敏),
hǎi (海), huǐ (悔), huǐ (诲), huì (晦)
- 田 (Za-01) avec ses dérivés miáo (苗), mǔ (亩), nán (男)
léi (雷), léi (累), lěi (累), lèi (累), lǐ (里), lì (里)
- 果 (Za-07) avec ses dérivés huái (踝), huǒ (夥), luǒ (裸)
- 里 (Zm-01) avec ses dérivés mái (埋), mán (埋),
lí (厘), lí (狸), lì (里), lǐ (理), lǐ (俚), lǐ (鲤), lì (里), liáng (量), liàng (量)

série à laquelle on rattachera le cas où le rattachement de la consonne mobile 'h' à la famille 'm'-'r'-'n'-'l', se justifie parce qu'il permet, par migration, de minimiser la mesure de dispersion d'un spectre:

- 虍 (Ji-01) avec ses dérivés hǔ (虎), nüè (虐), lǔ (虏), lù (虑),

voire de donner sa cohérence au spectre composite exceptionnel suivant, issu de la simplification:

- 又 (Sg_01) / 𠄎 (Si-01) avec ses dérivés huān (欢), nán (难),

Le cas de 每 est ambigu dans la mesure où l'on peut, à dispersion phonétique égale, faire appel au 'w' pour faire le lien entre le 'm' et le 'h' sans migration du 'h' (voir en ce sens la carte phonétique donnée à la fin de l'annexe sur les "exemples de calcul de dispersion phonétique"). Mais le nombre significatif de caractères commençant par 'm' et 'h', respectivement 7 et 4, comparé au caractère unique portant l'initiale 'w' (侮) plaide en faveur d'un rapprochement direct entre 'm' et 'h'.

Le cas de 𠄎 est aussi ambigu dans la mesure où l'on peut vouloir rattacher le 'h' comme de coutume à la consonne initiale 'x', si l'on accepte l'accroissement qui en résulte de la dispersion phonétique. De plus, au delà de ce cas particulier, beaucoup des marques ci-dessus ne sont pas significatives en elles-mêmes. On ne peut pas en général dans un spectre phonétique prendre en compte plus de deux exemples d'une consonne initiale et souvent il ne contient pas plus de deux consonnes initiales de la famille 'm'-'h'-'r'-'n'-'l'.

C'est en fait l'accumulation de tels exemples partiels qui tranche sur le hasard. Là où la phonétique rapproche simplement le 'n' d'une part au 'm', comme nasale, de l'autre au 'l', comme apico-alvéolaire, l'approche par décodage révèle des rapprochements complémentaires, d'abord entre le 'm' et le 'l' (13 exemples), déterminant une première famille 'm'-'n'-'l', et ensuite entre ce nouveau regroupement 'm'-'n'-'l' et, d'une part, le 'h' (7 exemples) et, de l'autre, le 'r' (13 exemples), ce qui permet de synthétiser toute la famille.

On notera que Karlgren [opus cité] accepte cette analyse, même si c'est parfois partiellement, pour les caractères (repris sous leur forme simplifiée) 貳 (K9), 任 (K935), 乃 et 尔 (K13 et K 14), 蓼 (K546), 儿 (K12), 而 (K10), 若 (K938), 女 (K675), 每 (K607), 果 (K479). Comme pour le rapprochement entre 'l' et les prédorso-palatales, Karlgren suppose un ancêtre commun pour expliquer le rapprochement, pour lui inattendu, entre le 'h' et le 'm' dans le spectre de 每, de même qu'entre le 'h' et le 'l' dans le spectre de 果. Le 'h' serait un cousin, plutôt qu'un frère de 'm' et de 'l',

-6- Conclusion

En dépit des inévitables erreurs qui sont encore à trouver dans une compilation de cette ampleur, cet article donne au fond de carte phonétique présenté dans [Décodage des caractères chinois, déjà cité] et rappelé ci-dessus dans la section 'calcul de la mesure de dispersion' les justifications quantitatives qu'appelaient les différences constatées avec les leçons de la phonologie. C'est aux spécialistes de cette dernière qu'il convient de trouver une explication pertinente, sans doute en approfondissant les hypothèses de Karlgren sur l'évolution de

la prononciation des dialectes chinois au cours de leur longue histoire.

