

Com/P 4/1967/MC-Grenoble

Section d'Automatique Documentaire

du

CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE

Analyse et classification automatiques en chinois

par Maurice Coyaud

(Communication au Congrès International de Linguistique mathématique - Grenoble,
Août 1967 -)

On se propose d'abord une rapide revue des travaux soviétiques et américains concernant l'analyse automatique des textes scientifiques chinois ; une seconde partie de la communication est consacrée à un projet de classification nominale (par les particules "numérales" : measure words, classifiers) à l'aide de machine.

1. Essai de synthèse des recherches en analyse automatique.

Dans un premier temps, les recherches ont concerné l'analyse dépendante (ou binaire) des textes chinois : cette analyse était effectuée en fonction de la langue de sortie : l'anglais (travaux de Reifler, Washington), le russe (travaux de Zvonov, 1961).

Mais on s'est aperçu qu'il était plus rentable et plus intéressant de considérer le problème de l'analyse automatique sans tenir compte de la langue de sortie. C'est principalement ces recherches, sur l'analyse indépendante, qui retiendront l'attention (néanmoins, on tiendra compte des remarques pertinentes de Zvonov, issues de son travail sur l'analyse dépendante du chinois, en particulier sur les constructions par encadrement : ramochnye konstrukcii).

Les problèmes de l'analyse syntaxique du chinois ont été définis dans leurs grandes lignes par Voronin (1958), Wang (1963) et Hashimoto (1965). Selon Voronin, on commence par découper la phrase, si elle est complexe, en propositions simples ; on établit ensuite les groupes syntagmatiques et enfin les fonctions logiques : sujet et prédicat. La phrase complexe est décomposée

au moyen de bornes formelles comme la ponctuation, certaines conjonctions ; certaines propositions n'ont pas de frontières formellement bien déterminées : par exemple, les prépositives (ou déterminatives développées avec de : cf. Nikitina, 1965, et le rapport sur la technique Fulcrum de Garvin, 1965), et les relatives sans mot de relation ; ces dernières sont de deux sortes : ou bien introduites par des verbes comme penser, (re)connaître, etc. ; ou bien, en elles, le mot relié au verbe précédent en qualité de complément, sujet inversé ou partie nominale du prédicat est en même temps le sujet d'un groupe de mots plus éloigné. Avant de procéder à la synthèse de la traduction en langue de sortie, l'analyse devra spécifier le nombre des noms (catégorie non obligatoire, et peu fréquente en chinois) et le temps des verbes (compte tenu des cinq aspects du verbe chinois, des verbes modaux et des expressions de temps).

La catégorisation en parties du discours joue le rôle essentiel dans ces opérations et on la justifie en détail (Zerebin 1961 a, Voronin 1961) ; la nature du corpus joue aussi son rôle, et notamment les difficultés introduites par la présence de wényán dans les textes de mathématiques, apportant une concision qui complique la tâche de l'analyste, puisque le système grammatical wényán ne s'incorpore pas, mais se superpose au système báihuà.

A côté des problèmes généraux qui font l'objet des articles qui viennent d'être cités, on trouve un nombre appréciable de monographies étudiant notamment :

- 1) les ambiguïtés de catégorie grammaticale (Zerebin 1961 b : réduction de 5 à 3 types d'ambiguïté, en raison du passage d'un système d'analyse dépendante à une analyse autonome ; Nikitina 1966 ; Zvonov 1961 b : ambiguïtés des mots-outils yòng, yǐ, jiāng, yóu, wèi). La résolution de ces ambiguïtés est jugée trop difficile par Ertel et al. qui la supposent résolue avant le démarrage de leur technique d'analyse syntaxique ;
- 2) les constructions avec de (Fulcrum Report, 1965) ;
- 3) la reconnaissance automatique des groupes nominaux (Peng, 1966) ;
- 4) la recherche du sujet (Nikitina, 1965 : place du sujet ; types de sujets : propositionnel, nominal etc... ; doubleage du sujet par un pronom ou un nom généralisant ; sujets homogènes ; sujets absents ; sujets de subordonnées de détermination avec de) ;
- 5) l'imbrication (Hashimoto, 1966) ;
- 6) l'interrogation (Elliot, 1965) ;
- 7) des déterminants définis et indéfinis (Annear, 1965) ;
- 8) des conjonctions de subordination employées dans les phrases hypothétiques, notamment rúqǔ, tāngshì, tāngguò, jiāshì, yàoshì et leurs corrélateurs nà, jiù, cái, biàn, yě (Coyaud, 1967). Les différents facteurs déterminant

l'emploi de telle ou telle conjonction sont étudiés. Il n'est pas certain que l'emploi de ces conjonctions soit dicté uniquement par des facteurs dialectaux (par exemple jiǎshì viendrait du Sìchuān, yàoshì du Shānsī, les composés de tāng du Zhèjiāng) ; on sait d'autre part que jiù s'emploie quand la principale est brève, nà(mò) quand la principale est longue ; mais les autres critères, déterminant le choix des corrélateurs n'ont pas été dégagés.

