

Décodage des caractères chinois

philippe.coueignoux@eprivacy.com

Ph Coueignoux 19/11/20

-0- Table des Matières

-1- Introduction:

*un problème posé par les caractères chinois
une façon originale de traiter le problème*

-2- Les marques comme éléments constitutifs d'un code *mise en évidence par répétition dans un corpus donné analyse du rôle par écart par rapport à la moyenne*

-3- Difficultés et aides présentées par le codage à retrouver *liberté du codeur pour représenter les sons du chinois nature ambiguë de cette représentation lien de proximité entre sons décomposition des sons*

-4- Traitement du problème par déchiffrement *points d'ancrage fond de carte phonétique et cartes phonétiques mesures de dispersion et de densité approche itérative illustration de l'approche itérative justification de l'approche itérative*

-5- Application de la méthode itérative *la règle fondamentale du codage phonétique application naturelle hypothèses vraisemblables*

-6- Conclusion

-1- Introduction

un problème posé par les caractères chinois

Présenté comme le résultat d'une suite ordonnée de traits, le caractère chinois reste pour beaucoup un dessin dont le calligraphe exprime la beauté formelle selon son inspiration artistique. Mais ceux qui étudient l'écriture chinoise perçoivent très vite l'évidence que la plupart des caractères sont aussi composés de signes qui réapparaissent dans plusieurs d'entre eux

On pourrait être tenté d'établir un parallèle avec l'écriture des langues occidentales basée sur l'alphabet latin. Pour prendre un exemple, au caractère chinois correspondrait le mot français et au signe répétitif dont on vient de parler correspondrait la lettre, signes et lettres étant eux-mêmes composés de traits. Ce serait une erreur à plusieurs titres.

En français, la décomposition d'un mot en traits est toujours analysable en deux étapes cohérentes, d'abord du mot à la lettre et ensuite de la lettre au trait. Autrement dit, pour un mot donné, la suite des traits se segmente toujours selon la suite des lettres, le premier trait d'une lettre suivant le dernier trait de la lettre précédente. En chinois par contre, un premier obstacle vient du fait que, s'il existe toujours un ordre des traits d'un caractère, la disposition bi-dimensionnelle des signes dans le caractère qu'ils constituent n'est pas susceptible d'être décrite selon un ordre naturel. Et surtout, même si l'on impose aux signes un début d'ordre, comme de haut en bas et de droite à gauche, il n'est pas rare d'observer des caractères pour lesquels la suite de ses traits est incohérente avec la suite de ses signes constitutifs. Dans le caractère 街 (jiē / rue), le tracé des traits du signe 行, que l'on retrouve dans le caractère 衡 (héng / peser), est interrompu par les traits formant le signe central 圭. Et rares sont les débutants qui reconnaissent dans le caractère 哀 (āi / la peine) le signe 衣 car là encore l'unité de ce dernier est rompu par les traits du signe central 口. Le trait fonde la lettre en français, il mine le signe constitutif en chinois. D'autre part, la lettre joue un rôle capital dans l'écriture d'un mot français car elle permet, au moins en théorie, de le prononcer correctement. Le français a en effet une écriture alphabétique qui utilise un petit nombre de signes, les lettres de l'alphabet, pour représenter tous les sons élémentaires de la langue parlée. Ceci n'est pas le cas des signes constitutifs de l'écriture chinoise, dépassant le millier et qui ont leur origine dans la représentation graphique stylisée ou abstraite de la réalité. Par exemple le signe 口 ci-dessus dessine une bouche et le signe 行 des traces de pas et ne jouent aucun rôle phonétique dans les caractères 哀 et 街 où nous les avons respectivement rencontrés.

N'ayant pas le statut central de la lettre, les signes répétitifs du chinois ont donc reçu une attention moindre dans l'étude et l'enseignement. On les réduit le plus souvent à des anecdotes grâce à leur pouvoir pictographique

évocateur indéniable, comme lorsque l'on met en valeur des flammèches dans le caractère 火 (huǒ / feu). Plus systématiquement un dictionnaire en utilise un sous ensemble pour faciliter sa consultation. Les signes retenus, appelés clés, y jouent surtout un rôle sémantique pour indexer des groupes de mots selon une classification naturelle. Le signe pour la bouche 口 sert par exemple à indexer des onomatopées comme 啊 (á,ǎ,à,a / ah) ou des cris comme 吠 (fèi / aboyer) ou encore des sentiments qui s'expriment oralement comme 嗔 (chēn, être en colère) ou 哀 (āi, la peine). Mais cette indexation est un objectif purement utilitaire qui n'hésite pas à rentrer en contradiction avec la structure réelle des caractères. Dans 哀 (āi, la peine), de nombreux dictionnaires isolent les deux premiers traits pour en faire sa clé 亠 alors que le caractère est composé des deux éléments 衣 et 口, pourtant eux-mêmes des clés.

Les signes constitutifs des caractères chinois ont néanmoins leur importance propre et en particulier ils jouent aussi un rôle phonétique. Ce fait, connu depuis toujours comme étant à l'origine de la grande majorité des caractères, dits idéo-phonogrammes (形声), est étudié depuis longtemps [Caractères chinois : étymologie, graphie, lexiques, Léon Wieger, 1899]. Seulement la façon dont les signes déterminent le son des caractères dont ils sont constituants est loin d'être évidente et surtout n'a pas de pouvoir déterministe prédictif. Plutôt que de détermination du son on parlera d'influence phonétique et, en l'absence de règles de prononciation, comme pour les lettres du français, on devra se contenter de parler de système de codage et reconnaître que le code sous-jacent est inconnu. Retrouver ce code est le problème traité dans cet article.

une façon originale de traiter le problème

Ayant assimilé la représentation phonétique dans les caractères chinois au résultat d'un code inconnu, la présente étude se présente en effet comme une entreprise de décodage selon les règles courantes du déchiffrement. On notera que la cible qu'il s'agit de décoder n'est pas l'enregistrement phonique brut des sons mais une autre représentation phonétique, qui utilise par contre un code complètement connu, le pinyin.

En partant des signes constitutifs des caractères chinois comme éléments de base du code recherché, on montrera d'abord comment espérer déchiffrer le code inconnu selon une approche classique en la matière, qui s'appuie sur l'analyse des écarts par rapport au hasard et l'on étudiera plus précisément le contexte très particulier dans lequel on prétend appliquer cette démarche.

On présentera alors les trois outils sur lesquels s'appuie la méthode de décodage proposée, à savoir les points d'ancrage, pour commencer l'analyse, les cartes phonétiques, pour représenter ses résultats au fur et à mesure, et les mesures de dispersion et de densité phonétiques, pour quantifier ces résultats.

On pourra alors définir comment appliquer ces outils selon une méthode progressant par itérations successives,

en illustrant le processus par celui que l'auteur a lui-même suivi. Ce déroulement dépend de la façon dont le décodeur choisit d'organiser les faits mais conduit à un résultat global unique et optimal, sous forme d'une règle simple de propagation de l'influence phonétique s'appliquant sur le fond de carte commun à toutes les cartes phonétiques.

On terminera par une critique interne de ce résultat, en renvoyant sa confrontation aux résultats déjà connus dans les domaines de la phonologie et de la pédagogie à d'autres articles.

-2- Les marques comme éléments constitutifs du code

Parce que leur présence donne en général des indications sur le sens ou le son du caractère où ils apparaissent, les signes constitutifs de l'écriture chinoise seront appelés 'marques' dans ce qui suit. On rapprochera ce terme de celui de 'marqueur', utilisé en linguistique pour désigner les morphèmes dont la fonction est grammaticale. Dans le contexte de cet article, la marque ressort d'un niveau d'analyse plus fin, intermédiaire entre le morphème et le trait, et sa fonction n'est plus grammaticale mais au niveau sémantique et phonétique ni aussi bien déterminante, comme celle du 's', marqueur du pluriel en français et en anglais, mais seulement indicative, voire facultative. Si l'on préfère éviter ce terme de 'marque', on pourra lui substituer 部件 (bùjiàn), dont le ministère de l'Éducation du gouvernement chinois a publié la liste officielle dans [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>], liste identique, à quelques variantes près, à celle adoptée pour cette étude. Parler de '部件' souligne le rôle premier, faire partie, parler de 'marque' le rôle second, faire signe.

En anticipant sur les résultats de cette approche qui ignore donc le niveau du trait et, on le verra, la notion de ton, l'on dira par exemple que le caractère 火 est un tout indivisible en parties, soit une seule marque signalant à la fois un sens, 'le feu', et un son, 'huo'. Par contre le caractère 灰 est composé de deux parties, la marque 扌, qui indique 'la main' mais n'a pas d'influence phonétique, et la marque 火 qui, en l'occurrence, joue exactement le même rôle qu'avant. La cendre est en effet du feu que l'on peut prendre en main. On notera toutefois que le son qu'elle représente, 'hui', n'est pas identique mais voisin. Dans cet article, on s'intéressera uniquement au rôle phonétique des marques et l'on se donnera pour but de préciser le sens qu'il faut donner à cette notion de vicinité, qui sans être l'identité n'en est pas pour autant arbitraire.

mise en évidence par répétition dans un corpus donné

La première étape dans un déchiffrement consiste à compiler un corpus de connaissance sans lequel aucune mesure, qu'elle soit statistique ou non, ne puisse être faite. En l'occurrence l'étude porte sur les caractères

mentionnés par le 'Pocket Oxford Chinese Dictionary, troisième édition, 2003, Oxford University Press. Plus précisément seuls les caractères simplifiés sont admis au corpus. Les auteurs ne donnant pas de chiffres, le corpus considéré est estimé compter entre 5 000 et 7000 caractères. Les résultats obtenus ne seront donc valides que pour ce choix arbitraire mais la démarche suivie pourra être appliquée à d'autres corpus.

La caractéristique essentielle d'une marque est d'apparaître de façon répétée dans le corpus étudié. On conçoit donc bien que, plus le corpus est large, plus l'ensemble des marques pris en considération est grand. D'un autre point de vue, tout signe rencontré dans un caractère, y compris ce caractère lui-même, mais que l'on ne retrouve dans aucun autre caractère du corpus reste une marque potentielle dans un corpus plus large. Même sans sortir du corpus choisi l'énumération des marques pose trois difficultés dont il faut avoir conscience.