Mais quelle que soit cette explication, le fond de carte proposé fait de l'état actuel de l'influence des marques sur les sons des caractères chinois une synthèse qui se veut optimale et qui, pour la majeure partie des sons et des marques, en donne une visualisation suffisamment simple pour aider à la mémorisation des caractères, spécialement pour leur écriture.

-7- Annexes

histogramme de la distribution des marques sur les sons

nombre des représentations d'un son	nombre de cas (exceptions incluses)	nombre de cas (exceptions exclues)	nombre de cas (marques exclues)	
0	2	2	297	
1	43	46	86	
2	33	40	17	
3	45	42	4	
4	46	52	1	
5	45	43		
6	30	28		
7	26	29		
8	23	18		
9	21	22		
10	18	19	total des sons	405
11	18	15		
12	9	7		
13	5	5		
14	6	5		
15	4	6		
16	8	5		
17	3	3		
18	2	2		
19	2	4		
20	4	0	sons concernés	
21	0	0	(caractères exceptions	
22	0	1	inclus	exclus)
23	2	2	li - xian	xian- yan
24	1	3	zhi	jian - yu- zhi
25	1	1	yu	xi
26	2	1	jian - yan	qi
27	2	0	qi - xi	
28	0	0		
29	1	1	shu	shu

30	0	0		
31	0	0		
32	0	0		
33	0	0		
34	0	0		
35	0	0		
36	0	0		
37	0	1		shi
38	1	0	shi	
39	1	1	yi	yi
40	0	0		
41	0	1		ji
42	0	0		
43	0	0		
44	1	0	ji	
total des sons	405	405		

Le total des sons pinyin dépend des auteurs. Le tableau ci-dessus inclut deux sons rares 'yai' et 'zhei' pour lesquels le corpus ne contient d'ailleurs pas de caractères.

Il est important de rappeler que ces chiffres sont avant tout indicatifs car ils dépendent de conventions que l'on peut interpréter de façon plus ou moins objective. En l'occurrence il faut signaler quatre sources de variations:

- les variations graphiques qui se ramènent à la même marque principale comme 羊, 𦍋, 𦍌
- l'influence phonétique limitée de certaines marques principales comme 丩 et 𠂇
- l'influence phonétique limitée de certaines marques secondaires comme 江
- l'incohérence phonétique de quelques marques issues de la simplification comme 办

De façon plus précise, l'histogramme ci-dessus ne tient pas compte des variations graphiques. Par exemple les deux variantes 𠂇 (Sa-01) et 𠂈 (Sb-01) qui influencent le même son 'yi' respectivement dans les caractères 刈 et 义 ne sont comptées que pour une unité dans la liste du son 'yi'. L'histogramme omet de plus l'influence de quelques marques au profit de descendants ou d'ascendants plus prolifiques.

C'est le cas de plusieurs marques principales primitives, c'est à dire qui ne contiennent pas d'autre marques qu'elle-même, lorsque leur influence s'exerce par l'intermédiaire d'un descendant phonétique unique comme 𠂇 (Aq-01) sur le son 'qian'. La liste de 'qian' est limitée à 𠂈 (Ta-02) et exclut son ascendant 𠂉. C'est encore le cas de 丩 (An-01) vis à vis du son 'xiao', au vu de son descendant 肖 (Ym-01). On notera que ce dernier cas revient en fait à accepter a posteriori la marque 丩 comme une variante graphique de la marque 小 (Qb-01) alors

qu'elles ont été prises comme suffisamment distinctes sur le plan graphique pour mériter d'être traitées séparément comme le sont les paires 人(Ma-01) / 亻 (Bb-01) et 水(Qg-01) / 氵 (As-01).

C'est aussi le cas de certaines marques secondaires qui n'admettent pas d'autre descendant phonétique qu'elle-même, comme 江, jiāng (As-02), le son 'jiang' étant alors représenté par son ascendant phonétique 工, gōng (Ce-01).

Enfin quatre marques, 办/辦, bàn (En-14), 开/開, kāi (Rn-01), 隶/隸, lì (Ur-02) et 蓄, sè (Cb-03), ont la particularité de porter des sons incompatibles avec leur famille adoptive de dérivées phonétiques.

- pour 开, 'xing' (形, 刑) et 'yan' (研, 妍)
- pour 办, 'xie' (协, 胁), héritée de la marque 劦 (xié) elle aussi simplifiée en 办,
- pour 蓄, 'qiang' (墙, 蔷), héritée de la marque 冫 (qiáng) éliminée de 牆,
- pour 隶, 'dai' (逮), héritée de l'ancienne marque 隸 (dài).