2. Classification nominale.

Le chinois contemporain possède une centaine de particules "numérales" déterminant chacune une classe nominale. Le mot "numérale" est gênant, car ces particules ne s'emploient pas nécessairement entre un nombre et le nom, mais aussi entre un démonstratif et le nom ou des adjectifs précédant ce nom. Ceci dit, nous conserverons le mot "numérale", qui est traditionnel ("classificateur" serait trop général). Le nombre des numérales a été souvent exagéré (plusieurs centaines, pour Zhao Yuanren ou Hockett) mais les inventaires ne visaient sans doute pas strictement l'état de langue du chinois contemporain. Pour le quóyǔ, Herrfurth (1964) considère que 48 numérales sont d'usage courant ; un inventaire fondé sur un corpus plus large (Coyaud, 1967 b) relève 84 numérales au sens restreint et 16 numérales indiquant des capacités d'usage ordinaire. Le travail que nous présentons ici a un intérêt propre, qui est théorique ; mais il a aussi un intérêt pratique (§ 2.1).

2.1. Les particules numérales et l'analyse automatique.

En linguistique appliquée, la classification des noms par les numérales a une double importance. D'abord, en pédagogie : il est indispensable d'enseigner avec précision à celui qui veut apprendre le chinois avec quelle(s) numérale(s) chaque nom se combine (l'ignorance de ces combinaisons est aussi pesante que celle du genre des noms en français ou en allemand). D'autre part, en analyse automatique, il est nécessaire que l'automate dispose d'un dictionnaire complet, attribuant à chaque nom la (ou les) numérale(s) qui lui correspond(ent). De tels dictionnaires n'existent encore que de façon partielle. L'information consistant à indiquer les couples numérales-noms dans le dictionnaire automatique a une double utilité :

a) permettre de résoudre des polysémies lexicales ; par exemple yí qè yue "un mois" et yí lún yue "une lune", ne diffèrent que par la particule numérale, qè s'appliquant au mois, et lún à l'astre. Le mot yue est homographe et homophone.

b) permettre de résoudre des ambiguïtés de catégories grammaticales. Les particules numérales sont pour la plupart des homographes pouvant souvent fonctionner comme noms et parfois comme verbes (dài "emporter", juān "enrouler", etc.). Un essai d'établissement automatique des syntagmes nominaux (avec

particules numériques; nécessairement) a été fait par Peng (1966).

On pourrait ajouter à ces deux rubriques une troisième, représentant un champ d'études encore inexploré : l'utilisation de la classification nominale en tant que groupes d'affinités sémantiques, permettant de choisir entre plusieurs corrélations acceptables d'un point de vue purement grammatical (1).

2.2. Classification nominale à l'aide des numériques (problème théorique):

L'étude préliminaire au projet esquissé ci-dessous fait l'objet d'un rapport détaillé (Coyaud, 1967 b) dont nous nous contenterons ici de développer les implications dans la perspective d'un traitement par machine. Le problème consiste à établir avec détail et précision la classification des noms en chinois, à l'aide des particules numériques. En particulier il s'agit de dresser la liste des ensembles disjoints, celle des ensembles conjoints, déterminer exactement les éléments appartenant à chacun de ces ensembles, et en particulier établir la liste des éléments appartenant aux zones d'intersection entre ensembles pris deux à deux, et n à n . La tâche est assez considérable pour justifier une aide mécanique, si l'on se propose effectivement de classer environ 3,000 noms. Nous avons entrepris le travail à la main pour environ 500 noms. Nous donnerons la physionomie générale de la classification telle qu'elle apparaît à l'issue de ce travail. Nous ne nous sommes pas limités à une perspective synchronique (description du système classificatoire des noms du chinois contemporain (CC) ; le problème de l'évolution de ce système est important. On relève en effet peu de numériques en chinois archaïque (cf. Jaxontov et Huang Zaijūn) ; par contre, en chinois ancien (CA) de la période 200.600 de notre ère, Liú Shirú a pu relever 141 numériques propres et 54 numériques exprimant des capacités (sans être pour autant des unités de mesure codifiées, qui ne nous intéressent pas ici). En quoyú ou chinois contemporain (à partir de 1920 environ), notre inventaire comprend 100 numériques (84 propres et 16 de capacité). Les deux listes de numériques, pour le CA et le CC, sont données en annexe. Pour le CA, nous n'avons retenu que les 141 numériques au sens propre du mot, qui seront comparées aux 84 numériques propres du CC pour ce qui est de la classification des noms qu'elles déterminent. Les 141 du CA n'incluent pas toutes les 84 du CC. Ces deux ensembles sont en intersection. Les différences individuelles sont données dans notre rapport détaillé. Indiquons ici simplement que sur les 141 du CA, 80 n'apparaissent pas en CC ; inversement, sur les 84 du CC, 24 n'étaient pas attestées dans les corpus de CA compilés par Liú Shirú.