Une première ambiguïté provient de la plasticité graphique du signe. Si l'on s'attachait à l'identité graphique absolue, le nombre de marques indépendantes exploserait car, ne serait-ce que pour des questions de mise à l'échelle dans l'espace fixe alloué à chaque caractère, la même forme géométrique varie par homothétie. On cherchera donc dans cette étude un compromis où certaines variations de forme sont ignorées, d'autres sont reconnues comme des signes distincts mais regroupées comme une seule marque, d'autres enfin sont prises comme des marques distinctes. Ce compromis ne peut se justifier que d'un point de vue subjectif mais sans affecter la méthode suivie ultérieurement.

Se référant à la rose des vents pour décrire la position relative des marques à l'intérieur du caractère qu'elles constituent, on dira par exemple que les marques à l'ouest des caractères 玻 et 环 sont identiques à la marque 王, même si leur forme est légèrement distordue par rapport aux marques à l'est de 狂 et 汪. Par contre on considèrera les marques du nord-ouest dans 差 et 羞 comme différant de la marque 羊, tout en renvoyant le fichier de 𠂇 à celui de 羊. où les listes de leurs dérivés sont confondues. Enfin les marques de l'ouest des caractères 法 et 汉 sont considérées comme des instances de la marque 彳, distincte de la marque 水, et leurs dérivées forment deux listes séparées dans leurs fichiers respectifs.

Un deuxième point est d'ordre théorique. De nombreux auteurs considèrent qu'une marque doit avoir eu une existence autonome au cours de l'histoire en tant que caractère même si elle n'apparaît plus qu'à l'intérieur des caractères du corpus choisi. Cette restriction de nature diachronique ne se justifie pas dans notre approche. D'ailleurs plusieurs marques, comme 彳 citée plus haut, et 灬, sont des clés reconnues par tous en tant que telles alors que leur forme stylisée les éloigne visuellement du caractère donné comme leur origine historique et n'apparaît qu'en composition.

La dernière difficulté est capitale. Le critère de répétition ne distingue pas à priori une marque des sous ensembles de traits qui la constituent, et en particulier de ses traits individuels. Si une marque est répétée, les traits qui en forment le signe graphique aussi. De fait, par souci d'efficacité globale en l'absence des capacités de l'informatique, les compilateurs de dictionnaires ont préféré ignorer la majorité des marques en les indexant par l'un de leurs traits. Dans notre approche, en présence d'une répétition dans un caractère donné, seul le plus grand ensemble de traits à avoir ce caractère répétitif sera pris comme marque. Avec cette règle, un trait peut toujours faire office de marque mais dans des cas beaucoup plus restreints. Prenons par exemple le caractère 谈 . Il est clair qu'il comprend la marque 火 . Mais comme cette marque est elle-même incluse dans la marque 炎 qui se compose 'd'un plus grand ensemble de traits', 谈 sera considéré comme un dérivé direct de 炎 et non pas de 火 et l'on ira de 火 à 谈 uniquement par dérivations successives. De même on se gardera de lire la marque 讠 dans 谈, car ce trait est simplement une partie de la marque 讠 qui a 'un plus grand ensemble de traits'. En définitive on lira uniquement les deux marques 讠 et 炎 dans le caractère 谈.

Constatant ainsi qu'une marque en tant que telle peut être un signe constitutif d'une marque plus complexe, on est conduit à repérer des liens de dérivation des marques entre elles selon un processus itératif d'addition qui crée des liens de descendance et d'héritage [Lire le chinois - l'apport de l'analyse, Ph Coueignoux 15/04/19, non publié]. Par exemple le caractère 礪 dérive de la marque 薄, qui descend de 溥, qui descend de 專, qui descend de 甫.

analyse du rôle par déviation par rapport à la moyenne

A partir du corpus choisi, et en tenant compte des réserves ci-dessus, on peut donc dresser la liste exhaustive de toutes les marques sous-jacentes et, pour chacune d'entre elles, la liste de tous les caractères dans lesquels elle apparaît. Cette étape représente un travail de très longue haleine et le résultat final sera toujours entaché d'erreurs sans que la méthode elle-même soit à mettre en cause. Dans la suite de cet article, tous les faits cités se basent sur la compilation réalisée par l'auteur et disponible sur demande.

L'on sait par ailleurs que la présence d'une marque à l'intérieur d'un caractère peut influencer ce dernier quant au sens (influence sémantique), au son (influence phonétique) et au signe (influence graphique) [Lire le chinois - l'apport de l'analyse, Ph Coueignoux 15/04/19, non publié]. Par définition même des marques comme signes constitutifs, la dernière influence est toujours active mais la difficulté vient du fait que rien a priori ne décèle la présence des deux premiers types d'influence car la présence d'une marque peut être purement graphique. Selon Bernhard Karlgren par exemple [Analytic Dictionary of Chinese and Sino-Japanese, Bernhard Karlgren, Dover

Publications, 1974 (from the 1923 original)], dans le caractère 封 (fēng) (caractère K30), la marque de l'ouest 圭 (guī) a capturé par influence graphique la marque originelle 丰 (fēng), dont l'influence phonétique est désormais perdue et l'on vérifiera a posteriori que 圭 ne joue dans 封 qu'un rôle graphique et n'est donc qu'un signe vide, non signifiant. Le cas du caractère 查 (chá) (caractère K1070) est légèrement différent. Dans ce dernier, la marque du sud 旦 (dàn) a capturé par influence graphique la marque originelle 且 (qiè), qui n'apparaît plus que dans une variante inusitée 查. Mais a posteriori on avancera que le rôle phonétique de 且 n'est pas vraiment perdu, il a simplement été transféré à 旦, capable de l'assumer via ses dérivés 坦 et 袒 (tǎn).

Le problème de décodage à résoudre est de déterminer la présence et la nature d'une influence phonétique pour chaque marque et chaque caractère qui en dérive.

On fera pour ce faire quatre hypothèses fondamentales sur la nature de l'influence phonétique:

- plasticité
- maximisation
- unicité
- intentionnalité

La première hypothèse ne fait que renommer le problème posé. Ce dernier n'existe en effet que dans la mesure où l'on est d'accord pour attribuer un rôle phonétique potentiel aux marques tout en reconnaissant l'absence de détermination prédictive. On admet ainsi que le son est plastique, c'est à dire qu'il peut se transmettre tout en se déformant de proche en proche. Et l'on refuse d'en dire plus sur ces transformations sinon d'affirmer qu'elles ne sauraient être totalement arbitraires. C'est justement au décodage de retrouver selon quelles règles deux sons peuvent être suffisamment proches l'un de l'autre pour que l'on puisse parler d'influence phonétique.

La deuxième hypothèse n'est pas nécessaire mais elle est naturelle. Il faut bien placer une limite entre exiger l'identité des sons avant de reconnaître une influence phonétique ou au contraire accepter comme valide n'importe quel saut d'un son à l'autre. Une fois cette limite codifiée sous forme de règles, on se donnera la liberté de la voir appliquée partout où un fait contraire ne s'y oppose pas. Rien ne permettra de prouver sans erreur possible les résultats ultimes qui découlent de cette hypothèse sinon un principe d'équité: une fois le codage positivement prouvé dans son principe, la charge de la preuve revient à celui qui refuse de l'appliquer dans un cas particulier. On peut voir là le premier effet de la démarche statistique qui sous-tend la méthode présentée dans cet article. Sachant que près de 90% des caractères chinois ont une composante phonétique [Edoardo Fazzioli "Caractères Chinois", publié en français chez Flammarion en 1987, p18], on a neuf fois plus

de chance d'avoir raison a priori en affirmant plutôt qu'en niant une influence phonétique là où elle est possible.

La troisième hypothèse tempère la précédente. A maximiser ainsi l'influence phonétique, le codage proposé risque d'en multiplier les possibilités. En adoptant le principe selon lequel une seule marque peut avoir une influence phonétique dans un caractère donné, on engendrera des revendications contradictoires de telle sorte que les marques se limiteront mutuellement leur propre champ d'influence. Par exemple on verra que le son du caractère 閔 (mǐn) peut être théoriquement influencé aussi bien par la marque ㄇ (mén) que par la marque 文 (wén) mais on écartera une de ces influences au nom de ce principe d'unicité. Le choix d'éliminer ㄇ plutôt que 文 sera guidé par d'autres considérations, expliquées le moment venu. De même le son 'péi' de 賠 sera attribué à l'influence de 音 (pòu) et non pas à celle de 貝 (bèi). Il faut noter que cette hypothèse d'unicité ne s'applique pas à l'influence sémantique. Dans un caractère donné, rien n'empêche à la marque phonétique de jouer aussi un rôle sémantique.