On a rendu compte de cette incohérence en éliminant les caractères et les sons concernés des dérivées phonétiques catalogués dans les fichiers des marques 开, 办, 蓄 et 隶. Mais on a quand même fait le choix de les compter comme des marques dans la liste des représentations de leur propre son alors qu'on aurait pu tout aussi bien les prendre comme des caractères faisant exception.

La dernière colonne donne la distribution des caractères faisant exception à la règle, c'est à dire des caractères qui ne sont pas des marques en eux-mêmes et n'ont donc pas de descendance phonétique mais n'ont pas non plus d'ascendance phonétique selon une marque constitutive, par exemple 好 (hǎo). On voit que pour les trois quarts des sons (297), l'influence des marques est suffisante pour expliquer la prononciation de tous les caractères concernés, le nombre total des caractères exceptions étant de $1 \times 86 + 2 \times 17 + 3 \times 4 + 1 \times 4 = 136$

Le tableau mentionne explicitement les douze sons qui sont représentés par le plus de marques et de caractères. Cela permet aussi de mieux comprendre comment s'articulent entre elles les colonnes de ce tableau. Ainsi le son le plus diffus ou prolifique, 'ji', est représenté par 44 signes, soit 41 marques et 3 caractères par exception. Il compte donc pour 1 respectivement sur les lignes 44 (colonne caractères inclus), 41 (colonne caractères exclus) et 3 (marques exclus). Par contre le son 'yi' n'a pas d'exception et compte donc pour 1 sur la ligne 39 aussi bien pour la colonne caractères inclus pour que celle caractères exclus, et pour 1 sur la ligne 0 pour la colonne des exceptions (marques exclus).

histogramme de la distribution des sons sur les marques

nombre des sons d'une marque	nombre de cas
0	98
1	499
2	317
3	151
4	83
5	46
6	17
7	13
8	13
9	10
10	2
11	0
12	2
13	2
14	1
15	0
total des marques	1254

Encore une fois les chiffres ci-dessus ne sont pas à prendre à leur valeur exacte pour les raisons exposées dans la section précédente.

98 marques ne correspondent à aucun son, rassemblent trois types différents.

- soit quatre marques 'vides' qui ne sont pas des caractères et n'ont aucun caractère dérivé direct. Les trois premières sont admises pour des raisons purement historiques, à savoir 冫 (Fa-01), 辶 (Nd-01), et 鬥 (U1-01).

Bien qu'elle soit 'vide', la quatrième apporte malgré tout une origine phonétique commune à deux marques qui ne sont toujours pas des caractères mais qui à leur tour transmettent cette influence à leurs propres descendants;

à savoir 冫 (Bd-01), la simplification de 臣 (Uc-01) en composition dans les marques 冫 (Bd-02) et 冫 (Sg-04), dont dérivent in fine des caractères usuels comme 覽 (lǎn), 藍 (lán) et 堅 (jiān),

- soit des marques de rôle purement sémantique, comme 彳 (As-01) et 艹 (Ra-01), clés traditionnelles,

- soit des marques qui sont ou ont été des caractères, avec un son propre, mais qui ne figurent pas dans le corpus en tant que caractère et qui n'y ont pas non plus de dérivés phonétiques, comme 冉, rǎn (Zj-01).

499 marques n'influencent qu'un seul son, regroupent deux types.

- les marques dites 'stériles' du point de vue phonétique car elles n'ont, à part elles-mêmes, aucun descendant phonétique dans le corpus, comme 𠄎 / yuē (Xj-01), groupe dans lequel se rangent les pseudo-marques admises uniquement pour indexer les caractères correspondants (cf ci-dessus la section sur le 'nombre de marques'), comme 𠄎 / shū (Ba-03),

- les marques 'monosons' qui ont une vraie descendance sur un son unique, comme 星 / xīng (Cp-01).

Lorsqu'une marque influence deux sons, ces deux sons, sauf rare exception, commencent soit par la même consonne, soit par deux consonnes phonétiquement voisines, sur la même colonne ou la même ligne du fond de carte phonétique caractéristique de l'écriture chinoise [Décodage des caractères chinois, déjà cité]. C'est le cas de 𠄎 (Kh-01) avec les sons 'min' et 'mian', 包 (Eg-01) avec les sons 'bao' et 'pao' et de 左 (Jt-02) avec les sons 'zuo' et 'duo'.

histogramme de la distribution des 'coudes phonétiques' sur les marques

[On se reportera au tableau présenté en mode paysage dans un fichier séparé pour des raisons de mise en page.]