(1) Pour une utilisation d'un réseau d'affinités sémantiques en analyse automatique de textes français, cf. Coyaud, Siot-Decauville, L'analyse automatique des documents, Mouton, Paris, 1967.

2.21. Relations entre classes nominales en CA.

Les ensembles simples disjoints, au nombre de 11, sont définis par les numérales : fān, qǔ, jiàn 2; kōng, luàn, niǔ, pò, qì, tà, wén, zhì 2. Chacun des 130 autres ensembles est en intersection avec au moins un ensemble de ce groupe. Il y a quatre groupes d'ensembles disjoints parmi ces 130 (l'élément commun est indiqué après) :

- a) bǐng 1, fú 2 : éventail
- b) chàng, qēng, kè : unités de temps
- c) jié 1, jù 1 : verset

d) tous les 123 restants. Le diagramme donné en annexe contient ces ensembles. Une seule intersection a été indiquée à chaque fois (afin d'éviter l'encombrement). Bien d'autres intersections sont attestées. Par exemple, nous n'en avons donné qu'une seule pour qè, alors qu'en CA, d'après les exemples de Liú Shìrú, l'ensemble défini par qè est en intersection avec quinze autres ensembles, définis par les numérales suivantes (l'élément commun est donné après) :

- míng, yuán, jiè, wèi 1, rén, tóu : humains
- fā, jù 2, zhì 3 : flèche
- shù, bēn, zhū, shuāng, cóng : arbres
- zuò : statue de divinité

2.22. Relations entre classes nominales en CC.

En CC (quóyǔ) les résultats fournis par notre enquête donnent 19 classes disjointes et 80 classes en intersection, dont les éléments forment un graphe connexe (voir le second diagramme). L'ensemble défini par qè est en intersection avec 25 ensembles (ces intersections ne sont pas toutes indiquées dans le diagramme), définis par les numérales suivantes (les éléments communs sont donnés dans Coyaud 1967 b) : bā, bù, chuàng 2, dào, dǐng, qēn, jiē, jià, jiàn, jiàn 1, kē 1, kùai, liàng, miàn, míng, piàn, sùo, tiào, tóu, wèi 1, yuán 1, zhì 3, zhēng, zhōng, zuò.

2.23. Le qèhúa.

L'importance exacte du qèhúa est difficile à évaluer d'après ce sondage sur la classification de 500 noms à peine. Signalons que méi en CA semble jouer le rôle de qè en CC. L'augmentation du nombre des substitutions par qè de numérales du CA au CC n'est pas énorme, elles sont passées de 15 à 25. Mais 6 numérales seulement sont identiques dans les groupes des 15 CA et 25 CC.

2.24. Classification à l'aide de machine.

Les résultats donnés ci-dessus ont été obtenus à la main, au terme d'une étude ingrate et incomplète portant seulement sur les classes d'appartenance de 500 noms environ en CA et en CC. Afin d'obtenir des résultats exempts d'erreurs (ce que ne prétendent pas être ceux qui ont été présentés ci-dessus) et portant sur un échantillon important de noms (environ 3.000), on envisage une aide mécanique. Le but recherché est l'établissement correct des classifications de noms en CA et en CC : classes disjointes, classes sécantes, énumération exhaustive des éléments de chaque classe et de chaque zone d'intersection 2 à 2 et n à n. C'est seulement à l'issue de ce travail que l'on pourra avoir une idée précise de la classification des noms par les numérales en chinois, et de son évolution diachronique (en particulier de l'évolution de la numérale générale qè : le qèhua). En outre, le programme sera assez général pour permettre une étude analogue des classifications nominales dans d'autres langues à particules numérales, comme le laotien, le vietnamien, le thai, le malais (Herrfurth), le japonais et le maya du Yucatan (cf. la 2ème partie de notre bibliographie). Ce n'est qu'alors que des comparaisons précises (et non plus vagues comme toutes celles qui ont été présentées jusqu'à présent) pourront être entreprises.