Enfin l'hypothèse d'intentionnalité est à la base de la méthode par déchiffrement. Elle consiste à dire qu'en l'absence d'influence règne le pur hasard. Ceci entraîne qu'inversement tout écart vis à vis de la moyenne correspond à une intention humaine, consciente ou non, et donne un indice sur la nature du code inconnu que l'on recherche. Par exemple l'une des méthodes de chiffrement les plus simples consiste à transposer les lettres de l'alphabet dans un message donné. Si un message source est en français et suffisamment long et si, d'autre part, la lettre Z est la plus fréquente dans le message codé, on en déduit que la lettre Z code la lettre E car cette dernière est de loin la plus fréquente en français. Dans le cadre de cette étude, comparons les deux marques 口 et 方 . De la première dérivent 246 caractères du corpus, couvrant 158 sons pinyin hors tonalité, le son le plus fréquent (he) correspondant à 8 caractères. De la deuxième dérivent 14 caractères, couvrant 3 sons pinyin, le son le plus fréquent (fang) correspondant à 11 caractères. Autrement dit le son 'fang' concentre 78.5% (11/14) du spectre phonétique potentiel de 方 contre 3.2% (8/158) du spectre de 口 pour le son 'he'. De plus la marque 口 arrive à 64.2% (158/246) du nombre maximal de sons qu'elle peut couvrir contre 21.4 (3/14) pour la marque 方 . Ainsi, tant en terme de couverture que de concentration, la marque 方 s'écarte de façon significative du comportement phonétique de la marque 口 , que l'on peut prendre en première approximation pour le résultat du pur hasard [Notes: en fait, d'une part, la marque 口 a une influence phonétique sur trois caractères dont le son est 'kou', une réalité masquée de prime abord; de l'autre, le pur hasard n'exclut pas, au contraire, que plusieurs caractères dérivés de la même marque aient le même son sans qu'il y ait d'influence. Mais, pour une marque donnée, le taux de concentration sur un son qu'elle n'influence pas y sera toujours plus proche de 1/n où n est le nombre de caractères dérivés et le taux de couverture plus proche de 100% en l'absence d'influence phonétique notable. Pour plus de rigueur, la modélisation du hasard tiendrait compte de la fréquence relative des sons sur

l'ensemble du corpus et, à cause de sa complexité, ne serait pas à sa place dans cet exposé dont elle ne modifierait pas les conclusions]. Attribuer en conséquence à la marque 方 une influence phonétique sur le son 'fang' est un premier pas vers le déchiffrement.

La démarche précédente est probabiliste. Elle ne démontre pas une influence phonétique de 方 sur le son 'fang', elle la déclare extrêmement probable. Par nature, selon la théorie de la reconnaissance des formes, elle est sujette à deux types d'erreur. Le premier est d'affirmer une influence alors que c'est contraire à la réalité. Le deuxième est de nier une influence qui existe bien. La théorie démontre que plus une démarche cherche à éliminer les erreurs d'un certain type, plus elle entraîne des erreurs de l'autre type. Selon l'hypothèse de maximisation invoquée ci-dessus, la présente démarche cherchera à éliminer le plus possible les cas où une influence phonétique réelle serait ignorée, au prix d'accepter à tort certaines influences.

-3- Difficultés et aide présentées par le codage à retrouver

La difficulté dans le décodage d'un message donné provient pour majeure partie du type de codage utilisé. Bien que personne n'en soit responsable, le codage de l'influence phonétique par les marques présente deux difficultés. On verra qu'une autre perspective, plus positive, offre deux types d'aides avec lesquelles bâtir une méthode de décodage.

liberté du codeur pour représenter les sons du chinois

Même si ce n'était pas le cas lorsque l'écriture chinoise se développait initialement [L'emploi des phonétiques dans l'écriture chinoise, Laurent Sagart 2006], aujourd'hui celui qui crée un nouveau caractère a normalement de nombreuses possibilités pour représenter un son chinois. Ce qui fait qu'à un son cible peut correspondre plusieurs marques susceptibles de le coder dans un caractère donné. C'est ce qui se passe lorsque l'on veut écrire en chinois un mot étranger, soit pour étendre le vocabulaire spécifique à un métier, soit pour transcrire un nom propre [Parlons Chinois, Zhitang Yang-Drocourt, 2007 L'Harmattan, pp 355-356].

Le processus de décodage en est d'autant plus compliqué. Si l'on reprend par exemple le chiffrement par transposition des lettres de l'alphabet, mais en utilisant deux lettres pour coder chaque lettre du message original, on voit que l'on dispose de 26 fois plus de signes qu'il n'est strictement nécessaire. On peut donc facilement coder la lettre E en puisant aléatoirement dans un stock de paires de lettres. Ce faisant on élimine le pic de fréquence relatif qui trahissait inmanquablement la présence du E dans le message codé.

A cette liberté dans le codage du son se rajoute le fait que le codeur n'est même pas dans l'obligation de coder

par une marque le son d'un caractère donné, ce qui est le cas de 染 dont aucune des trois marques constitutives n'a d'influence phonétique discernable.

nature ambiguë de cette représentation

La réciproque est également vraie, au sens où la même marque est susceptible de coder plusieurs sons différents. Prenons par exemple la marque 包 . En dérivent 19 caractères, 11 de son 'bao' et 8 de son 'pao'. Pour les mêmes raisons qui lient la marque 包 au son 'fang' par écart au hasard pur, la marque 包 est liée aussi bien au son 'pao' qu'au son 'bao'.

En fait c'est un phénomène de même type qui explique pourquoi la présente étude ne tient aucun compte du ton des caractères car la corrélation entre marque et tonalité est trop faible pour être exploitée. Dire que 包 est liée au son 'bao' est un raccourci pour désigner les sons 'bāo', 'báo', 'bǎo' et 'bào'. Certes, la répartition des 11 caractères dérivés de son 'bao' entre les quatre tons (6+1+1+3) n'est pas uniforme mais toute tentative pour lier 包 à 'bāo' seul se heurterait à une répartition (6,5), de corrélation trop indéterminée pour être utile.

A cette ambiguïté dans la nature du rôle phonétique d'une marque, on rattachera le fait qu'une marque présente dans un caractère donné peut n'y jouer aucun rôle phonétique même si cette marque est phonétiquement active dans bien d'autres caractères. C'est le cas de la marque 圭 dans le caractère 封 cité plus haut.

lien de proximité entre sons

L'exemple de la marque 包 n'apporte pas que des difficultés. En l'occurrence elle offre aussi une aide cruciale dans le processus de déchiffrement. Car si l'influence phonétique d'une marque n'est pas univoque, limitée à un seul son, les différents sons qui peuvent y être associés ne sauraient l'être par pur hasard mais parce que ces différents sons sont proches les uns des autres. Cette affirmation renvoie évidemment aux processus historiques d'évolution phonétique qui expliquent la situation présente et dans lequel le hasard joue un rôle minimum. Mais elle se dégage aussi de l'analyse brute du code en question. Si donc 包 est liée aux deux sons 'bao' et 'pao' et à eux seuls, ce ne peut être l'effet du pur hasard mais parce que, en toute probabilité, ces deux sons sont proches l'un de l'autre pour une raison que l'analyse du code révèle même si elle ne peut pas l'expliquer.

décomposition des sons

Avant d'aborder le processus de décodage proprement dit, le cas précédent doit être rapproché du cas d'autres marques dont l'influence phonétique s'étend aussi sur deux sons. Après la paire 'bao/pao' liée à 包, voici les paires 'ba/pa', 'ban/pan', 'bang/pang' liées respectivement aux marques 巴, 半 et 旁. Il paraît alors logique de postuler l'indépendance de la consonne initiale du son du reste de la syllabe qu'il représente et de raffiner la conclusion précédente en déclarant que, dans les quatre paires de sons proches citées, ce sont en fait les

consonnes 'b' et 'p' qui sont proches l'une de l'autre.

Ceci permet alors d'examiner la marque 卑, dont les caractères dérivés couvrent les sons 'bai' (1), 'bei' (2), 'bi' (3), 'pai' (1) et 'pi' (2), sons dont la dispersion apparente échappait jusqu'alors à tout décodage. En projetant les sons concernés sur la seule consonne initiale, on retrouve effectivement la paire 'b/p' avec une distribution en deux pics (6+3) inexplicable par le seul fait du hasard. Mais si, dans ces conditions, l'on admet l'influence phonétique de 卑 sur les cinq sons de ses dérivés, cela entraîne que l'on considère aussi comme proches les unes des autres les trois terminaisons 'ai', 'ei' et 'i'.

De même, l'examen de la marque 并, dont les dérivés couvrent les sons 'beng' (1), 'bing'(3), 'pin' (2) et 'ping' (2) et présentent ainsi les deux pics (4+4) de la paire 'b/p', conduit encore à séparer, dans la terminaison, la voyelle 'e' ou 'i' de la consonne finale 'n' ou 'ng', et, selon ces deux nouvelles projections, à déclarer 'e' proche de 'i' et 'n' proche de 'ng'.

En conclusion on est fondé à analyser la proximité phonétique en trois dimensions indépendantes, la consonne initiale, une partie médiane, qui peut être voyelle ou diphtongue, et la partie finale du code pinyin des sons, deux parties au plus pouvant ne pas exister dans un son donné.

-4- Traitement du problème par déchiffrement

points d'ancrage

La méthode de déchiffrement adoptée se base donc sur une série de détails dont l'existence est contraire aux lois du pur hasard et qui révèlent par conséquent un élément significatif du codage inconnu. Elle suit ainsi un procédé usuel, comme celui qui attaque un chiffrement par transposition des lettres de l'alphabet. Il détecte d'abord la lettre chargée de coder le E, trahie par son pic de fréquence, Z dans l'exemple cité plus haut, et cherche ensuite les paires de lettres correspondant aux mots très courants 'et', 'ne', 'se' pour faire des hypothèses sur le code attribué à T, N et S.

Dans le cas présent, ces points d'ancrage sont constitués par les marques dont le spectre phonétique formé par le son des caractères dérivés présente un ou plusieurs pics de fréquence relative significatifs. La collecte de ses points d'ancrage est faite systématiquement sur la table exhaustive des marques et de leurs dérivées constituée au préalable. Elle est rendue plus productive par la décision de projeter les sons sur les trois dimensions 'consonne initiale', 'partie médiane' et 'consonne finale', mesure justifiée dans la section qui précède.

L'analyse des points d'ancrage apporte donc deux types de renseignements sur le code inconnu, illustrés par les exemples déjà traités:

- la projection en trois dimensions du spectre phonétique d'une marque donnée
 - 'f', 'a' et 'ng' pour 方
 - 'b' et 'p', 'ao' et '-' (dénotant l'absence de consonne finale) pour 包
- la proximité phonétique d'éléments de sons, dans chaque dimension, par exemple:
 - 'b' et 'p' selon les marques 包, 巴, 半, 旁, 卑 et 并
 - 'ai', 'ei' et 'i' selon la marque 卑
 - 'e' et 'i' selon la marque 并
 - 'n' et 'ng' selon la même marque 并

Cette présentation n'a pas pour but de donner la liste exhaustive des points d'ancrage ainsi trouvés. On notera seulement deux points de synthèse importants:

- le grand nombre de marques, en général elles-mêmes dérivées de marques plus élémentaires, dont le spectre phonétique se limite à un seul son, comme 哥 et 鬲 (ge), 勾 (gou), 郎 (lang), 星 (xing)..., soit environ 40% du total [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation]. L'analyse phonétique de ces marques est évidemment trivial.

