

## **OUTILS INFORMATIQUES POUR LA LINGUISTIQUE HISTORIQUE BANTU**

Jean-Marie HOMBERT, Médard MOUELE et Lai-Won SEO

*Three computer tools useful for doing comparative linguistics are presented. They have been developed on MacIntosh and are applied to Bantu languages. They deal with the presentation of a linguistics atlas (including audio data), the organization of a proto-bantu lexicon and testing of diachronic rules.*

### **1. INTRODUCTION**

L'étude des changements phonétiques fait appel à deux types de données : données réellement diachroniques pour les langues (ou groupes de langues ) pour lesquelles nous disposons de documents écrits couvrant une période historique suffisamment importante et des données de type dialectologique qui nous permettent de faire des hypothèses diachroniques à partir d'une coupe synchronique.

Dans les deux cas le comparatiste est amené à manipuler (collecter, dupliquer, trier) une quantité souvent impressionnante de données avec les contraintes inhérentes à ce genre d'exercice: perte de temps et dégradation de la fiabilité au cours des manipulations successives. Il est évident que l'informatique peut trouver ici un champ d'application idéal puisque l'essentiel du travail consiste à faire des tâches répétitives sur des volumes importants de données.

L'utilisation de l'informatique permettra donc de "gagner du temps"(au moins si l'on omet de comptabiliser l'investissement initial nécessaire à la saisie informatique des données et à la mise au point du logiciel de traitement de ces données) mais surtout elle amènera une plus grande rigueur et une plus grande fiabilité. Certains comparatistes ont très tôt pris conscience de ces avantages puisque dès le début des années soixante-

époque préhistorique des cartes perforées- des comparatistes américains (Naim,1962;Swadesh,1963) proposaient un système automatisé de comparaison. Depuis cette époque ,un grand nombre de projets comparatistes ont été développés (cf. bibliographie), très différents les uns des autres quant à leurs ambitions (allant de la simple mise en forme des données à la production de formes reconstruites), aux zones linguistiques concernées (amérindienne, indo-européenne, sino-tibétaine, nilo-saharienne et niger-congo)et à l'état d'avancement des travaux(variant de simples propositions méthodologiques à la réalisation de logiciels (quasi) commercialisables

Dans notre laboratoire nous nous intéressons à cette problématique depuis 1986.; d'abord sous forme d'une réflexion méthodologique sur un petit groupe de langue bantu (Bancel, 1986) puis par la mise en oeuvre d' une approche de type système expert(Bancel et al,1988) avec une première implémentation de la phase préliminaire de mise en forme des données sur Micromega32(programmation en C). Nous nous sommes également intéressés aux problèmes posés par le codage informatique des segments (Bancel, 1988 ; Puech, 1988); l'utilisation d'un codage adapté étant évidemment un élément essentiel permettant la prise en compte de classes naturelles lors de l'étape de formalisation des règles diachroniques. Plus récemment les spécificités des matériels compatibles PC et Macintosh (Coûts,rapidité,convivialité) nous ont amenés à poursuivre des développements en parallèle sur ces deux types de matériel; un premier projet,sous la responsabilité de J.M. Hombert, tente d'exploiter au profit du comparatiste les avantages du monde Macintosh (ergonomie,applications multimédia.);l'autre projet sous la responsabilité de G. Puech est développé sur compatible PC avec le SGBBD Oracle. Une interface écrite en langage C est développée de manière à rendre transparente par un linguiste la saisie et l'interrogation des données sans obliger l'utilisateur à l'apprentissage du langage SQL qui tout en étant simple de conception, se révèle délicat à manier pour les requêtes complexes.

Le second aspect du projet porte sur la formalisation de règles de reconstruction à partir d'une exploitation systématique des termes "comparables" contenus dans la base de données

. Les réalisations présentées dans cet article ne concernent que les travaux effectués dans le cadre du premier projet; une présentation du second projet fera l'objet d'un article dans le prochain numéro de cette revue

## **2. PRESENTATION DES LOGICIELS**

### ***2.1. Atlas linguistique parlant***

Les atlas linguistiques constituent une source de données extrêmement importante sur les changements phonétiques. Malheureusement sous leur forme traditionnelle, ils sont difficilement utilisables (difficulté de manipuler ces ouvrages volumineux et de produire de nouvelles cartes synthétiques). Nous avons développé un logiciel permettant un traitement informatique des données lexicales collectées dans le cadre du projet ALGAB (Atlas Linguistique du Gabon) ; pour cela nous avons utilisé quatre logiciels disponibles commercialement : Carto 2D, FileMaker, HyperCard et Signalyze. Carto 2D permet de digitaliser les fonds de cartes nécessaires ; les données lexicales sont structurées grâce à FileMaker, la présentation des cartes linguistiques se fait sous HyperCard et enfin des mesures acoustiques peuvent être effectuées à l'aide du logiciel Signalyze.


