

PROBLEME DE LA CLASSIFICATION DES CARACTERES CHINOIS
(THE CHINESE CHARACTER CLASSIFICATION PROBLEM)

Marion R. Finley, Jr.
Professeur agrégé
Département de Mathématiques - Université Laval

RESUME*

Le problème de la classification des dessins à deux dimensions que forment l'ensemble des 40,000 caractères chinois est présenté. La classification traditionnelle selon des formes significatives appelées les radicaux est esquissée ainsi que des classifications qui en sont dérivées. En utilisant la notion de grammaire formelle à deux dimensions, une classification qui donne pour chaque caractère une formule quasi-algébrique est indiquée. Mentionnées aussi sont des retombées sur le développement d'appareils de composition des textes en chinois en utilisant les techniques de l'affichage graphique et les mini-ordinateurs. (The classification problem is presented for the set of 40,000 two-dimensional patterns known as the Chinese characters. The traditional classification according to certain meaning patterns termed radicals is sketched together with some variants derived from it. Using the notion of two-dimensional formal grammar, a classification is outlined yielding a quasi-algebraic formula for each character. Also mentioned are spin-offs on the development of Chinese text composition devices using graphic-display techniques and mini-computers.)

(*) Travail appuyé par l'octroi no. 2651 du Ministère de l'Education du Québec et par IBM (Canada).

1. AVANT-PROPOS

La langue chinoise est parlée par environ huit cent millions d'habitants de cette planète. Elle représente l'une des plus grandes, des plus stables, et des plus resplendissantes civilisations qu'a jamais connue l'humanité. Le pays où résident la plupart de ces huit cent millions d'âmes domine de plus en plus les manchettes de nos journaux.

Les fruits de cette grande civilisation, tels l'acupuncture et d'autres sciences de la santé ne font que commencer à se faire connaître au monde moderne. Quels sont les fruits toujours cachés dans les écrits chinois, exprimés au moyen de cette écriture si compliquée, si énigmatique, si difficile à rendre en langue européenne?

C'est dans l'esprit de vouloir ouvrir au monde la littérature chinoise, contemporaine aussi bien que classique, qu'une étude des problèmes du traitement par ordinateur de la langue chinoise a été lancée à l'Université Laval dans le cadre du projet intitulé "Etude de systèmes d'information à très grande capacité" fondé par le Ministère de l'Éducation du Québec et par IBM Canada. Des dictionnaires prototype chinois-anglais-français ont été implantés sur le système à temps partagé APL/LAVAL. Pour plus de renseignements sur ce projet, voir l'article intitulé "Recherche sur le traitement des banques de données en utilisant l'APL/LAVAL" (Finley, 1974).

2. QU'EST-CE QU'UN CARACTERE CHINOIS?

Un caractère chinois est un idéogramme, c'est-à-dire un dessin à deux dimensions qui véhicule une idée. Dès leur invention, vers 2500 avant J.C., les caractères ne représentaient que des objets comme un homme, une maison, un feu, l'eau, un arbre, etc., sans aucune indication de la prononciation des mots correspondant à ces objets. Mais les scribes se sont aperçu qu'ils devraient modifier les caractères pour traiter des idées abstraites et pour indiquer leur son. Quant au problème de développer des idéogrammes pour représenter des idées abstraites, les scribes y ont très bien réussi. Quant à la prononciation, c'est là le point le plus faible de l'écriture. Il reste toujours vrai que les idéogrammes ne donnent qu'une maigre aide à la prononciation.

Le développement de l'écriture chinoise est donné en beaucoup de détails ailleurs, voir par exemple Forrest (1965), chapitre 2. Aux fins de cet article, que cela suffise de mentionner qu'un idéogramme moderne consiste en principe en deux parties dont le radical ou la partie significative (qui implique le sens du caractère) et la partie phonétique qui donne une certaine indication de la prononciation du mot correspondant (voir Figure 1).

A propos de la prononciation des mots correspondants aux caractères, il est à noter que ce sont des mots d'une seule syllabe: donc, un caractère - une syllabe et on peut bien dire que la langue chinoise classique est monosyllabique. Ceci donne une compacité énorme à la langue classique, car toute l'information nécessaire est contenue dans

Figure 1. Quelques caractères chinois.

<u>forme antique</u>	<u>forme moderne (écrite)</u>	<u>forme moderne (imprimée)</u>	<u>signification</u>
			homme
			fil
			femme
			montagne, colline
			chariot, véhicule

les caractères. Pour illustrer ce point, un court texte et sa traduction française est montré dans la Figure 2.