- la grande diversité des proximités phonétiques avérées par les marques dont le spectre phonétique recèle des pics multiples. La maîtrise de cette diversité est l'essentiel du décodage.

fond de carte phonétique et cartes phonétiques

Pour mieux suivre les considérations abstraites qui suivent sur la méthode d'analyse que présente cet article, le lecteur est prié de se reporter à l'exemple donné ci-dessous pour l'illustrer (section ***illustration de l'approche itérative***). Il n'oubliera pas que cette méthode n'est pas une démonstration mathématique mais décrit comment atteindre un résultat par approximations successives, le point de départ étant une liste nécessairement partielle de points d'ancrage et le cheminement intermédiaire pouvant varier largement selon le point de départ et celui qui met la méthode en pratique. Seul compte le résultat final de ces cheminements car l'on assure qu'il est unique et optimal.

L'analyse des points d'ancrage initiaux pose d'emblée un problème de représentation. Si l'on se concentre sur les 23 consonnes initiales, l'on peut adopter en théorie une matrice de rapprochement 23x23, symétrique, dont les lignes et les colonnes donnent la liste des consonnes et chaque case dénote par un 1, la présence et par un 0, l'absence d'une proximité entre les deux consonnes correspondant à la ligne et la colonne concernées.

Voici, pour l'exemple en question, une matrice 23x23 qui représente les rapprochements constatés à un certain stade de l'illustration de la méthode:

	b	c	ch	d	f	g	h	j	k	l	m	n	p	q	r	s	sh	t	w	x	y	z	zh
b	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
c	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
ch	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
d	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1
f	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
h	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
j	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0
k	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
l	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
n	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
p	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
r	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
s	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
sh	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0
t	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
w	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0
x	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0
y	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0
z	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
zh	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

On constate que la matrice 23x23 est 'creuse', c'est à dire comporte surtout des 0. Par contre on peut y déceler un phénomène inverse. La marque 分 présente un pic significatif sur 'f' (9 dérivés sur 17) et un pic secondaire constitué par la paire 'b/p' (6). On a donc pu rapprocher 'f' de 'b' et de 'p' et, si l'on place les trois consonnes 'b', 'p' et 'f' sur des lignes et colonnes consécutives de la matrice de rapprochement, on voit apparaître une sous matrice carrée 'pleine' de 1, sans aucun 0.

	b	f	p
b	1	1	1
f	1	1	1
p	1	1	1

Cette proximité mutuelle permet ensuite de repérer une influence phonétique de 方 ('fang') sur ses dérivés 旁 et 彷 ('pang'), même en l'absence de pic incontestable sur 'pang' parmi tous ses dérivés (2 dérivés sur 14). C'est en effet une conséquence des principes de maximisation et d'unicité énoncés plus haut. Car, si l'on admet comme l'hypothèse la plus probable que 旁 et 彷 sont bien sous l'influence phonétique d'une marque (maximisation) et d'une seule (unicité), 方 est la seule marque possible.

On vérifie encore que la marque 贛, dont le spectre phonétique n'était pas assez caractérisé a priori (2 'pen', 1 'ben' et 1 'fen'), présente a posteriori une grande unité d'influence phonétique.

Le même phénomène surgit à l'examen du spectre phonétique des marques 共 (pics sur 'g' et 'h'), 亢 (pics sur 'k' et 'h') et 工 (pics sur 'g', 'h' et 'k'), déterminant une autre 'famille' de consonnes 'g', 'k' et 'h' caractérisée par la sous matrice pleine ci-dessous:

	g	k	h
g	1	1	1
k	1	1	1
h	1	1	1

On adopte alors une autre forme de représentation de la proximité, plus compacte et plus parlante, le fond de carte phonétique. Cette dernière est encore une matrice à deux dimensions mais le rôle respectif des rangées et des cellules est inversé. Cette fois, les consonnes sont réparties entre les cellules et c'est l'appartenance de deux consonnes à la même ligne ou à la même colonne qui dénote la proximité. Les deux familles (b, p, f) et (g, k, h) y figurent par exemple chacune sur une colonne. En acceptant de répéter la même consonne dans sa colonne et dans sa ligne et en continuant d'admettre la présence de nombreuses cellules vides, on construit ainsi un fond de carte phonétique putatif représentant visuellement lui aussi les proximités connues mais à l'aide d'une matrice beaucoup plus compacte.

Il est important de noter qu'à ce stade le concept de fond de carte phonétique n'est qu'un outil de travail qui ne se

justifie que par sa capacité à représenter synthétiquement des faits avérés ou simplement présumés. Voici par exemple, tirée de l'illustration de la méthode, le fond de carte phonétique, plus compact et plus parlant, qui remplace la matrice de rapprochement ci-dessus. Comme cette dernière, ce fond de carte n'a jamais été qu'une étape provisoire. Pour les faits sur lesquels il s'appuie, on se reportera aux détails donnés ci-dessous à titre historique dans le paragraphe "illustration de la méthode itérative".

					r	y		
						y	w	
	h						w	
							m	n
z	g	b	d	zh		j		l
	k	p	t					
		f						
c				ch		q		
s	s		sh		sh	x		

figure 1

C'est pour respecter la compacité que l'on a limité le fond de carte à deux dimensions, comme la matrice de rapprochement; car on aurait pu choisir d'aligner les consonnes initiales proches les unes des autres sur trois dimensions, voire plus. Ce choix limite donc à deux, par ligne ou colonne, le nombre de façons selon lesquelles deux consonnes peuvent être déclarées proches l'une de l'autre. Il faut noter que ce choix n'est pas neutre car il recoupe la nature intrinsèque des consonnes initiales établie par la phonologie [Standard Chinese (Beijing), Lee Wai-Sum et Zee Eric, *Journal of the International Phonetic Association*. 33 (1): 109–112, 2003].

On étend facilement le même concept de fond de carte pour représenter la proximité dans la partie médiane des sons. Pour des raisons de place et de priorité, cet article n'approfondira le décodage de la médiane que de manière secondaire. Enfin la finale fait simplement l'objet d'une matrice colonne à trois cases: '-', 'n' et 'ng'.

Pour une marque donnée et à un stade donné d'application de la méthode, on peut alors projeter les sons reconnus comme faisant partie de son spectre phonétique sur les fonds de carte utilisés en 'coloriant' les cellules correspondantes. On obtient ainsi ce que l'on appelle la carte phonétique de la marque. Voici par exemple pour 分 la partie de sa carte concernant les initiales:

					r	y		
						y	w	
	h						w	
							m	n
z	g	b	d	zh		j		l
	k	p	t					
		f						
c				ch		q		
s	s		sh		sh	x		

carte des initiales de la marque 分 sur le fond de carte de la figure 1

mesures de dispersion et de densité

On peut encore associer un aspect quantitatif au fond de carte en lui associant une mesure de dispersion de l'influence phonétique d'une marque donnée. Cette mesure se définit comme la somme de trois termes, correspondant respectivement aux matrices des initiales, médianes et finales:

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

Pour chaque matrice, son terme est calculé en couvrant d'abord toutes les cellules coloriées par une série de rangées, lignes ou colonnes, reliées de façon à ce que l'on puisse connecter deux cellules coloriées quelconques par un parcours le long des rangées ainsi choisies. On calcule ensuite la dispersion sur chaque rangée et l'on fait le total sur l'ensemble des rangées. Les règles de calcul sont définies de sorte qu'une dispersion aléatoire des consonnes sur la matrice sous l'effet du hasard corresponde à une mesure maximum.

Par opposition le minimum absolu, 0, correspond à une seule cellule coloriée, soit l'absence totale de dispersion. Ensuite, le long d'une même rangée, la dispersion augmente avec le nombre de cellules coloriées, mais de façon dégressive (par exemple 0 pour 'b', 1 pour 'b' et 'f', 1,5 pour 'b', 'p' et 'f', etc...). En effet l'accumulation le long d'une rangée est contraire au hasard. Enfin on augmente encore la dispersion en appliquant une pénalité lorsque l'on 'pivote' ('fait un coude') pour relier une ligne et une colonne comprenant des cellules coloriées mais sans que la cellule commune à la ligne et à la colonne soit elle-même coloriée. En effet il est aussi contraire au hasard qu'une consonne du spectre phonétique fasse précisément le lien entre deux autres consonnes du spectre placées sur les deux rangées auxquelles elle appartient. Autrement dit, plus un spectre est fait de cellules connectées les unes aux autres en passant par des pivots eux-mêmes partie du spectre, plus ce dernier est cohérent et moins grande est sa mesure de dispersion.

De nombreuses configurations de rangées sont possibles qui, toutes, couvrent l'ensemble des cellules du spectre par une série connectées par des pivots et chacune donne lieu à un calcul de mesure. La mesure retenue est le minimum de tous les cas possibles, obtenue en prenant une configuration de rangées qui recouvre bien toutes les cellules du spectre mais évite les pénalités autant que possible. Par exemple pour calculer la dispersion de la marque 允 dont le spectre couvre les sons 'zhen', 'chen' et 'dan', la configuration de gauche, comportant trois rangées et deux pivots, dont un pénalisé, est écartée au profit de celle de droite, plus économe et sans pénalité:

	h							h							
z	g	b	d	zh			j	z	g	b	d	zh			j
	k	p	t						k	p	t				
		f								f					
c				ch			q	c				ch			q
s	s		sh		sh	x		s	s		sh		sh	x	

tentatives de calcul de la mesure de dispersion de la marque 允 sur un extrait de sa carte phonétique

Pour plus de détails sur cette mesure de dispersion, on renvoie à [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation]. On signalera simplement ici qu'aux lignes et colonnes de la matrice des médianes, s'ajoute pour établir une connexion une rangée privilégiée formée des consonnes simples 'a', 'e', 'i', 'o', 'u' et 'ü'.

