

Constitution d'une base de données lexicales des langues sinitiques et contribution à l'étude de leur phylogénie

YIN Yuanhao (22100358)

INALCO, SDL (M2)

Parcours : Linguistique : langues, terrains, variations, typologie

- инaлco

Example 1

R. 李. Lǐ (1987) et Xióng & Zhāng (2012) :

- 1 Mandarin 官话
- 2 Jin 晋语 (séparé du mandarin)
- 3 Wu 吴语
- 4 Min 闽语
- 5 Hakka 客家话
- 6 Yue 粤语
- 7 Xiang 湘语
- 8 Gan 赣语
- 9 Hui 徽语
- 10 Pinghua et Patois 平话和土话

Exemple 2

X. 李. Lǐ (2012) :

- ① Mandarin 官话
- ② Wu 吴方言
- ③ Xiang 湘方言
- ④ Min 闽方言
- ⑤ Yue 粤方言
- ⑥ Gan 赣方言
- ⑦ Hakka 客家方言 (distinct du Gan pour des raisons culturelles)
- ⑧ Pinghua 平话
- ⑨ Hui 徽州话 (de nature mélangée et transitionnelle)

Exemple 2

X. 李. Lǐ (2012 : 511) :

Tonème du MC	A	B	C	D
Mandarin	[p ^h]			[p]
Yue	[p ^h]	[p ^h]/[p]		[p]
Xiang	[b]	(majorité)/[p]		[p ^h] (majorité)/[b]
Wu			[b]	
Hui (Jixi)		[p ^h] (majorité)/[p]		
Gan			[p ^h]	
Hakka			[p ^h]	
Pinghua			[p]	
Min		[p] (majorité)/[p ^h]		

Les initiale obstruantes sonores du MC dans les langues siniques (prenons l'initiale
[*b-] comme exemple) sans la stratification

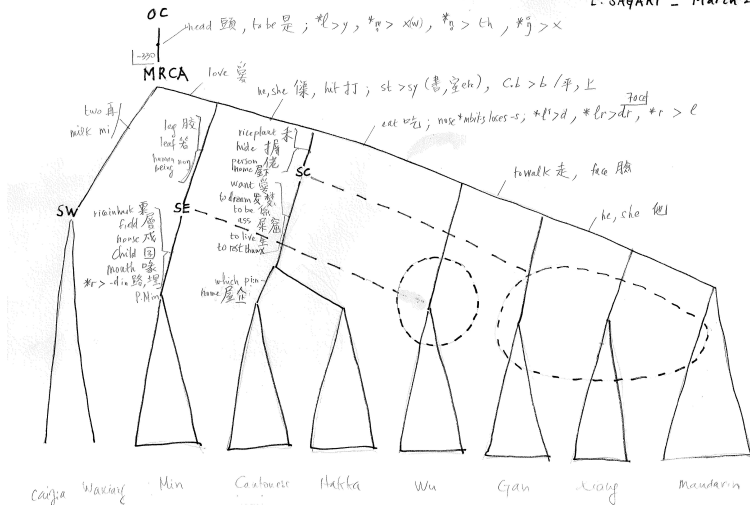
La classification des langues sinitiques

Les critères pour la phylogénie linguistique, une classification tant synchronique que diachronique :

- Lexical
- Innovation commune
- Stratification

Sagart (2011) :

L. SAGART - March 2011



Critères phonétiques de la stratification

Sagart & Xu (2001), Chén (2003) et Chén (2005) :

- Correspondances phonétiques systématiques entre les doublets/triplets/quartets... sur une syllabe dans différents mots (“wen-bai”)
- Connaissances de l’évolution phonologique du chinois (OC, MC, etc.)
- Critère (étendu) de cohérence

Example

“日” du Wu de Wenzhou :

- couche I : [ne], “三四 ~” (Yóu & Yáng 1998 : 158), “trois ou quatre jours”
- couche II : [niai], “生 ~” (Yóu & Yáng 1998 : 1471), “anniversaire” (couche héritée)
- couche III : [zai], “~ 记” (Pān 1998 : 54), “journal intime”

Situations exceptionnelles

- Cáo (2022) :
Les **mots grammaticaux** peuvent subir le changement phonétique exceptionnel (“fossilisation”, “lénition”, etc.).
e.g. “去” (“aller”, verbe directionnel), “渠” (prénom 3SG), “家” (suffixe), etc.
- Lǐ (1985) :
Les **toponymes** peuvent garder les prononciations archaïques.
e.g. “富”[p^hu] de “大富车” de la province du Hebei (Wáng 2017 : 36–37)

Une base de données déjà établie par Wu & List (2023) Traitement des données :