Vu qu'il y a environ 40,000 caractères, dont 10,000 sont effectivement employés et que de ces 10,000 environ, 4,000 sont vraiment nécessaires et que d'autre part il n'existe qu'environ 1,500 syllabes distinctes, on voit que plusieurs idéogrammes auront exactement la même prononciation (le ton y compris, voir Figure 1). Si on essayait de parler la langue classique, on aboutirait à bien des ambigüités. Pour contourner ce problème, la langue parlée et la langue écrite moderne qui se rapproche de la langue parlée, se servent des mots consistant en deux à six syllabes. Donc, la langue moderne n'est plus strictement monosyllabique, cependant elle reste assez compacte en comparaison avec les langues européennes. Ceci mène à la complication que certains caractères ne sont employés que dans certaines combinaisons avec d'autres caractères. Etant donné quelques lignes d'un texte, il faut distinguer entre les occurrences des caractères seuls et les combinaisons. Souvent seul le contexte peut nous aider dans ces cas.

Un soulagement pour le novice ici c'est que la grammaire et la prononciation du mandarin, comme on appelle la langue officielle de la Chine, sont assez faciles à apprendre pour les occidentaux.

La transcription en caractères romains employée dans cet article est le pinyin, celle adoptée par le gouvernement de la République Populaire de la Chine. Ce dernier a introduit des formes simplifiées pour environ 2,000 des idéogrammes ayant une haute fréquence d'occurrence.

3. LA CLASSIFICATION DES CARACTERES CHINOIS

3.1 Le problème

Le problème se pose: comment arranger les caractères chinois afin d'en créer des dictionnaires? Vu qu'il n'y a pas d'alphabet, mais plutôt des dessins, comment les ordonner? C'est un problème qui a vexé les anciens et, malgré tout, il continue à nous vexer aujourd'hui car toute solution proposée comporte certains désavantages importants. Bien que certaines solutions fonctionnent (plus ou moins), il reste que l'acte de trouver un nouveau caractère dont on ignore la signification peut être des fois une tâche difficile et frustrante, même pour l'expert.

3.2 Classifications non-synthétiques

Il y a deux façons de classer les caractères chinois - soit les méthodes traditionnelles et leurs formes dérivées présentées dans cette section et quelques méthodes récentes à décrire dans la section suivante. Les méthodes de la première classe sont basées sur des principes ad hoc et souffrent tous de certains désavantages majeurs, par exemple, la classification est incomplète, elle est difficile à appliquer, ou ses classes d'équivalence sont trop grandes, et ainsi de suite. Les méthodes de la deuxième classe, par ailleurs, sont plutôt basées sur des techniques formelles et synthétiques de description des objets deux-dimensionnels.

Figure 2. Compacité de la langue classique.

(a)	知	zhī	connaître
(b)	之	zhī	le, la, les (pronom du cas objectif, 3 ^e personne)
(c)	者	zhě	participe (voir la terminaison -ing de l'anglais)
(d)	不	bù	non, pas
(e)	如	rú	semblable, comme
(f)	好	hào	aimer
(g)	之	zhī	voir (b) ci-haut
(h)	者	zhě	voir (c) ci-haut

traduction: ceux qui le savent ne sont pas égaux à ceux qui l'aiment.

b a(c) d d e g f(h)

ou bien aussi:

celui qui le sait n'est pas égal à celui qui l'aime.

b a(c) d d e g f(h)

Seulement les mots soulignés paraissent dans le texte chinois. Remarquer que la langue ne fait aucune distinction quant au nombre ou au genre. Le texte chinois n'a que 8 mots tandis que le texte français en a 13. Noter aussi que les deux premiers caractères (a) et (b) possèdent exactement la même prononciation, bien qu'ils aient de significations complètement différentes.

3.2.1 Classification selon le nombre des traits. Une façon de classer les caractères se suggère tout de suite, soit par le nombre de traits employés pour dessiner le caractère. (voir Figure 3). Ce nombre peut varier de un à un maximum d'une trentaine et sa valeur moyenne se situe vers vingt. Cette façon de classer les caractères donnerait trente-deux classes d'équivalence, où chaque classe contiendrait en moyenne $10,000/32 = 300$ caractères (tout en se limitant à 10,000 caractères maximum!) Donc, on rejette cette classification.