哥 et 鬲 (ge), 勾 (gou), 郎 (lang), 星 (xing) reçoivent ainsi une dispersion de 0. Par contre les marques 包 et 方 recevront une dispersion de 1, parce que leur 'tache' phonétique s'étend sur deux consonnes de la colonne contenant la famille (b, p, f) tout en restant focalisée sur les deux autres dimensions, respectivement les initiales 'b' et 'p', la médiane unique 'ao' et la finale unique '-', les initiales 'p' et 'f', la médiane unique 'a' et la finale unique 'ng'. 贲 aura encore une dispersion de 1.5 (triplet (b, p, f) sur la même colonne), 卑 une dispersion de 2.5 (1 pour la paire 'b/p' plus 1.5 pour le triplet de médianes (ai, ei, i) se terminant en 'i') et 并 une dispersion de 3 (1 chacun pour les paires 'b/p', 'e/i' et 'n/ng').

Il faut insister sur le fait que cette mesure n'est qu'un outil de travail qui ne se justifie que par sa capacité à refléter de façon synthétique les faits bruts, à savoir la liste des cellules coloriées de la carte phonétique et l'écart par rapport au hasard. D'autres définitions sont possibles à côté de ce qu'il convient d'appeler 'la mesure de Coueignoux' en faisant référence à [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation].

A cette mesure de dispersion d'une marque, on associera une mesure de densité en divisant par la dispersion le nombre de dérivés phonétiques trouvés dans le corpus choisi:

$$\text{Dens} = (\text{Nb de caractères}) / \text{Disp}$$

Mettant à part les marques de dispersion 0, de densité infinie, une marque est donc d'autant plus dense que ses pics phonétiques sont relativement élevés, c'est à dire que son spectre ne peut être le fruit du hasard. 包 a ainsi en définitive une densité de 19, 方 de 13, 并 de 3, 卑 de 2.5.

approche itérative

A ce stade on dispose de tous les outils nécessaires pour poursuivre le décodage et l'on en a déjà esquissé la méthode dans ce qui précède.

L'on est en effet passé de la marque 包, dont la densité remarquable reflète un écart au hasard irréfutable, à celui de la proximité intrinsèque entre les consonnes 'b' et 'p'. Puis de cette proximité cartographique représentée clairement sur un fond de carte, l'on est revenu à l'examen de marques comme 卑 et 并 dont la dispersion et la densité étaient peu significatives a priori. On a alors repris l'analyse cartographique car 卑 et 并 viennent d'une part renforcer la proximité mutuelle des consonnes 'b' et 'p' et de l'autre suggérer d'autres proximités cartographiques, en l'occurrence sur les deux autres dimensions des médianes et des finales.

Un autre aller-retour entre marques et cartographie a alors conduit, à partir de la marque 分, à adjoindre 'f' à la paire 'b/p' et ensuite à compléter par ricochet l'analyse des marques 方 et 贛.

C'est ce processus itératif entre examen des marques et construction du fond de carte phonétique qu'il convient maintenant de généraliser en deux étapes à parcourir en boucle:

- partant des points d'ancrage suggérés par l'examen du spectre phonétique des marques
améliorer le fond de carte phonétique et ses mesures associées
- partant d'un fond de carte phonétique donné, avec les mesures associées,
examiner les marques pour en révéler plus exhaustivement le spectre phonétique.

Bien que les parcours déroulés selon ce processus varient selon les données phonétiques que l'on choisit de prendre en compte au fur et à mesure, tous convergent vers un fond de carte unique (si l'on exclut les transformations triviales par transposition ou changement d'ordre des lignes et colonnes) et optimal sur lequel on peut représenter, en en minimisant la dispersion et maximisant la densité, le spectre maximal de chacune des marques.

Voici présentées par exemple sur le fond de carte optimal auquel aboutit la méthode, les cartes phonétiques des

marques 并 et 分 citées plus haut. Sur chacune des trois sous-cartes, couvrant initiales, médianes et finales, le coloriage des cellules utilise une teinte grisée et les rangées utilisées pour calculer la dispersion sont indiquées en marge par une case noire. La densité renvoie à la liste de caractères dérivés figurant dans le dictionnaire de marques compilé par l'auteur et n'est donnée ici qu'à titre informatif.

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

并

dispersion 3
densité 2,67

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

分

dispersion 5
densité 3

Lorsque la marque ne correspond qu'à un seul son ou quelques sons dont la consonne initiale appartient à la même colonne du fond de carte et les parties médianes et finales sont très proches, on se contente de donner la liste de ces sons. Par exemple les cartes phonétiques des marques 包 et 方 citées plus haut ne comporteraient que quatre cellules grisées, respectivement ('b', 'p') et ('p', 'f') en initiale, 'ao' et 'a' en médiane, '-' et 'ng' en finale. On les remplacera par les listes condensées suivantes:

b	ao
p	

carte phonétique de

包

dispersion 1
densité 19

p	a	ng
f		

carte phonétique de

方

dispersion 1
densité 13

Le progrès opéré au fur et à mesure des itérations sur la boucle d'analyse se mesure de la façon suivante:

- à la suite de la première étape, plus un fond de carte phonétique et ses mesures associées diminue la dispersion et augmente la densité du spectre phonétique des marques analysées jusque là, meilleur il est car il accroît la cohérence de l'analyse

- à la suite de la deuxième étape, plus un fond de carte phonétique et ses mesures associées justifie d'étendre l'influence phonétique des marques sur des caractères dérivés ignorés jusque là, meilleur il est, selon notre hypothèse de maximisation.

Ainsi un fond de carte qui rapproche le 'p' du 'b' en les plaçant sur une même colonne est meilleur qu'un autre qui les mettrait sur deux rangées séparées. Il minimise en effet la dispersion phonétique a priori de 包, 巴, 半, 旁, 卑 et 并 en éliminant un pivot dans la représentation de leur spectre et par conséquent le coût d'une des deux rangées qui lui sont associées.

De même un fond de carte qui rapproche 'f' de 'b' et 'p' en le rajoutant à leur colonne est meilleur car il permet d'étendre a posteriori le nombre des dérivés phonétiques de 方 et de 贛. Si l'on admet ce rapprochement, on rattachera en effet au spectre de 方, limité a priori à ('fang'), les deux caractères 旁 et 彷彿 ('pang') et, au spectre de 贛, circonscrit a priori à ('ben') et ('ben'), le caractère 憤 ('fen').

Cette méthode itérative étend donc progressivement l'analyse en enrichissant le contexte qui relie les faits connus entre eux puis en révélant des faits nouveaux que ce nouveau contexte rend significatifs et ainsi de suite. Encore une fois on se gardera de penser que le cheminement ainsi trouvé est unique et la carte 9x9 de la figure 1 présentée en exemple n'a qu'une valeur anecdotique. Mais l'on avance que tout processus déroulé selon cette approche convergera vers le fond de carte optimal unique utilisé pour présenter les cartes phonétiques des marques 并 et 分 et présenté ci-dessous dans les figures 3bis, 4 et 5. Par contre l'extension finale du spectre phonétique d'une marque donnée restera parfois sujet à discussion car, pour une marque donnée et en l'absence de pic, les dernières influences phonétiques rattachées à son spectre ne sont que des possibilités dont la seule justification statistique est fondée sur l'hypothèse de maximisation, une démarche qui finit par devenir subjective. Autant l'extension de l'influence phonétique de 分 (fēn) sur 貧 (pín), dont le son 'pin' n'a pas d'autre instance dans les dérivés de 分 ne semble contestée par personne, autant celle de 里 (lì) sur 埋 (mái, mán) semble sujet de controverses.

On notera incidemment qu'une marque n'a pas besoin d'être ou d'avoir été un caractère en elle-même pour se voir reconnaître une influence phonétique. Ainsi la carte phonétique de la marque 毘, dérivée sémantique de 匚 et phonétique de 比 couvre clairement les sons 'bi' avec son dérivé 篋 et 'pi' avec son dérivé 媿, alors que le caractère 毘 n'est pas attesté.

illustration de l'approche itérative

Ce qui suit est simplement destiné à renforcer le caractère intuitif de la méthode itérative qui vient d'être exposée. Dans la mesure où ce cheminement n'est pas unique, la découverte des faits qui y est rapportée n'a aucune valeur normative sauf a posteriori. Si le lecteur s'intéresse plus à la teneur et à la justification des résultats qu'à la méthode qui y a conduit, il peut donc, au moins en première lecture, sauter cette section sauf à y repérer les figures 3bis, 5 et 5 déjà citées qui en synthétisent les résultats. En lieu et place de la justification historique qu'elle apporte pas à pas, il se reportera aux justifications quantitatives développées a posteriori dans [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation]. S'il désire au contraire étudier cette section, il s'en facilitera la lecture en se référant au dictionnaire des marques établi par l'auteur, qui donne pour chaque marque tous les caractères du corpus qui en dérivent.