- Filtration des morphèmes non saillants, e.g. les suffixe “子”, “头”, etc.
- Annotation des couches : 1 v.s. 2/3 pour Min et 2 v.s. 1/3/4 pour non Min
- Correction de l'étymologie, e.g. “拨” > “昇” (“donner”) du Wu de Suzhou
- Concepts à exclure (mal traduit, trop de variation, étymologie non déterminée, etc.)
- COGIDs partielles (des syllabes) > COGIDs pleines (des mots)

Méthodes phylogénétiques

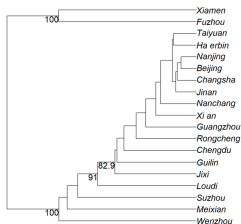
Hall (2018 : 74–75) et Pellard & Ryder & Jacques (2023 : 15–16) :

		Avantage	Inconvénient
Méthodes de diastances	UPGMA	rapide	présupposition non pertinente
	NJ		
	NN		
Méthodes de caractères	MP	“rule of thumb” populaire	“long branch attraction” chronophage
	ML		
	BI		

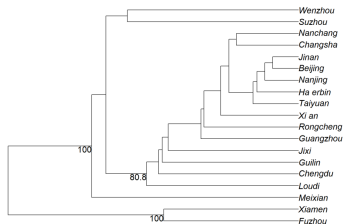
On applique les méthodes d’UPGMA, d’NJ et de MP aux données suivant le tutoriel de M. Pellard¹.

¹<https://tpellard.github.io/phylolinguistique/>

UPGMA



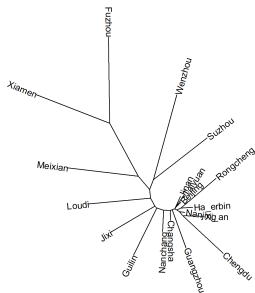
(a) Avec emprunts



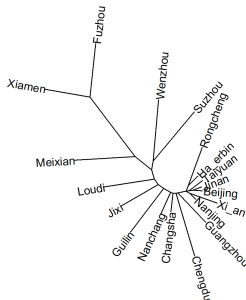
(b) Sans emprunts

FIG. : Arbres d'UPGMA testés avec la technique de BS

NJ



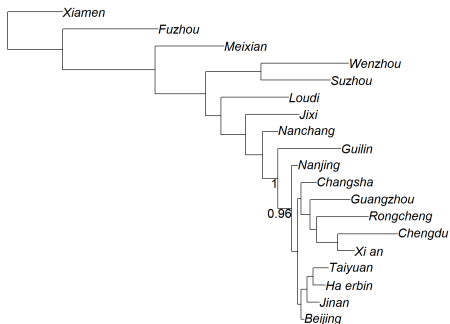
(a) Avec emprunts



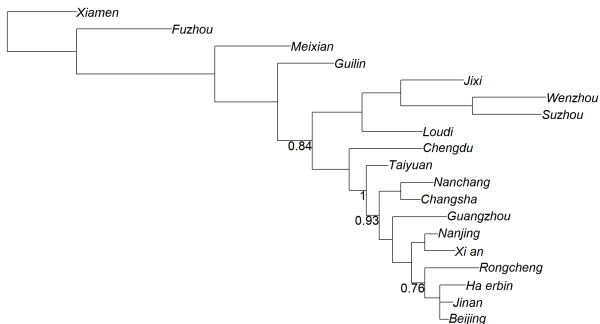
(b) Sans emprunts

FIG. : Arbres générés avec la méthode d'NJ **in al co**

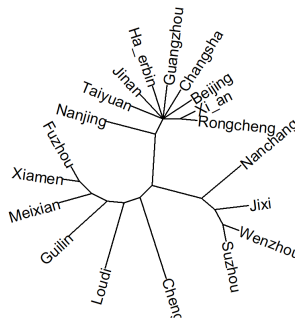
MP avec Ratchet



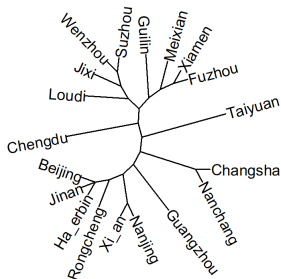
MP avec Ratchet



MP avec BaB



(a) Avec emprunts



(b) Sans emprunts

FIG. : Arbres consensus strict des arbres MPi n a l c o

Observation

- Les Min de Fuwhou et Xiamen constituent un clade.
- Les Wu de Suzhou et Wenzhou constituent un clade.
- Le Hakka se sépare après le Min et ne constituent jamais un clade avec le Gan de Nanchang.
- Les Xiang de Changsha et Loudi ne constituent jamais un clade.