3.2.2 Classification selon les radicaux. Les chinois ont tôt reconnu le fait qu'ils pouvaient classer les caractères selon leurs parties significatives. L'empereur Kang Xi trouva 214 radicaux par moyen desquels la plupart des caractères pouvaient être identifiés. Donc, pour trouver la définition d'un caractère dans un dictionnaire arrangé de cette façon, on n'a qu'identifier son radical, ensuite regarder dans la liste de tous les caractères ayant ce radical et trouver le caractère. Dans cette liste les caractères sont ordonnés selon le nombre de traits. Donc, les classes d'équivalence contiennent en moyenne $10,000/214 = 50$ caractères, ordonnés selon le nombre de traits. (Voir Figure 4).

Les difficultés de cette méthode sont les suivantes:

- (1) il n'est pas facile de se rappeler des 214 radicaux,
- (2) des fois il est difficile d'identifier le bon radical parmi ceux contenus dans un caractère, (3) dans certains caractères, le radical est obscur à cause d'un changement de sa forme.

Malgré ces inconvénients, cette méthode reste la plus populaire. Une fois maîtrisé les radicaux, on peut retrouver sans trop de peine la plupart des caractères cherchés.

3.2.3 Classification selon la forme géométrique. Il y a eu quelques tentatives de classer les caractères selon leur forme géométrique dont le "système à quatre coins" inventé par Y.W. Wong en 1926 et ses formes dérivées. Ce système est trop compliqué à expliquer ici; voir Simon, 1949.

Une variation employée par les russes se sert du trait le plus bas et le plus à droite. Ce système est très pratique car on n'a plus de mémoriser les radicaux, mais les classes d'équivalence sont trop grandes. Voir Grigoriev, 1956.

3.2.4 Classification de Lin Yu Tang. Lin Yu Tang, auteur et écrivain assez bien connu dans l'occident pour ses traductions des oeuvres classiques chinoises vient de sortir un dictionnaire (Lin Yu Tang, 1972) en se servant d'un mélange des méthodes géométriques et la méthode des radicaux. A partir d'un "alphabet" de traits et en utilisant un nombre restreint des radicaux traditionnels, il dérive un code que représente la classe d'équivalence du caractère. (Voir Figure 5). Les classes contiennent environ vingt caractères en moyenne avec une dispersion assez étroite, i.e. il n'y a que trois ou quatre classes ayant une cinquantaine de caractères.

Figure 3. Dessin d'un caractère.(a) Les traits fondamentaux

Tout caractère chinois consiste en certains traits choisis parmi l'ensemble indiqué ci-dessous en placés dans un carreau imaginaire tout en suivant la règle générale suivante: dessiner d'en haut en bas, de gauche à droite, les longs traits horizontaux avant les longs traits verticaux (voir exemples dans (b)).