Cet histogramme est basé directement sur les cartes phonétiques de chaque marque et non pas, comme dans les deux histogrammes précédents, sur la liste donnant, pour chaque son, les marques qui l'influencent. On ne peut donc pas le comparer directement avec celui de la dispersion des sons sur les marques. En effet le nombre de sons des dérivés n'a que peu de rapport avec le nombre de coudes. Et de nombreuses marques qui ne sont pourtant pas 'aphones' sont malgré tout mentionnées dans la liste des sons comme influençant 0 son, soit parce qu'il n'y a pas de caractère dérivé portant son influence phonétique, soit parce que cette influence a été carrément omise au profit d'un ascendant ou descendant phonétique plus important. On renvoie aux commentaires donnés dans les deux annexes précédentes.

Les lignes de ce tableau correspondent aux 24 familles présentées sur les colonnes du tableau des marques principales (cf ci-dessus la section 'essai de classification graphique').

Les titres des colonnes de ce tableau appellent les commentaires suivants:

- pour la définition des marques dites 'variantes' et 'dupliquées', voir la section 'essai de classification graphique'.
- on appelle 'aphones' les marques qui n'ont jamais été des caractères et n'ont jamais reçu de prononciation
- pour la définition des marques dites 'stériles' et 'monosons', voir la section précédente.
- on appelle 'simples' les marques multisons dont les consonnes initiales s'étendent sur une seule ligne ou une seule colonne de la matrice des consonnes initiales sans recourir à un coude. Cette catégorie inclut les marques qui n'influencent que deux sons.
- les quatre titres suivants, à savoir 'unicoudes', 'bicoudes', 'tricoudes' et 'multicoudes' parlent d'eux-mêmes

marque tricoude: 寸 (Dg-01)

b	g	z	zh	d	j	a	.	ai	ao	.	.
p	k	c	ch	t	q	.	e	ei	.	.	.
f	h	s	sh	.	x	.ia	ie	i	i.o	iu	.
m	.	.	r	n	l	.	.	.	o	ou	.
w	y	ua	ue	u.i	uo	u	.
- no initial consonant -						.	üe	.	.	.	ü
.	-- er -- no vowel --					

carte phonétique de

寸

dispersion 13,5
densité 1,26

Il n'existe qu'une seule marque ayant plus de trois coudes, 又 (Sg-01), avec sa variante 𠄎 (Si-01). Cette marque a en effet hérité par simplification de pas moins de trois autres marques, 𠄎, 莫 et 董. La carte phonétique résultante en porte témoignage mais reste cohérente, avec quatre coudes, respectivement en q ('qu' (取), 'quan' (权 et 劝)) qui articule entre les sons originaux 'you' (友 et 又) et 'yu' (驭) et le son 'tan' (叹), en t qui articule entre les sons 'qu' et 'quan' et le son 'nan' (难), le n qui articule entre le son 'tan' et le son 'han' (汉) et enfin le h qui articule entre le son 'nan' et le son 'guan' (观).

b	g	z	zh	d	j	a	.	ai	ao	.	.
p	k	c	ch	t	q	.	e	ei	.	.	.
f	h	s	sh	.	x	.ia	ie	i	i.o	iu	.
m	.	.	r	n	l	.	.	.	o	ou	.
w	y	ua	ue	u.i	uo	u	.
- : pas de consonne initiale : -						.	üe	.	.	.	ü
.	-- pas de voyelle -- 'er' --					

carte phonétique de

又 / 𠄎

dispersion 11
densité 0,91

Enfin les six marques suivantes présentent des incohérences.

Le son de 虬 sī (Ji-06) n'est pas voisin de son seul dérivé 虬 chǐ.

Le son de 夾 jiā (Na-02) n'est pas voisin de son seul dérivé 映 shǎn.

Le spectre phonétique de 妥 (Su-07) comporte deux sous-ensembles, respectivement sur les sons 'tuo' (妥) et 'nei' (馁) d'une part et de l'autre sur le son 'sui' (绥), sans qu'il soit possible d'établir un lien entre les deux.