Le fond de carte phonétique 9x9 ci-dessous que l'on a déjà présenté a été effectivement utilisé à un certain stade de l'application de la méthode par l'auteur pour représenter la proximité des consonnes initiales.

					r	y		
						y	w	
	h						w	
							m	n
z	g	b	d	zh		j		l
	k	p	t					
		f						
c				ch		q		
s	s		sh		sh	x		

figure 1

Au delà des familles déjà répertoriées (b, p, f) et (g, k, h), il représente les proximités suivantes, selon les colonnes:

- entre 'z' et 'c' selon la marque 曾, et renforçant l'unité phonétique de la marque 早
- entre 'z' et 's' selon la marque 臬
- entre 'c' et 's' selon la marque 司
- entre 'g' et 's' selon la marque 公 (a)

- entre 'd' et 't' selon la marque 炎
- entre 'sh' et la paire 'd/t' selon la marque 是 (b)
- entre 'zh' et 'ch' selon la marque 知, et renforçant l'unité phonétique de la marque 专
- entre 'r' et 'sh' selon la marque 上 (c)
- entre 'y' et 'x' selon la marque ㄩ (应 et 兴) (d)
- entre 'y' et 'j' selon la marque ㄨ (舆 et 具) (d)
- entre 'j' et 'x' selon la marque 夹
- entre 'q' et 'x' selon la marque 肖 (e)
- entre 'w' et 'm' selon les marques 亡, 门 et 未
- entre 'n' et 'l' selon la marque 良

et selon les lignes:

- entre 'r' et 'y' selon la marque 右 (c)
- entre 'y' et 'w' selon la marque 元
- entre 'h' et 'w' selon la marque 勿
- entre 'm' et 'n' selon la marque 尔 (f)
- entre 'z' et 'zh' selon la marque 乍
- entre 'z' et 'j' selon la marque 且
- entre 'g' et 'j' selon la marque 句
- entre 'zh' et 'd' selon la marque 者
- entre 'j' et 'l' selon les marques 林 et 金
- entre 'sh' et 'x' selon la marque 肖 (d), et renforçant l'unité phonétique des marques 小 et 向

notes

(a) cette marque n'a qu'un seul pic significatif avec les trois caractères 松, 讼, et 颂, le premier confirmé par le choix qu'on en a fait pour simplifier le caractère 鬆, mais sur le son 'song' distinct du son originel 'gong' de la marque 公. Ceci est équivalent à la présence de deux pics significatifs.

En fait le voisinage phonétique entre 's' et 'g' n'est pas direct comme le laisse entendre la figure 1, mais indirect comme le finalisera la figure 3 de la matrice des initiales. Et l'on retiendra que pour des raisons pratiques la figure 3 est remplacée dans le fond de carte optimal par la figure 3bis, le cas de 公 faisant exception.

(b) cette marque a trois pics, respectivement sur 'shi' (是 et 匙), 'di' (堤 et 提) et 'ti' (提 et 题). Si l'on trouve trop

faible la portée statistique de ces rapprochements, on en trouvera une confirmation plus robuste dans [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation].

(c) ces deux rapprochements sont très peu significatifs et sont pris à ce stade de façon hypothétique, faute de mieux, pour 'caser' la consonne 'r' quelque part en attendant mieux!

(d) ces deux rapprochements sont fondés sur l'analyse de caractères simplifiés en terme de marques qui n'ont jamais eu d'existence indépendante.

(e) la marque 肖 présente trois pics significatifs sur 'x', 'sh' et 'q' (8+5+4 sur un total de 19).

Elle n'entraîne pas à elle seule la création d'une famille (q, x, sh) sur le modèle des familles (b, p, f) et (g, k, h), mais plutôt de deux paires 'x/sh' et 'x/q' entre lesquelles 'x' joue un rôle pivot. Ce choix illustre le caractère quelque peu arbitraire du cheminement de la méthode. En l'occurrence il sera préservé dans le fond de carte optimal mais un choix différent aurait fini par être rectifié comme le seront les choix moins heureux qui ont présidé à l'emplacement des cellules pour le 'l', le 'm' et le 'n' dans la figure 1.

Il n'est par contre pas possible à ce stade de lever toutes les ambiguïtés créées par les cinq pics significatifs de la marque 占 sur 'zh', 't', 'd', 'n' et 'z' (8+4+3+2+2). Il faudra attendre plus de faits indépendants pour clarifier la carte de cette marque.

(f) porté par les caractères 弥 (mí) et 你 (nǐ), dérivés de 尔, ce rapprochement est à ce stade nullement significatif et reflète plutôt une hypothèse souhaitée par l'auteur!

La carte de travail de la figure 1 représente donc des proximités avérées par des points d'ancrage indiscutables, comme les familles (b, p, f) et (g, k, h) et de nombreuses autres proximités nouvelles mais aussi des faits présumés dont la justification est légère, voire inexistante. Rien ne suggère à ce stade un lien entre 'c' et 'ch', 's' et 'sh' à part le code pinyin lui-même. A nouveau rien ne justifie de mettre 'q' sur la même ligne que 'c' et 'ch', ni 'b' sur la même ligne que 'z', 'g', 'd', 'zh', 'j' et 'l'. Quant à la ligne 'k', 'p', 't', elle n'est au stade considéré qu'un a priori sans preuve transposé par l'auteur de la grammaire du grec ancien [Précis de grammaire grecque, Ch. Maquet et F. Flutre, Hachette, 1925], c'est à dire une influence déguisée de la phonologie !

En revanche certains rapprochements restent pendants, faute de rentrer facilement sur la matrice:

- entre 'r' et 'n' selon la marque 若
- entre 'q' et 'sh' selon la marque 全
- entre 'h' et 'm' selon la marque 每

Ce fond de carte de travail n'a donc pas le niveau optimal visé et a effectivement évolué par la suite en un nouveau fond de carte présenté en figure 2 comme une matrice 7x6 et résultant de trois bouleversements

majeurs:

- le basculement des consonnes 'm', 'n', 'r' et 'l' pour former une ligne homogène qui rende compte à la fois de la proximité du 'n' avec la paire 'd/t' et du 'm' avec la famille (b, p, f), sans perdre pour autant les liens avérés entre les paires 'm/n', 'l/n' et 'l/j'.

- une compaction des lignes qui mette les consonnes 'c', 'ch' et 'q' sur la ligne (k, p, t)

- une compaction des colonnes qui mette les consonnes 'sh' et 'r' avec la paire 'zh/ch'.

		w			y
		m	n	r	l
	h				
z	g	b	d	zh	j
c	k	p	t	ch	q
s	s		sh	sh	x
		f			

figure 2

Avec un pic principal sur la paire 'd/t' et un pic secondaire sur 'n', la marque 丁 incite à inclure n dans la famille (d, t, sh) inspirée par la marque 是. Cette nouvelle famille (d, t, sh, n) étend par ricochet le spectre phonétique des marques 鸟, 水 et 女 et permet d'organiser le spectre de la marque 占, dont l'ambiguïté défiait jusqu'alors l'analyse, par croisement des deux familles (z, zh, d) et (d, t, n) au niveau du pivot 'd'. En parallèle avec le développement précédent, l'adjonction de 'm' à la famille (b, p, f) est justifié par la marque 必, de pic principal 'bi', au vu de son pic secondaire sur mi (秘, 泌), et de la marque 林, dont le spectre phonétique gagne en cohérence si l'initiale 'm' (麻, 梦) rejoint la famille (b, f) (彬, 梵, 焚).

Confirmant les liens entre les pairs 'm/n' et 'l/n', la nouvelle famille (m, n, l) donne alors par ricochet son extension maximale aux spectres des marques 米, 里 et 牛. Si dans cette nouvelle famille on inclut 'r', on tient maintenant compte des faits de proximités entre 'r' et 'n' portés par la marque 若, mentionnés plus haut comme laissés en suspend, ainsi que par la marque 乃. Cela suggère aussi de réinterpréter le son pinyin exceptionnel 'er' comme équivalent à la consonne initiale 'r' prise isolément, sans partie médiane ni finale. Cette possibilité illumine alors le cas significatif mais jusqu'à lors obscur des marques 而, 尔, et 耳, qui combinent ce son exceptionnel 'er' avec des sons commençant par 'n', 'm' et 'l' ainsi que par la consonne 'r' ordinaire. L'identité 'er'='r' ne fait que traduire une régularisation statistique du pinyin en tant que code sans préjuger de la réalité

phonétique sous-jacente qu'il cherche à transcrire. On peut toutefois chercher à rattacher ce phénomène au fait qu'en français les consonnes de la série 'l', 'm', 'n', 'r' se distinguent de la plupart des autres, comme 'b', 'c', 'd', 'j', 'k', 'q' ... en ce que la prononciation de leur nom ne fait pas appel à une voyelle finale.

On notera que la création d'une famille (m, r, n, l) oblige à verser le 'l' dans la colonne (j, q, x) pour conserver la proximité de la paire 'j/l'. Cette modification se trouve justifiée car elle minimise alors la dispersion phonétique de la marque 金, révélée comme le représentant le plus complet d'une famille (j, q, x, l, y). Pour conserver de même la proximité de la paire 'm/w', la famille (b, p, f) s'agrandit en une famille (b, p, f, m, w), qui minimise en effet la dispersion phonétique de la marque 敏 (繁) et de la marque 亦 (变, 蛮, 弯), le 'm' faisant office de pivot entre cette dernière famille et la famille (m, r, n, l).

La création d'une ligne (c, k, p, t, ch, q) provient de la prise en compte de la marque 欠 dont le spectre phonétique s'étend sur les consonnes 'k', 'ch' et 'q'. Elle étend par ricochet l'influence sur les consonnes 'k' et 'q' des marques 卡, 空, et 高, la consonne 'k' jouant le rôle pivot pour cette dernière entre la colonne (g, k, h) et la nouvelle ligne (c, k, p, t, ch, q). De même elle étend sur les consonnes 'c' et 't' le spectre de la marque 天 (蚕) et sur les consonnes 'ch' et 't' celui de la marque 堂 (瞳). Elle permet aussi d'organiser le spectre de la marque 召 sur les consonnes 'ch' (超) et 't' (迢) d'une part et 'zh' (召) et 'd' (貂) de l'autre. On remarque que, sur la ligne (c, k, p, t, ch, q), la consonne 'p' reste très indépendante de ses proches, en ce sens qu'un spectre connecté par cette rangée, comme celui de 欠, ne s'étend pas au 'p'. C'est aussi le cas de la consonne 'b' dans la famille parallèle (z, g, b, t, zh, j), le cas de la marque 更 faisant figure d'exception, car son spectre, connecté par cette rangée, inclut bien le 'b' (便) à côté du 'g' (更) et du 'j' (梗).

La création de la colonne (r, zh, ch, sh) se justifie parce qu'elle minimise la dispersion des marques 只, 旨 et 朱, et permet par ricochet d'étendre l'influence de nombreuses marques dont le spectre joue aussi sur la colonne (d, t, sh, n), comme 台, 啻, 尚.