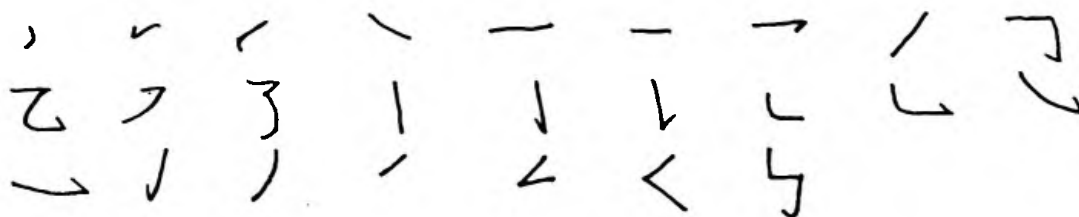
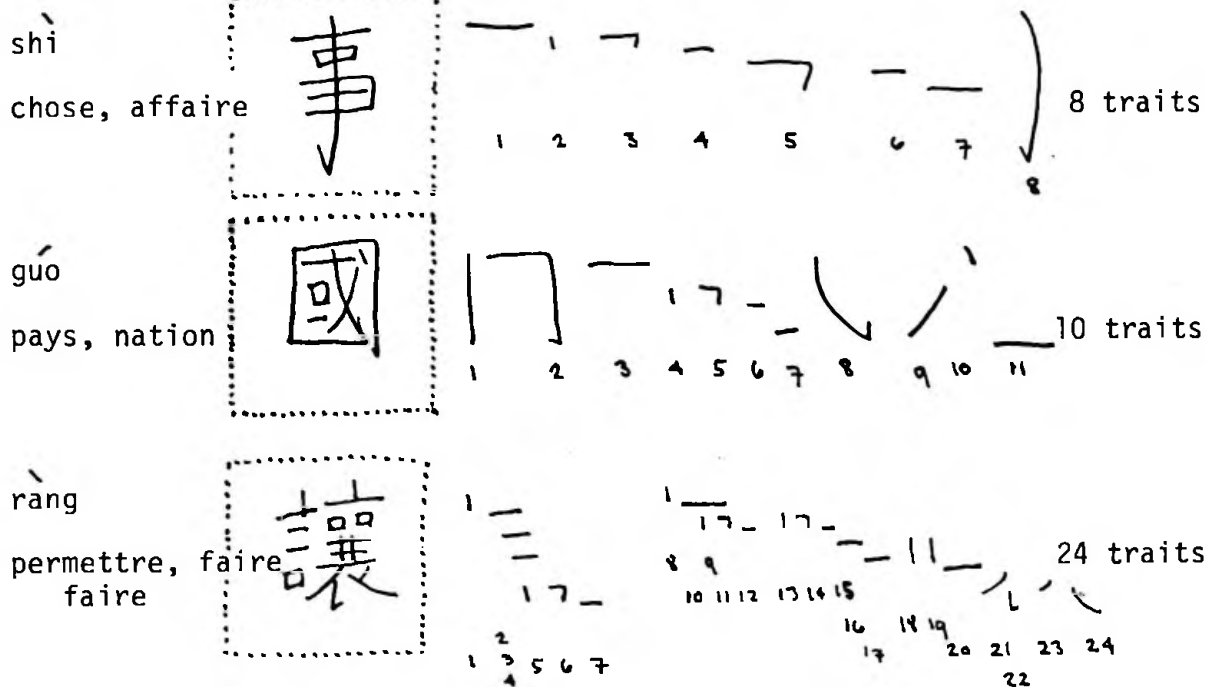
(b) L'écriture de quelques caractères

Figure 4. Classification Kang Xi.(a) Exemples des radicaux (voir Fenn, 1972)

Les radicaux sont ordonnés selon le nombre de traits. Ceux ayant le même nombre de traits sont rangés selon un ordre établi par la tradition. Il y a 214 radicaux.

	<u>1 trait</u>		
1 ⋮	一 ⋮	一 i	un
	<u>3 traits</u>		
30	口	口 kou	bouche
31 ⋮	囗	囗 wei	clôture
42 ⋮	小	小 xiao	petit
46 ⋮	山	山 shan	montagne
	<u>7 traits</u>		
147 ⋮	見	見 jian	voir, percevoir
149 ⋮	言	言 yan	mot(s)

(b) Organisation d'un dictionnaire chinois (voir Fenn, 1972)

Afin de trouver la définition d'un caractère inconnu, on identifiera son radical, ainsi obtenant un chiffre entre 1 et 214. Ensuite on cherchera sous la liste correspondante à ce radical le caractère, s'il est contenu dans cette liste, pointerà alors à la page où se trouve la définition.

Exemple:

shou 說 (parler, 14 traits) radical no. 149, (yán-mots)

Le radical en est numéro 149 sous la liste correspondant à ce radical on trouve une centaine d'entrées ordonnées selon le nombre de traits de la composante non-radical (ici 7):

<u>caractères</u>	<u>page où se trouvent les définitions</u>
言	623
言	144
言	⋮

Fig. 4 (suite)

話

:

201

:

説

469

:

le caractère cherché se
trouve à la page 469

調

249

:

Figure 5. Classification de Lin Yu Tang

Le principe c'est de déterminer de la forme géométrique du caractère lui-même la cote numérique de classification. La procédure procède ainsi: on regarde le trait le plus en haut et le plus à gauche, ensuite le trait le plus en bas et le plus à droite. Le Dr. Lin propose un "alphabet" qui inclut tous les traits qui pourront se trouver dans ces deux positions-son "alphabet" avec toutes les inflexions des traits y comprises consiste en une centaine de formes. Il reste cependant vrai qu'il est probablement plus facile de trouver un caractère en suivant sa méthode qu'un suivant le système Kang Xi.

Exemple: 說 shuō (parler)

trait supérieur à gauche: 丩 cote: 60

inflexion-cas spécial du radical 言 cote: 60A

trait inférieur à droite: 乚 cote: 70

Donc, la cote du caractère est 60A70. Il y a une quinzaine de caractères ayant cette cote. Les entrées dans le dictionnaire sont ordonnées selon l'ordre alphabétique de leurs cotes.