Le spectre phonétique de 串 (Tx-01) comporte deux sous-ensembles, respectivement sur les sons 'chuan' (串) et 'huan' (患) d'une part et de l'autre sur le son 'cuan' (窜), sans qu'il soit possible d'établir un lien entre les deux.

Le spectre phonétique de 册 (Yt-01) comporte deux sous-ensembles, respectivement sur les sons 'zha' (栅) et

'shan' (珊, 刪 et 珊) d'une part et de l'autre sur le son 'ce' (册), sans qu'il soit possible d'établir un lien entre les deux.

Le spectre phonétique de 皆 (H1-04) comporte deux sous-ensembles, respectivement sur les sons 'jie' (皆) et 'xie' (偕 et 諧) d'une part et de l'autre sur le son 'kai' (揩 et 楷), sans qu'il soit possible d'établir un lien entre les deux.

A ces six exceptions on peut rattacher si l'on veut les quatre marques mentionnées dans l'annexe ci-dessus intitulée 'histogramme de la distribution des marques sur les sons', à savoir 办/辦, bàn (En-14), 开/開, kāi (Rn-01), 隶/隸, lì (Ur-02) et 嗇, sè (Cb-03), qui peuvent être présentées comme des 'erreurs' de simplification. Elles ont été régularisées en isolant arbitrairement de sa famille phonétique le son propre à la marque.

On peut naturellement chercher à résoudre ces quelques incohérences. Il suffit pour cela d'éliminer les cas de son singulier en leur déniaient tout lien avec le reste de la carte phonétique (cas de 'chi' par rapport à 'si' pour 屮, cas de 'sui' par rapport au groupe 'tuo'-'nei' pour 妥, cas de 'huan' par rapport au groupe 'chuan'-'cuan' pour 串) et d'adjoindre le cas de 夾 et de 册 à celui des trois marques 'en erreur' pour lesquels le son original de la marque est tenu comme étant dissocié de sa famille phonétique. Mais le cas de 皆 présente deux groupes et il est difficile de justifier d'en éliminer un. Et de toute façon, il faut savoir admettre que tout système engendré par l'activité humaine peut toujours violer ses propres règles, a fortiori des règles implicites qui sont supposées régir l'activité en question.

Par contre on renvoie le lecteur à [La structure phonétique du chinois - quelques résultats originaux, déjà cité] pour la façon dont d'autres incohérences apparentes sont légitimement expliquées par deux règles simples:

- la règle dite du 'chemin submergé' fait appel à des caractères hors du corpus pour rétablir une connection manquante sur la carte phonétique, par exemple le spectre de 爰 (If-07) s'articule autour de la consonne initiale 'q' basée sur le caractère rare 爰 (qūn), faisant coude entre les sons 'jun' (俊 ...) et 'cun' (皃),
- la règle dite du 'grand-père' permet à l'influence phonétique de sauter une génération et de justifier de l'influence phonétique d'une marque sur un caractère dérivé d'une marque intermédiaire elle-même dérivée de la marque initiale, par exemple 数 (shù) est influencé par son grand-père 女(nǚ) (Su-01) et non pas par son ascendant direct 娄 (lóu) (Pe-03).

exemples de calcul de dispersion phonétique

Ce calcul, qui ne concerne pas les primitives aphones, donne naturellement une dispersion nulle à tous les

monosons, comme les exemples déjà cités 日 / yuē (Xj-01), 书 / shū (Ba-03) et 星 / xīng (Cp-01).

Au delà des monosons, le cas le plus simple est illustré par 冂 (Uk-01), de dispersion 1 :

					.							.
b	g	z	zh	d	j	a	.	ai	ao	.	.	.
p	k	c	ch	t	q	.	e	ei	.	.	.	-
f	h	s	sh	.	x	.ia	ie	i	i.o	iu.	.	.
m	.	.	r	n	l	.	.	.	o	ou.	n	.
w	y	ua	ue	u.i	uo	u.	.	.
- no initial consonant -						.	üe	.	.	.	ü	ng
.	-- er -- no vowel --						.

carte phonétique de
冂
dispersion 1
densité 8

On rappelle que la densité est le rapport entre le nombre de caractères du corpus dérivés de la marque et sa mesure de dispersion, soit, pour 冂, respectivement huit et un.