Discussion

Les statuts anormaux du Yue de Guangzhou et du mandarin de Chengdu :
Deux types de biais résultant de l'écriture dans les enquêtes sur le terrain


	Avec emprunts	Sans emprunts
Beijing	76%	76%
Jinan	74%	75%
Guangzhou	70%	69%
Rongcheng	69%	68%
Changsha	62%	59%
Nanchang	62%	59%
Jixi	61%	61%
Loudi	58%	58%
Chengdu	56%	55%
Suzhou	54%	53%
Meixian	52%	51%
Wenzhou	50%	51%
Xiamen	40%	36%
Fuzhou	22%	32%


e.g. “给” v.s. “界” (“donner”) du Yue de Guangzhou


Discussion


- La nativisation étymologique des emprunts (Aikio 2007)
e.g. “太阳” v.s. “日头/热头” (“soleil”) du Yue de Guangzhou (Zhān 2002 : 389) et du Wu de Suzhou (Yè 1988 : 346)
- Situation opposée : les emprunts détectables partagent la même forme que les mots remplacés
e.g. “四” (*sijH) [sɿ] (“quatre”) v.s. “死” (*sijX) [si] (“mourir”) du Wu de Suzhou (Liú & Wáng & Bǎi 2007 : 96)

Bibliographie III



 Chén, Zhōngmǐn 陈忠敏. 2005. Lún yǔyīn céngcì de shíjiān xiānhòu 论语音层次的时间先后[Quelques notes sur les étapes chronologiques des strates sonores]. *Yǔyán yánjiū jíkān* 语言研究集刊[Bulletin of Linguistic Studies] (1). 123-132.

 Hall, Barry G. 2018. *Phylogenetic trees made easy*. Fifth Edition. New York : Oxford University Press.

 Lǐ, Róng 李荣 (éd.). 1987. *Zhōngguó yǔyán dìtú jí* 中国语言地图集[Atlas des langues de la Chine]. Hongkong : Longman.

 Lǐ, Rúlóng 李如龙. 1985. Dì míng zhōng de gǔyīn 地名中的古音[Les prononciations archaïques dans les toponymes]. *Yǔwén yánjiū* 语文研究[Linguistic Research] (1). 30-30.

Bibliographie V

-  Pān, Wùyún 潘悟云. 1998. Wēnzhōu huà yīndàng 温州话音档[Enregistrement du dialecte de Wenzhou]. In Hóu, Jīngyī 侯精一 (éd.), *Xiàndài hànyǔ fāngyán yīnkù* 现代汉语方言音库[Phonobase des dialectes chinois modernes. Shanghai : Shanghai Educational Publishing House.
-  Pellard, Thomas & Robin J. Ryder & Guillaume Jacques. 2023. The Family Tree model. In Ledgeway, Adam et al. (éds.), *The Wiley Blackwell companion to diachronic linguistics*. preprint 2023-05-11. https://www.academia.edu/101656989/The_Family_Tree_model.

Bibliographie VI

-  Sagart, Laurent. 2011. Classifying Chinese dialects/Sinitic languages on shared innovations.
Paper presented at the Séminaire Sino-Tibétain du CRLAO (2011-03-28). Downloadable at <https://www.academia.edu/19534510> (accessed February 4, 2018).
-  Sagart, Laurent & Shixuan Xu. 2001. History through loanwords : the loan correspondences between Hani and Chinese. *Cahiers de linguistique Asie orientale* 30(1). 3-54.
-  Shěn, Míng 沈明. 1994. *Tàiyuán fāngyán cídiǎn* 太原方言詞典 [Dictionnaire du dialecte de Taiyuan]. Nanjing : Jiangsu Education Publishing House.

Bibliographie VII

-  Wáng, Jiājiā 王佳佳. 2017. *Héběi yì xiàn dì míng de yǔ yán xué kǎo chá* 河北易县地名的语言学考察[*Une étude linguistique des toponymes dans le district de Yi, province du Hebei*]. Bohai University. ([mém. de mast.]).
-  Wu, Mei-Shin & Johann-Mattis List. 2023. Annotating cognates in phylogenetic studies of Southeast Asian languages. *Language Dynamics and Change* 1(aop). 1-37.
-  Xióng, Zhèng huī 熊正辉 & Zhèn xīng 张振兴 Zhāng (éds.). 2012. *Zhōngguó yǔ yán dì tú jí (Dì èr bǎn) : Hànyǔ fāng yán juǎn* 中国语言地图集 (第二版): 汉语方言卷[*Atlas des langues de la Chine (2^{ème} version) : Volume des dialectes du chinois*]. Beijing : Commercial Press.