3.3. Classifications synthétiques

3.3.1 Nécessité d'une description formelle des caractères chinois.
 Pour revenir à la question primordiale, c'est-à-dire, de comment classer les caractères chinois, il serait utile de demander "qu'est-ce que c'est qu'un caractère chinois?" Mais cette fois la question se pose du point de vue de faire une tentative de donner une réponse formelle à la question en utilisant les outils formels à notre disposition. Donc, on cherche une description formelle des éléments de E_Z , l'ensemble de tous les caractères chinois.

Les propriétés d'une "bonne" description des éléments de E_Z devraient être au moins les suivantes:

- Complétude: chaque objet de E_Z possède une description
- Consistance: une description donnée ne s'applique jamais à plus qu'un objet
- Unicité: un objet n'aura qu'une description
- Simplicité: les descriptions devraient être faciles à déduire des objets
- Compacité: les descriptions devraient être compactes, i.e. facile à saisir et manipuler.

Si on réussissait à trouver une bonne description pour E_Z , on pourrait alors employer les descriptions comme la base de classification pour E_Z . Noter que dans ce cas les classes d'équivalence ne contiendraient qu'un seul élément, le caractère décrit. Si la description était vraiment commode, on pourrait peut-être créer une machine-à-écrire chinoise assez compacte et utile.

Dans les sections à suivre, quelques résultats de l'auteur seront présentés. Ces résultats sont basés sur Finley (1973) et Leban (1971, 1973). D'autres auteurs ont employé des idées semblables (e.g. Nagao, 1972).

3.3.2 Description formelle de E_Z . Lorsqu'on écrit un caractère chinois, on s'imagine d'abord que le caractère sera imbriqué dans un carreau, ensuite suivant l'ordre d'écriture pour le caractère, on met dans ce carreau les traits l'un après l'autre en leurs relations spatiales l'un par rapport aux autres et en rapport avec le carreau (voir Figure 6). Les traits sont choisis d'un ensemble qui consiste en une vingtaine de classes de traits (voir Figure 3).

Donc, il y a un alphabet fini d'objets qui servent à bâtir n'importe quel élément de E_Z en suivant certaines règles de composition. Cette idée nous mène à proposer une grammaire à deux dimensions (Narasimhan, 1964, Finley 1973, Nagao 1972)

$$G = \langle S, L, x \rangle$$

- où S = l'alphabet des traits (symboles terminaux)
- V = l'alphabet des symboles non-terminaux
- L = liste finie des règles de production en deux dimensions
- x = le symbole du départ

Les règles de L, au lieu de concaténer les symboles comme l'on fait en une dimension, les compose en deux dimensions où on remplace la concaténation par une opération de configuration, $\sqcup$, avec ses inflexions.

Des exemples de cette méthode sont présentées dans Figure 6. Pour plus de renseignements, voir les références déjà citées. Pour une bonne discussion des grammaires formelles comme telles, voir Hopcroft et Ulmann (1969) ou Bochmann (1973). Noter que cette méthode donne une dérivation formelle de chaque caractère et que chaque caractère peut être représenté par une formule quasi-algébrique qui, à son tour, peut être ramenée en notation suffixe sans parenthèse. E.g., le caractère pour gei de Figure 6 aura la formule suivante:

$$((t1 \sqcup t2 \sqcup t3)) \sqcup (t4 \sqcup t5 \sqcup t6)) \sqcup (((t7 \sqcup t8) \sqcup t9) \sqcup ((t10 \sqcup t12) \sqcup t11))$$

qui devient

$$t2 \sqcup t3 \sqcup t1 \sqcup t4 \sqcup t5 \sqcup t6 \sqcup t7 \sqcup t8 \sqcup t9 \sqcup t10 \sqcup t12 \sqcup t11 \sqcup$$

Maintenant pour créer un inde des caractères chinois, les caractères seront d'abord ramenés en forme suffixe sans parenthèse et ensuite ordonnés selon l'ordre "alphabétique" des traits et des opérateurs qui paraissent dans les formules.

3.3.3 Discussion et raffinements. Il est évident que, si les opérations de configuration employées dans cette méthode ont vraiment compris tout ce que fait l'humain lorsqu'il écrit les caractères chinois, la grammaire engendre non seulement tous les éléments de E_7 mais aussi tout candidat possible comme élément de E_7 , i.e., $L(G) \supset E_7$. Aussi, chaque caractère aura une seule unique description. Donc, ces descriptions pourraient servir comme étiquettes pour les éléments de E_7 .