N.B. Les numéros de référence de l'index des numérales du CA donné en annexe se rapportent aux paragraphes de notre rapport détaillé (1967 b), reproduisant pour le CA l'ordre de présentation de Liú Shirú (1965, p. 76 à 223).

Index des numérales du chinois ancien

beì ("cent chars")	輩	1.12 K
běn	本	1.12 B
biǎn	編	1.13 E
bǐng 1 ("manche")	柄	1.12 C
bǐng 2 ("pâté")	餅	1.13 H
bó ("jonque")	舶	1.13 I
bù	部	1.22 D
cáo	曹	1.12 K
cà	冊	1.13 E
céng	層	1.12 J
chàng	唱	1.13 L
chēng	稱	1.22 C
chóng	重	1.12 J
chù	処	1.12 L
chuáng 1 ("lit")	牀	1.12 D
chuáng 2 ("drapeau")	幢	1.12 E
coíng	叢	1.22 A
dào	道	1.12 I
dēng	等	1.12 K
dī	滴	1.12 F
diǎn	奠	1.12 G
dié 1 (trois champs)	疊	1.12 J
dié 2 (clef de la planche 90)	牒	1.12 J
dǔ	堵	1.13 C
duān 1 (clef <u>li</u>)	端	1.12 C
duān 2	段	1.12 H
dùn	頓	1.12 N

fú 1 ("surface")	幅	1.12 I
dù	朵	1.13 K
fā	发	1.21 B
fān 1 (clef des plumes)	翻	1.12 N
fán	幡	1.13 J
fāng	方	1.13 B
fáng	房	1.13 A
fēn	分	1.12 H
fēng	封	1.13 F
fú 2 (clef de la main)	拂	1.12 I
fù	副	1.22 B
gān	竿	1.13 K
gè	个	1.11
gēn	根	1.12 B
gēng	更	1.13 L
gǔ	股	1.12 G
guǎn	管	1.13 D
hán	函	1.13 F
háng	行	1.22 E
hé 1	合	1.22 B
héngsha	恒沙	1.13 M
hù	户	1.22 D
huì	会	1.13 M
jí	级	1.12 J
jì	剂	1.13 B
jiā	家	1.22 D
jià	架	1.21 B
jiān	间	1.12 H
jiǎn	简	1.13 E
jiàn 1	件	1.12 H
jiàn 2 ("verrou")	键	1.13 M
jiē	阶	1.12 J
jié 1 ("versifier", clef de l'homme)	偈	1.13 M

jiè	介	1.11
jiǐng	莖	1.12 B
jù 1 ("parole")	句	1.13 E
jù 2 ("préparer")	具	1.22 B
juǎn	卷	1.13 E
kān	堪	1.13 M
kē 1 (clef de la feuille)	顆	1.12 G
kē 3 (céréale et boisseau)	科	1.12 K
kè	刻	1.13 L
kǒng	孔	1.13 H
kǒu	口	1.12 A
kuài	块	1.12 G
lèi	类	1.12 K
lì	粒	1.12 G
liǎng	两	1.13 I, 1.21 A
liè	列	1.22 E
lǐng	領	1.12 C
liú	流	1.12 K
lǚ	縷	1.12 F
lǚǎn	縷	1.13 B
méi	枚	1.11
mén	门	1.12 K
miàn	面	1.12 I
míng	名	1.12 E
niǔ	鈕	1.13 G
pǐ	疋	1.13 I
piān	篇	1.13 E
piàn	片	1.12 G
pǐn	品	1.12 K
pò	璞	1.13 H
pū	鋪	1.12 I
qì	契	1.13 M

qiāng
 qū 1 ("région")
 qū 2 ("tordre")
 qū 3 ("corps")
 qún
 rén
 shān
 shēn
 shèng
 shì
 shǒu 1 ("tête")
 shù
 shuāng
 sōu
 suì
 suǒ
 tà
 tiáo
 tǐng
 tōng
 tóu
 wán
 wèi 1 ("rang")
 wèi 2 ("odeur")
 wén
 wù
 xí
 xù
 xuán
 yāo