On notera que l'analyse laissée en suspend de la marque 全 a été résolue non pas en rapprochant les consonnes 'q' et 'sh' mais par la prise en compte de la consonne 'ch' comme pivot. Par contre elle ne résout toujours pas le cas de la marque 每, toujours en suspend pour avoir suggéré un lien entre 'h' et 'm', et fait perdre les liens de la figure 1 entre 'h' et 'w' d'une part (元) et 'r' et 'y' de l'autre (右), qu'il serait souhaitable de rétablir.

A ce stade de l'analyse, il apparaît que la compaction du fond de carte phonétique semble aller dans le sens de l'optimisation globale qui est le but ultime. Il est donc naturel de chercher à poursuivre dans ce sens en compactant maintenant les consonnes initiales 'f' et 'h', isolées sur leur propre ligne dans la figure 2, mais adjacent d'une case vide sur une ligne voisine. Cela nous amène à la figure 3, soit une matrice 5x6 dont l'ordre relatif des colonnes et des lignes, qui ne joue aucun rôle dans les mesures de dispersion, a été aussi normalisé.

Les trois changements majeurs sont donc de placer

- la consonne 'h' dans la famille (m, r, n, l)
- mais aussi de dupliquer 'h' dans la famille (s, sh, x).
- la consonne 'f' dans la famille (h, s, sh, x)

Par convention on évite dans la figure 3 de dupliquer explicitement les consonnes mais en leur permettant, sur cette carte grâce à la normalisation de l'ordre des lignes et des colonnes, d'occuper par extension une cellule vide voisine, une possibilité rappelée par leur mise entre parenthèse explicite sur la cellule concernée. Cette convention est étendue aux consonnes 'w' et 'y', qui sont aux extrémités d'une ligne vide. Sans avoir à le mentionner explicitement pour ne pas alourdir la présentation, toute cellule de cette ligne peut aussi être occupée par 'w' ou 'y', rétablissant ainsi les liens entre 'h' et 'w' d'une part et 'r' et 'y' de l'autre.

b	g	z	zh	d	j
p	k	c	ch	t	q
f	(h)(s)	s	sh	(sh)	x
m	h		r	n	l
w					y
	pas	de	consonne	initiale	

figure 3

L'insertion du 'h' dans la ligne (m, r, n, l) est justifiée par l'examen de la marque 每 laissée en suspend jusque là, qui, comme la marque 黑, rapproche 'h' de 'm', ainsi que par la marque 户, qui rapproche 'h' de 'l'. Ceci permet en ricochet d'expliquer la dispersion de la marque 爰 qui bénéficie du rapprochement entre 'h' et 'n'.

La duplication du 'h' dans la famille (s, sh, x) est aussi facile à justifier grâce à la marque 行. Ceci étend par ricochet l'influence de la marque 亘 et celle de 巷, pour laquelle 'h' sert de pivot entre 'g' et 'x'. Ceci permet aussi d'organiser le spectre de la marque 刁 sur les deux paires en correspondance 'k/q' et 'h/x'.

Enfin l'insertion de la consonne 'f' dans la famille (h, s, sh, x) permet d'étendre le spectre des marques 林, en rapprochant 'f' de 's', 火, en rapprochant 'f' de 'h', et 犬 en rapprochant 'f' de 'x'. On notera en passant qu'en français la série 'f', 's', 'x' forme avec la série 'l', 'm', 'n', 'r' l'ensemble des consonnes dont le nom ne s'appuie pas en français sur une voyelle finale.

La figure 3 donne bien à la consonne initiale 's' la possibilité théorique d'occuper la cellule vide à sa gauche et donc, en devenant la voisine de 'g', de rendre compte du spectre phonétique (gong, song) de la marque 公 . Néanmoins c'est la consonne 'h' qui statistiquement est de loin la plus sollicitée pour occuper à sa droite cette même cellule vide. Du coup, quitte à faire une exception pour 公, le fond de carte optimal utilise en pratique une matrice des initiales légèrement différente donnée dans la figure 3a suivante:

b	g	z	zh	d	j
p	k	c	ch	t	q
f	h	s	sh	(sh)	x
m	(h)		r	n	l
w					y
	pas	de	consonne	initiale	

figure 3bis

On a ignoré jusqu'à présent la construction des fonds de carte phonétique pour les parties médiane et finale des sons, dont le résultat est donné dans les figures 4 et 5. La seule différence notable est que, pour la partie médiane, on représente la proximité par l'appartenance à la même ligne ou la même colonne mais aussi par l'appartenance à la diagonale descendante des voyelles simples (a, e, i, o, u, ü). Ce manque d'intérêt relatif reflète simplement la priorité donnée dans cet article à la consonne initiale eu égard à la plus grande complexité de son décodage.

a		ai	ao		
	e	ei			
ia	ie	i	i(a)o	iu	
			o	ou	
ua	ue	u(a)i	uo	u	
	üe				ü
(er)	-	(pas	de	Voyelles)	

figure 4

-
n
ng

figure 5

On souligne par contre que les conventions arbitraires imposées par le pinyin sur la transcription phonétique des sons, notamment de leur partie médiane, ne semble pas avoir de conséquence grave. Par exemple la forme initiale du fond de carte présentée en figure 4 contenait une ligne et une colonne spéciales pour la voyelle 'i' placée après une consonne des familles (z, c, s) et (zh, ch, sh, r), voyelle radicalement différente de la voyelle 'i' ordinaire. Au cours de l'optimisation cette ligne et cette colonne ont ensuite disparu par compaction, une forme d'optimisation déjà rencontrée pour la matrice des consonnes initiales. En effet les deux voyelles 'i' sont étroitement associées entre elles, comme le prouve le spectre de 支, avec deux pics respectifs de 4 et de 5 sur ces voyelles à l'exclusion de toute autre, et comme le confirme l'extension des spectres de nombreuses marques où domine l'une des formes de 'i' mais où l'autre est aussi présente, comme 其 (斯) et 示 (示), pour le 'i' ordinaire, 只 (积) et 旨 (稽) pour l'autre. Mais de plus chacune de ces formes revendiquent la même proximité avec les mêmes voyelles et diphtongues, comme on peut le voir pour la diphtongue 'ai' dans le spectre des marques 此 et 台 ou la diphtongue 'ie' dans le spectre de 失. Faire d'une forme un pivot pour l'autre n'est pas une solution car les spectres de 圭, où n'intervient que le 'i' ordinaire, et de 失, où n'intervient que l'autre forme de 'i', imposeraient alors des choix de pivot incompatibles. Attribuer la même cellule aux deux 'i' est donc optimal et justifie a posteriori la convention analogue du système pinyin.

On notera encore que les deux triphthongues 'uai' et 'iao' reconnues par le pinyin, sont classées en faisant abstraction de leur voyelle centrale comme si elles s'écrivaient 'ui' et 'io'. Dans le premier cas, ceci revient à faire partager à 'uai' et 'ui' la même cellule en suivant un raisonnement de compaction semblable au précédent pour les deux voyelles notées 'i'. Le spectre de 鬼 rapproche 'uai' (槐) de 'ui', celui de 不 rapproche 'uai' (坏) de la ligne et de la colonne se croisant en 'ui' et pour finir le spectre de 果 et de 火 demandent que 'uai' (踝) et 'ui' (灰) soient tous les deux pivots. Dans le deuxième cas au contraire, rien ne rapproche 'iao' de 'io' mais les deux médianes se rapprochent toutes deux des deux voyelles 'i' et 'o', comme le montrent pour 'io' les spectres de 京 (琼 avec 京), de 弓 (穹 avec 引 et 弓) et de 宏 (宏), et pour 'iao' ceux de 缴 (缴 avec 激), de 示 (标 avec 际 et 宗) et de 天 (沃). Leur faire partager la même cellule devient la seule solution possible si l'on veut aussi conserver la proximité de 'iao' avec 'ao' (尧) et celle de 'io' avec 'ia' (京). Tout revient donc à faire abstraction du milieu de la triphthongue, ce qui généralise de façon formelle le choix du pinyin de ne pas reconnaître en tant que telles les triphthongues 'iou' et 'uei' mais de les abrégé respectivement en 'iu' et 'ui' [Parlons Chinois, Zhitang Yang-Drocourt, 2007 L'Harmattan, pp 143] .

justification de la méthode itérative

On pourrait se demander s'il n'aurait pas été plus judicieux de laisser de côté la présentation et l'illustration de la méthode et de chercher au contraire à relier directement le résultat final à une liste exhaustive de points d'ancrage. Remplacer ainsi un cheminement fait de tâtonnements successifs par ce qui deviendrait alors une vraie démonstration serait en effet plus satisfaisant d'un point de vue théorique mais, d'une part, travestirait la démarche réelle de l'auteur et, de l'autre, biaiserait les réactions des critiques.

La seule critique acceptable n'est pas en effet de trouver un contre exemple au résultat atteint, car on trouvera toujours des exceptions dans la description d'un système d'origine humaine, mais bien de proposer un résultat global qui s'avèrerait meilleur que celui donné ici, c'est à dire qui, en même temps et mieux que l'auteur, maximiserait globalement l'influence phonétique individuelle de l'ensemble des marques sur leurs caractères dérivés et minimiserait globalement la dispersion du spectre phonétique individuel de l'ensemble des marques. Or on avance qu'une telle critique constructive sera amenée à emprunter cette méthode même pour arriver à un nouvel optimum par un autre cheminement.