Si l'on accepte comme étant vraie l'hypothèse énoncée ci-haut, le désavantage évident de ces descriptions c'est leur longueur: pour décrire le caractère gei, n'ayant que douze traits, 29 symboles ont été employés.

Pour éliminer ce problème l'auteur tente de trouver des opérations plus puissantes et un ensemble de formes, les traits simples y compris, plus vaste afin de simplifier les descriptions. L'esprit de cette recherche est toujours de soulager le fardeau de mémoriser les 214 radicaux traditionnels.

Bien sûr, une fois définie une forme, celle-ci peut être étiquetée par un symbole conventionnel quelconque et employés dans les expressions pour d'autres formes plus compliquées qui la contiennent comme sous-composante. Par exemple, la bouche dans gei peut être étiquetée B et employée comme telle dans la description de gei. Quelques simplifications sont esquissées dans Figure 7. Celles-ci représentent des recherches en cours sans aucune prétention que le problème ait été résolu.

Figure 6. Dérivation du caractère 給 gěi .

(voir les deux pages suivantes, s'il vous plait)

(a) Liste détaillée des "productions" engendrant le caractère. Les symboles t_1 à t_{12} représentent les traits marqués aux feuilles de l'arbre de (b).

(b) Arbre de dérivation.

Remarques:

Dans ces deux diagrammes, les inflexions de l'opérateur $\square$ sont les suivantes:

- | | |
|---------------|--|
| $a \sqcup b$ | structure a est gauche de structure b |
| $a \sqcap b$ | " en haut de |
| $a \sqcirc b$ | liée à |
| $a \sqcap b$ | le plancher de a est liée à la tête de b |
| $a \sqcap b$ | a est liée au milieu de b |
| $a \sqcap b$ | a est traversé par b |

Donc,

- | | |
|------------------|--|
| $a \text{ 人 } b$ | devient $a \circ \sqcap b$ |
| $a \text{ 人 } b$ | " $a \sqcap \circ b$ |
| $a \text{ 人 } b$ | " $a \sqcup b$ |
| $a \text{ 中 } b$ | " $((a \sqcap c) \sqcap b) \circ \sqcap d$. |

(a)

x	$\rightarrow$	$A \sqcup B$
A	$\rightarrow$	$C \sqcap D$
C	$\rightarrow$	$S1 \perp \boxtimes \circ C1$
$C1$	$\rightarrow$	$S2 \perp \boxtimes \circ S3$
$S1$	$\rightarrow$	$t1$
$S2$	$\rightarrow$	$t2$
$S3$	$\rightarrow$	$t3$
D	$\rightarrow$	$S4 \sqcup S5 \sqcup S6$
$S4$	$\rightarrow$	$t4$
$S5$	$\rightarrow$	$t5$
$S6$	$\rightarrow$	$t6$
B	$\rightarrow$	$E \sqcap F$
E	$\rightarrow$	$S7 \circ \boxtimes S8 \sqcap S9$
$S7$	$\rightarrow$	$t7$
$S8$	$\rightarrow$	$t8$
$S9$	$\rightarrow$	$t9$
F	$\rightarrow$	$S10 \boxtimes S11$
$S10$	$\rightarrow$	$S10 \perp \boxtimes S12$
$S10$	$\rightarrow$	$t10$
$S11$	$\rightarrow$	$t11$
$S12$	$\rightarrow$	$t12$