腔 1.12 A
 区 1.12 L
 曲 1.13 E
 躯 1.13 M
 群 1.22 A
 人 1.12 M
 扇 1.13 C
 身 1.13 M
 乘 1.12 I
 享 1.12 M
 首 1.13 E
 树 1.12 M
 雙 1.21 A
 艘 1.13 I
 榼 1.21 B
 所 1.12 L
 沓 1.22 B
 条 1.12 B
 榿 1.12 B
 通 1.12 N
 头 1.12 A
 丸 1.12 G
 位 1.13 A
 味 1.13 B
 文 1.13 H
 物 1.12 M
 襲 1.22 C
 绪 1.12 C
 懸 1.21 B
 腰 1.12 I

yuán 1 (pour les personnes)	員	1.13 A
yuán 2 ("ourlet")	緣	1.13 M
zhāng 1 ("bander un arc")	張	1.12 I
zhāng 2 ("texte")	章	1.13 E
zhèn	陣	1.12 N
zhī 1 (clef du bois)	枝	1.12 B
zhī 2	支	1.12 B
zhī 3 ("oiseau")	隻	1.12 F
zhǐ 1 ("papier")	紙	1.12 M
zhǐ 2 ("doigt")	指	1.13 H
zhì	帙	1.13 E
zhǒng	種	1.12 K
zhóu	軸	1.12 C
zhū	株	1.12 B
zhuàn	轉	1.12 J
zǐ	子	1.12 G
zūn	尊	1.13 M
zùo	座	1.12 D

Liste des Numérales du chinois contemporain

bǎ	把	fā	发
bān	班	fān 2	番 "retourner"
bāng	帮	fēng	封
bāo	包	fú 1	幅 "surface"
bēi	盃	fù	副
běn	本	gān	杆
bǐ	笔	gè	个
biān	编	gēn	根
bù	部	gǔ	股
céng	层	guān	罐
cháng	场	guǎn	管
chē	车	háng	行
chū	出	hào	号
chù	处	hé 2	"boîte"
chuàn	串	hú	壶
chúang 2	幢	jiā	家
dài 1	带	jià	架
dài 2	袋	jiān	间
dào	道	jiàn 1	"part"
dēng	等	jié 2	"noeud de bambou"
dī	滴	jù 1	"parole"
diǎn	点	juǎn	卷
dǐng	顶	kē 1	(clef de la feuille)
dūan 2	段	kē 2	(clef du bois)
dūi	堆	kǒu	口
dùì	对	kuài	块
dùn	顿	kǔn	捆
dǔo	朵	lì	粒
dùo	朵	liǎng	辆

liè	列
lǚ	縷
lún	輪
mén	門
miàn	面
míng	名
pán	盤
pī	批
pǐ	疋
piān	篇
piàn	片
píng	群
qún	群
shàn	扇
shēn	身
shēng	聲
shǒu 1	"tête" 首
shǒu 2	"main" 手
shuāng	雙
sōu	艘
sùo	所
tái	台
tào	套
tiáo	條
tǐng	挺
tǒng	筒
tóu	頭
wǎn	盪
wěi	尾
wèi 1	位
yaǎn	眼
zā	扎

zhǎn	盞
zhāng 1	"bander un arc" 張
zhèn	陣
zhī 2	支
zhī 3	隻
zhǒng	種
zhū	株
zhuāng	樁
zūn	尊
zùo	座