Pour un cas simple avec une série plus longue de consonnes initiales sur la même colonne, voici 𠄎 (Ta-02):

												.
b	g	z	zh	d	j	a	.	ai	ao	.	.	.
p	k	c	ch	t	q	.	e	ei	.	.	.	-
f	h	s	sh	.	x	.ia	ie	i	i.o	iu.	.	.
m	.	.	r	n	l	.	.	.	o	ou.	n	.
w	y	ua	ue	u.i	uo	u.	.	.
- pas de consonne initiale -						.	üe	.	.	.	ü	ng
.	- pas de voyelle -- 'er' -						.

carte phonétique de
𠄎
dispersion 2,875
densité 4,52

On notera que l'aurait pu accepter de décomposer le son 'yan' en 'y'+ 'ia'+ 'n', ce qui réduirait ici la dispersion à la seule contribution des initiales, soit 1,875, une facilité que l'on pourrait aussi étendre au son 'wa' décomposable en 'w' + 'ua'. Il s'agit là d'une reconnaissance a posteriori d'une certaine interdépendance entre consonne initiale et partie médiane.

Pour les cartes phonétiques suivantes, on est prié de se reporter à l'annexe précédente.

Pour un cas simple qui fasse intervenir les trois composantes, 分 (Em-01) a une dispersion de 5, somme de

1,5 pour les initiales avec la colonne 'b', 'p', 'f'

2,5 pour les médianes avec la diagonale 'a', 'e', 'i', soit 1,5, et la colonne 'ai', 'i', soit 1

1 pour la colonne des finales avec '-' et 'n'

Enfin le cas de 每 (Yk-01) illustre la deuxième façon de dupliquer une consonne initiale 'mobile'

b	g	z	zh	d	j	a	.	ai	ao	.	.
p	k	c	ch	t	q	.	e	ei	.	.	.
f	h	s	sh	.	x	.	ia	ie	i	i.o	iu.
m	.	.	r	n	l	.	.	.	o	ou.	.
w	y	ua	ue	u.i	uo	u.	.
- pas de consonne initiale -						.	üe	.	.	.	ü
.	- pas de voyelle - 'er' -					

carte phonétique de

每

dispersion 5,75
densité 2,09

Sa dispersion de 5,75 est la somme de

3,00 (1+1+1) pour les initiales

avec les colonnes 'm'-w' et 'h'-w dupliqué' et la ligne 'w'-w dupliqué' affectée du 1 minimum

1,75 (1,75) pour les médianes avec la colonne 'ai'-ei-'i'-ui'

1 pour les finales avec '-' et 'n'

Le choix de dupliquer la consonne initiale 'w' n'était pas obligatoire. On aurait pu tout aussi bien faire migrer la consonne 'h' sur la ligne du 'm' pour utiliser la colonne 'm'-w' et la ligne 'm'-h déplacé'. Mais la mesure n'aurait pas changé, car les initiales compteraient toujours pour 3 (1+2). Seul le nombre de coudes aurait diminué de un.

histogramme de la mesure de dispersion phonétique

dispersion phonétique	nombre de cas	nombre de cas après regroupement
0	488	[0-1[488
0.5	0	
1	189	[1-2[203
1.5	14	
2	90	[2-3[108
2.5	18	
3	60	[3-4[70
3.5	10	
4	66	[4-5[85
4.5	19	
5	61	[5-6[80
5.5	19	
6	28	[6-7[48
6.5	20	
7	18	[7-8[24
7.5	6	

8	24	[8-9[34
8.5	10	
9	14	[9-10[19
9.5	5	
10	10	[10-11[11
10.5	1	
11	7	[11-12[9
11.5	2	
12	4	[12-13[7
12.5	3	
13	3	[13-14[5
13.5	2	
total des marques	1191	1191

On a d'abord tronqué la dispersion sur une décimale puis en supprimant la décimale. Ce faisant on voit que l'histogramme devient plus régulier et significatif.

Cet histogramme, comme le précédent, se base directement sur les cartes phonétiques de chaque marque. Il est donc légitime de les comparer. On constate alors deux légères différences:

- 63 marques laissées de côté (1254 - 1191) à comparer aux 54 'aphones'

- 488 marques de dispersion nulle à comparer au total 497 (106 + 391) des 'stériles' et des 'monosons'

Ces différences reflètent les erreurs qui ont été commises dans certains fichiers du dictionnaire des marques, erreurs humaines inévitables de la part de l'auteur et qu'il s'efforce de corriger mais qui n'entachent pas la force des conclusions.