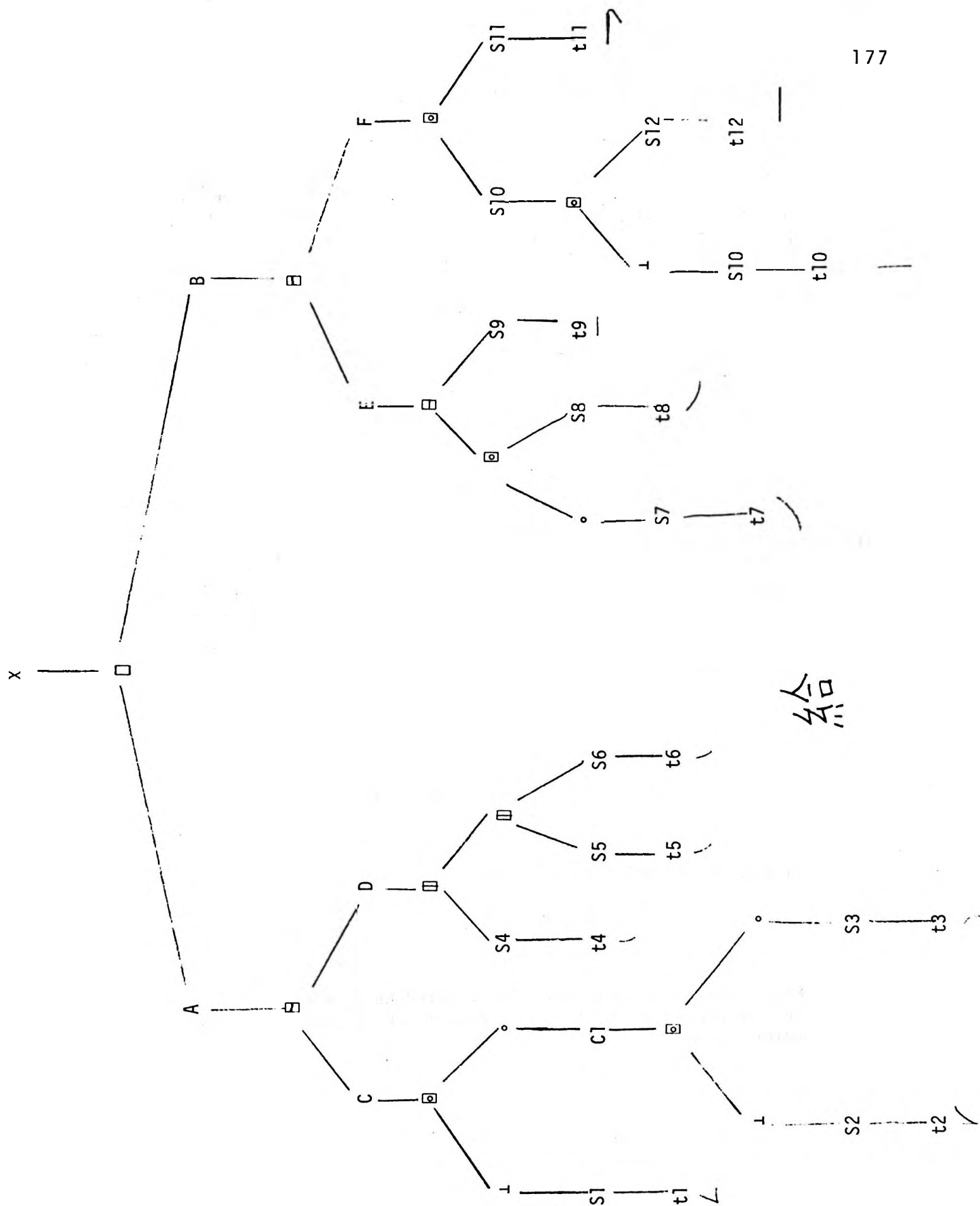


Figure 7 Simplification des descriptions(a) Opérateurs plus évolués

On cherche à généraliser l'opérateur $\square$. Par exemple, on peut l'étendre à des prédicats à n-places, e.g., $C(a,b,c)$ où C voudrait dire qu'on mette l'objet a à gauche et en bas de l'objet b , tandis qu'on mette l'objet c à droite et en haut:

si a était ///
 b était L
 c " F

alors,

$C(a,b,c)$ $C(\text{///}, \text{L}, \text{F})$ serait 馬 (mā ,
cheval)

(b) Extension de l'alphabet de traits

On cherche à étendre l'alphabet des traits fondamentaux comme présentés à la Fig.3. Surtout, on vise à inclure dans cette extension des structures qui sont communes à plusieurs caractères, telles certaines des radicaux et bien d'autres formes. Par exemple,

/// 4 "pieds"
 /// 3 "pieds"
 $\square$ la bouche(un radical)
 $\square$ la clôture(un radical)

et des formes telles les suivantes:

T 乚
 F $\text{}$
 F $\text{}$

Néanmoins, le nombre total de symboles ne devrait pas devenir trop grand, sinon on finira par avoir un "alphabet" trop dur à mémoriser et à manipuler.