Bibliographie

1. Analyse syntaxique.⁽¹⁾

- Annear, S.S.- (1965) English and Mandarin Chinese : Definite and Indefinite Determiners and Modifying Clause Structures, POLA 11, (p 1-55), Ohio State Univ.
- Coyaud, M.- (1967a) Les phrases hypothétiques en chinois contemporain (Problèmes de méthode). A paraître dans Hommage à A. Martinet. (A. Juilland ed.)
- Elliot, D.E.- (1965) Interrogation in English and Mandarin Chinese, POLA 11 (p 56-117), Ohio State Univ.
- Ertel, L.L., Garvin, P.L., Peng, F.L.L., Reitz, G.- (1965) Fulcrum technique for Chinese. English Machine Translation. The Bunker Ramo Corp. Canoga Park, California.
- Žerebin, M.E.- (1961a) Struktura i karakter slovnika dlja nezavisimogo analiza kitajskogo teksta. ITMiVT, Sbornik statjei, p 173 sq.
- (1961b) O razrešenii konverzionnoj omonimii pri nezavisimom analize kitajskogo teksta, ITMiVT, p 180 sq.
- Hashimoto, A.Y.- (1965) A Condensed Account of Syntactic Analysis of Mandarin. POLA 10, (p 1-27), Ohio State Univ.
- (1966) Embedding Structures in Mandarin, POLA 12, Ohio State Univ.
- Li, K. P.- (1964) Automatic Recognition of Mandarin Monosyllable. POLA 9, Ohio State Univ.
- Nikitina, T.N.- (1965) Obrabotka podležaščego pri MP s kitajskogo jazyka. NTI, 10- 1965 p 37-40.
- (1966) Grammatičeskaja neodnoznačnost' pri avtomatičeskom perevode s kitajskogo jazyka na russkij. NTI, 2- 1966 p 30-33
- Peng, F.C.C.- (1966) A Structural Recognition Routine for the Noun Phrases of Standard Chinese, in Computation in Linguistics, (p 207-251), Garvin ed. Indiana Univ. Press.

- Voronin, V.A.- (1961) Sistema nezavisimogo konfiguracionnogo analiza kitajskogo teksta pri masinno^vm perevode, ITMiVT, p 135 sq.
- Wang, S.Y.- (1963) Some Syntactic Rules in Mandarin, POLA 3, (p 32-53), Ohio State Univ.
- Zvonov, A.A.- (1961a) Analiz ramočnyx konstrukcij pri binarnom masinno^vm perevode s kitajskogo jazyka na russkij. ITMiVT, p 349 sq.
- (1961b) Analiz služebnyx slov pri binarnom masinno^vm perevode s kitajskogo jazyka na russkij. ITMiVT, p 360 sq.

note (1) - Les références de Žerebin, Voronin et Zvonov sont publiées dans les "Travaux de l'Institut de Mécanique de précision et de Technique du calcul, Académie des Sciences, 2e ed. Moscou, 1961 (ITMiVT).

2. Classification nominale.

Numérales du chinois.

- Coyaud, M.- (1967b) Classification nominale en chinois des particules numérales. Rap/P.2/SAD Polycopié, CNRS, SAD, Paris.
- Dragunov.- (1958) Xiandai hanyu yufa yanjiu, Pékin.
- Herrfurth, H.- (1964) Die Klassifikationsaffixe im Chinesischen, Beiträge zum Problem des Wortes im Chinesischen II, p 1-43. Akademie Verlag, Berlin,
- Huáng Zàijūn.- (1964) Cóng jiǎwén, jīnwén liàngcíde yīngyòng, kǎochá hanyu liàngcíde qǐyuan yú fāzhān. Zhōngguó yǔwén n° 6, p 432-441.
- Jaxontov, S.E.- (1965) Drevnekitajskij jazyk. ANSSSR, Moscou.
- Líng Quánfāng.- (1955) Yòng yǔ rénde gè tǐ liàngcí. Yǔyán zhīshí, n° 6.
- Liú Shìrú.- (1965) Wèi Jīn Nán Běi cháo liàngcíde yanjiu. Zhōnghuá shūjú. Pékin.
- Liú Shìshēng.- (1957) Tán liàngcí. Yǔyán xuéxí, n° 10.
- Wǎng Lì (Wang Liaoyi).- (1955) Yǒuguan rén wù hé xíngwéide xūci. Yǔ wén xuéxí cóngshū, Zhōngguó qīngnián chūbānshè, Pékin.
- Zhōng Bēnkāng.- (1959) Shìwù hé liàngcí. Yǔyán zhīshí, n° 8.

Numérales d'autres langues.

- Deloustal, R.- (1914) Des déterminatifs en annamite, BEFO, XIV, 5.
- Emenau, M.B.- (1951) Studies in Vietnamese Grammar. Univ. of Calif. Press, Berkeley.
- Haas, M.R.- (1942) The use of numeral classifiers in Thai, Language, XVIII, p 201 sq.
- Haguenhauer, Ch.- (1951) Morphologie du japonais moderne, Klincksieck, Paris.
- Hospitalier, J.J.- (1937) Grammaire laotienne, Imprimerie Nationale, Paris.
- Koutsoudas, A.- (1966) Writing Transformational Grammars. Mc Graw-Hill, New-York.
- Naganuma, N.- (1951) Basic Japanese Course, Tokyo.