-5- Application de la méthode itérative

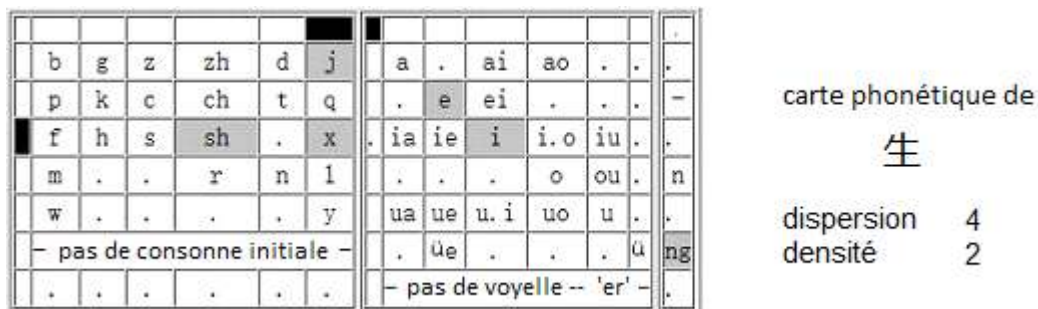
la règle fondamentale du codage phonétique

Le résultat essentiel de l'approche adoptée dans cet article est l'établissement du fond de carte phonétique contenu dans les figures 3, 4, 5 ci-dessus. On affirme que ce dernier permet une description optimale du codage de l'influence phonétique des marques au sens suivant. A toute marque peut être associée une carte phonétique particulière sur laquelle on a colorié son spectre d'influence de sorte que les sons de tous les dérivés phonétiques de cette marque se projettent sur cette aire coloriée. Apparaît alors la règle fondamentale du code qu'il s'agissait de déchiffrer: cette aire coloriée est connexe sur ses trois dimensions, initiale, médiane et finale, c'est à dire qu'aucune cellule coloriée n'y apparaît isolée mais peut être reliée aux autres en parcourant un trajet continu par lignes et colonnes (et la diagonale descendante pour les médianes). A chacun des coudes de ce parcours, la cellule située au croisement des deux directions concernées fait partie du spectre et fait ainsi office de pivot. Autrement dit on peut se passer d'ajouter des pénalités lors du calcul de la dispersion phonétique. Certes le résultat est imposé par la structure de la matrice des finales et les exceptions pour les médianes ne sont pas rares mais, pour les initiales, celles qui ne peuvent pas s'expliquer simplement se comptent quasiment sur les doigts d'une main, à savoir 6 sur 1254 marques au total [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation].

Ce résultat remarquable est évidemment trivial pour toutes les marques de dispersion nulle et de densité infinie. Pour les autres marques, leur carte phonétique est comme une signature visuelle immédiate. Voici par exemple la carte très simple de la marque $\hat{\Delta}$ qui occupe une seule colonne des consonnes initiales, et ne comporte donc aucun pivot. Qualitativement plus la mesure de dispersion est grande, plus l'influence phonétique de la marque se dilue sur une aire étendue; plus la mesure de densité est forte, plus ses pics sont accentués. Les aspects quantitatifs de ces mesures sont développés dans [Lire le chinois - analyse quantitative, Ph Coueignoux, en préparation] et les chiffres ci-dessous ne sont donnés que pour ne pas tronquer l'information disponible.



Voici une autre carte, pour la marque 生, qui, elle, présente un pivot, en l'occurrence la consonne initiale 'x' au croisement de la ligne (f, h, s, sh, x) et de la colonne (j, q, x, l, y) et qui relie ainsi les cellules 'sh' et 'j':



On notera que la 'mesure de Coueignoux' 'favorise' l'influence de la consonne initiale sur les colonnes par rapport aux lignes en comptabilisant 2 pour une paire en ligne au lieu de 1 pour une paire en colonne. Ceci reflète le fait que les spectres phonétiques se déclinent plus souvent le long d'une colonne que d'une ligne.

La règle fondamentale du codage phonétique peut s'interpréter de la façon intuitive suivante: l'influence d'une marque fait en quelque sorte tache d'huile et s'écoule le long des lignes et colonnes des matrices de la carte, ainsi que de la diagonale descendante de la matrice de la partie médiane. Cet écoulement fait un coude à chaque pivot et plus il faut de pivots pour assurer la connexité de la carte, plus l'influence phonétique est diffuse et complexe, le plus souvent témoin soit d'une grande ancienneté de la marque, soit d'un bouleversement récent dû à la simplification des caractères.

A la différence de la structure du fond de carte phonétique, les mesures de dispersion et, partant, de densité dites de Coueignoux ne s'imposent absolument pas. Mais on postule que les autres mesures que l'on pourrait adopter dans le même but se révéleront être cohérentes avec les données obtenues par celles-ci.

application naturelle

Armé du code de l'influence phonétique ci-dessus, il est alors possible de reprendre la liste exhaustive de toutes les marques et, pour chacune, de déterminer son aire d'influence phonétique maximale compatible avec ses caractères dérivés d'une part et de l'autre avec la règle de codage. Le dictionnaire des marques établi par l'auteur est l'aboutissement de cet exercice.

La plupart des marques ne pose pas de problème particulier mais il faut savoir prendre en compte trois phénomènes.

En premier lieu, dans un certain nombre de cas, un caractère peut prétendre à être le descendant phonétique des deux parents dont ils dérivent et il faut trancher en faveur de l'un d'entre eux pour respecter l'hypothèse d'unicité. On attribue alors l'influence phonétique au parent dont la mesure de dispersion augmentera le moins ou le spectre s'étendra le plus, ce qui maximise la cohérence globale du codage.

Le caractère 𠄎 (mǐn) par exemple correspondrait à un son unique à la fois pour la marque 𠄎, de spectre avéré {men, wen}, et pour la marque 𠄎, variante de la marque 文, de spectre avéré {wen}. Mais bien que le rattachement au spectre de 𠄎 n'augmenterait ce spectre que de 1 (pour la diagonale contenant la paire ('e','i')), donner la préférence à 文 permet d'admettre dans le spectre de ce dernier des caractères comme 吝 (lìn) et 𠄎 (liú), par proximité du 'l' avec le 'm', ou, si l'on a déjà décidé que le 'l' faisait partie des consonnes initiales de 文, de supprimer la pénalité de non connectivité de la carte de ses initiales en admettant dans le spectre le pivot 'm'. Noter que cette décision est en accord avec l'analyse de 文 par Karlgren, déjà cité (K 1315) et se trouve corroborée par l'existence d'autres caractères dérivés de 文 qui ne font pas partie du corpus exploité (𠄎, 𠄎, 𠄎, 𠄎, 𠄎) mais sont tous listés par la source mdbg.net sous le son 'min'. Noter aussi que Karlgren hésite mais n'écarte pas non plus la présence de 吝 (lìn) (K553) dans le spectre de 文. On voit comme la recherche de cohérence globale renforce les décisions qui, prises isolément, seraient tout à fait contestables. Voici en définitive les deux cartes phonétiques concernées.

A ce stade il est loisible d'étendre son influence phonétique sur la colonne 'zh'/'ch'/'sh', avec 肘 (zhǒu), 衬 (chèn), 守 (shǒu), 寿 (shòu) et finalement de la compléter par les sons 'si' de 寺 (sì) et 'shi' de 时 (shí). On obtient ainsi une carte cohérente compacte, de densité supérieure à 1 mais dont la dispersion a crû jusqu'à atteindre 13,5.

Or rien n'empêche de chercher à étendre encore cette carte au son 'fu' (fù) et à disputer victorieusement le son 'ru' (rǔ) avec 辰, au prix d'une augmentation supplémentaire de la dispersion à 15 sans diminution de la densité. On ne l'a pas fait uniquement parce qu'une dispersion phonétique de 15, même sans pénalité, semble remplacer la notion d'influence par celle d'hégémonie. Or l'hégémonie pure rejoint le hasard en ce sens que, dans les deux cas, tout devient arbitraire, tout leur est permis.

Rappelons enfin que tout système humain sort toujours du champ d'application de règles découvertes a posteriori, ne serait-ce que parce qu'en le faisant évoluer, on peut choisir de violer toute règle connue d'avance. Ceci vaut pour la règle fondamentale de codage ci-dessus. Par contre la majorité des exceptions sur la première dimension, la consonne initiale, s'explique en appliquant deux règles secondaires, qui, pour la première, accepte de prendre en considération des caractères hors du corpus et, pour la seconde, admet que l'influence phonétique peut sauter une génération et passer d'une marque au caractère dérivé d'une marque dérivée de cette marque. Pour plus de détails, on se rapportera à [Lire le chinois - quelques résultats originaux, Ph Coueignoux 23/05/19, non publié] (section 5).

hypothèses vraisemblables

Il résulte de ce qui précède que l'hypothèse de maximisation de l'influence phonétique permet au dernier stade de l'analyse de se passer de justification. On étend l'aire d'influence phonétique d'une marque non plus pour en augmenter la cohérence et donc la force d'explication statistique, mais simplement 'parce que ce n'est pas interdit'. Une telle démarche est évidemment à la fois subjective et sujette à critique et la frontière entre ce qui peut être justifié et ce qui ne l'est pas est floue.

Par exemple, pour la marque ci-dessous,

b	g	z	zh	d	j	a	.	ai	ao	.	.	carte phonétique de 羊 / 羴 / 羴 dispersion 9,5 densité 1,79
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' ---						
.	
.	
.	

l'inclusion du 'g' dans l'aire d'influence phonétique repose sur la coïncidence de réemploi du 'g' dans le caractère simplifié 盖 (gài) attribué à la marque 羊, en correspondance potentielle étroite avec un dérivé préexistant 羔 (gāo). Sachant que le pur hasard donne une probabilité inférieure à 1% de tomber à partir du son 'gai' sur un son de même initiale, de même finale et de médiane voisine, à savoir 'gao', on peut trouver vraisemblable que la simplification du caractère traditionnel, 蓋, s'est faite sur des considérations d'influence phonétique plus que sur de pures considérations graphiques. Le tracé de la frontière est une affaire de goût.

-6- Conclusion

L'approche qui vient d'être présentée se veut factuelle et pratique. Ses résultats doivent donc être confrontés d'une part aux résultats d'autres approches scientifiques et de l'autre aux pratiques pédagogiques. Ceci fait l'objet de deux articles précédents de l'auteur, intitulés "La structure phonétique du chinois - quelques résultats originaux" et "Lire le chinois - méthode", non publiés pour l'instant.

Dans l'ensemble l'approche retrouve l'essentiel des connaissances traditionnelles tirées de la phonologie. Elle donne aussi, avec le concept de carte phonétique, un support visuel original dont on peut penser qu'il sera utile à la mémorisation des caractères par les étudiants en les remplaçant d'emblée dans leur contexte phonétique.