3.4 L'AFFICIIAGE GRAPHIQUE ET LES MINI-ORDINATEURS

Une possibilité fascinante ici c'est de se servir de l'affichage graphique et des mini-ordinateurs pour soulager le problème de la composition de textes en langue chinoise. Une console visuelle munie d'une imprimante et branchée à un mini-ordinateur pourrait bien servir comme machine à écrire chinoise. Si les opérations sont suffisamment puissantes et si le nombre de formes n'est pas trop élevé, on peut envisager même un clavier de complexité à peu près égale à celle des claviers standards employés pour l'alphabet latin et, si le coût de l'équipement continue à baisser, même à des prix intéressants (1000 à \$2000). Donc, les implications d'une bonne classification comme esquissée dans la Section 3.3.1 sont d'une portée assez vaste, surtout si l'on considère que les japonais (qui emploient les caractères chinois) et les chinois eux-mêmes ne jouissent pas de la conveniencce qu'offre la machine à écrire!

Le software nécessaire pour appuyer l'affichage graphique des caractères chinois est d'un intérêt particulier, mais hors les limites du kprésent article. Un système expérimental fut développé sur l'APL/LAVAL muni d'une console Tektronix 4013. Un dictionnaire interactif des termes techniques de l'informatique est en cours de préparation sur l'APL/LAVAL.

4. CONCLUSION

En se servant d'une grammaire formelle à deux dimensions, un modèle pour la description formelle des caractères chinois a été développé. L'opération de concaténation des grammaires formelles ordinaires a été remplacée par celle de la configuration des objets en deux dimensions. Les descriptions obtenues sont ramenées en forme quasi-algébrique suffixe sans parenthèse. Elles peuvent être ordonnées selon un ordre précise sur les traits et les opérations. Ceci donne une classification des caractères. L'hypothèse primordiale est que ces descriptions aient capturé l'essentiel de ce que fait l'humain lorsqu'il écrit un caractère. L'évidence montre que cette hypothèse est probablement vraie.

Il est intéressant à noter comment une écriture qui date de l'antiquité puisse donner un défi si fort à la technologie la plus récente et la plus sophistiquée développée par l'homme, sans qu'on ne sache si celle-là aurait vraiment réussi à résoudre le problème.

REFERENCES BIBLIOGRAPHIQUES

- BOCHMANN, G.V. Langages formels, Département d'Informatique, Université de Montréal, 1973.
- BRANDT, J. Introduction to Literary Chinese, Ungar, New York 1954.
- CHANG, Hsi-Kuo. An interactive system for Chinese character generation and retrieval, IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-3, (mai 1973), pp. 257-265.
- DEFRANCIS, J. Beginning Chinese Reader, Parts I and II, Yale University Press, 1966.
- FENN, C.H. The Five Thousand Dictionary, Harvard University Press, 1972.
- FINLEY, M.R. On the formal description of Chinese characters, paper presented at the First International Symposium on Computers and Chinese Input/Output Systems, Academia Sinica, Taipei (August 1973).
- _____. Recherche sur les systèmes d'information en utilisant l'APL/LAVAL, en préparation, (1974).
- FORREST, R.A.D. The Chinese Language, Faber and Faber, London, 1965, pp. 36-47.
- GRIGORIEV, G.M. Kratkii Kitaïsko-Russkii Slovar' (Petit Dictionnaire Chinois-Russe, Short Chinese-Russian Dictionary), Maison d'Editions de Dictionnaires de Langues Etrangères et Nationales (Gosudarstvennoe Izdatielstvo Inostrannykh i Natsional'nykh Slovariei), Moscou, 1965.
- HOPCROFT, J.E. and J.D. Ullman, Theory of Formal Languages and Their Relation to Automata, Addison-Wesley, 1969.
- LEBAN, C. Graphemic synthesis: the ultimate solution to the Chinese Input/Output Problem, paper presented at the First International Symposium on Computers and Chinese Input/Output Systems, Academia Sinica, Taipei (August 1973).
- LIN, Yu Tang, Lin Yu Tang's Chinese-English Dictionary of Modern Usage, The Chinese University of Hong Kong, McGraw-Hill, Tokyo, 1972.
- NAGAO, M. Picture processing and data structure, in Graphic Languages, F. Nake and A. Rosenfeld: editors, North-Holland, (1972), pp. 48-68.
- NARASIMHAN, R. Labeling schemata and syntactic descriptions of pictures, Info. and Control, 7 (1964), pp. 151-179.
- SIMON, W. How to Study and Write Chinese Characters, Lund Humphries, 2nd revised edition, 1949.