Une liste de mots cartographiés est présentée à l'utilisateur (voir Carte 1). A partir de cette liste, il peut sélectionner la carte de son choix. Apparaît alors à l'écran l'ensemble de la zone linguistique étudiée avec la localisation des points d'enquête et pour chacun de ces points, la transcription phonétique du mot cartographié (voir Carte 2). En cliquant sur l'une quelconque de ces transcriptions, on peut alors entendre la forme sonore correspondante (préalablement digitalisée). Si une information phonétique plus précise est souhaitée, l'utilisateur peut alors faire appel au logiciel Signalyze (voir Bouton en bas, à droite de Carte 2) afin d'obtenir des mesures de fréquence ou de durée (voir Carte 3).

Ce logiciel permet au comparatiste de déterminer avec précision la distribution géographique d'un changement phonétique donné, d'identifier les éventuels contre-exemples et surtout de pouvoir accéder — et comparer — avec une grande rapidité aux formes sonores correspondant aux différents points d'enquête.

## 2.2. *Lexique Proto-Bantu (PBLEX)*

Les reconstructions du Proto Bantu proposées par Guthrie sont classées en fonction de la consonne initiale. Il est par conséquent difficile d'accéder à des formes lexicales proto bantu lorsque le critère de tri ne concerne pas cette position initiale. A l'aide du logiciel FileMaker, nous avons saisi l'ensemble des formes proto bantu reconstruites par Guthrie en décomposant l'information contenue dans chacune des entrées de Guthrie en 23 rubriques : "Numéro", "mot", "C1", "V11", "V12", "C2", "V21", "V22", "C3", "V3", "C4", "V4", "Ton11", "Ton12", "Ton2", "Ton3"; "Ton4", "Classe sg", "Classe pl", "appariement", "Traduction française", "Traduction anglaise" et "Commentaire".

Un exemple de présentation d'une fiche type est donné ci-dessous.

Numéro : 2164	 Nouvelle fiche																														
Mot : yoma																															
	<table border="0"> <tr> <td>e1</td><td>v11</td><td>v12</td><td>e2</td><td>v21</td><td>v22</td><td>e3</td><td>v3</td><td>e4</td><td>v4</td> </tr> <tr> <td>y</td><td>o</td><td></td><td>m</td><td>a</td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td colspan="2">Tons:</td> <td>H</td><td></td><td>B</td><td></td><td></td><td></td><td></td><td></td> </tr> </table>	e1	v11	v12	e2	v21	v22	e3	v3	e4	v4	y	o		m	a						Tons:		H		B					
e1	v11	v12	e2	v21	v22	e3	v3	e4	v4																						
y	o		m	a																											
Tons:		H		B																											
N.Classé Singulier : 7	7/8																														
Pluriel : 8																															
Traduction en français :																															
Traduction : place																															
Commentaire : NW,CW																															

Cette division en rubriques nous permet de faire des requêtes combinant des informations contenus dans l'une quelconque de ces rubriques. Ainsi, la fiche présentée résulte d'une interrogation de PBLEX des entrées contenant la voyelle a en position V11, la voyelle o en position V21, ayant le schème tonal HH et appartenant au genre 1/2. Il s'agit ici d'une application qui n'a impliqué aucune programmation complémentaire ; les facilités offertes par FileMaker étant suffisantes pour atteindre l'objectif recherché.

### ***3. Reconstruction Assistée par Ordinateur (RAO)***

Il est important pour un comparatiste de pouvoir tester de **manière systématique** les règles d'évolutions phonétiques qu'il propose. Le logiciel RAO, réalisé avec HyperCard permet de tester les règles de passages du proto bantu vers une langue bantu actuelle. L'utilisateur est d'abord confronté à un écran présentant des formes proto bantu (avec leur traduction) et des formes attestées dans la langue bantu moderne choisie (voir Carte 1). Il peut créer des règles lui permettant de passer de l'une à l'autre. A titre d'exemple, la Carte 2 illustre la création de la règle : consonne initiale p devenant consonne initiale f. Cette règle est alors appliquée à l'ensemble du corpus et ses effets apparaissent dans une nouvelle colonne (voir T Fang dans Carte 3). La comparaison des données provenant de l'application des règles (colonne T Fang) et des données réelles permet au comparatiste de proposer et de tester de nouvelles règles afin d'obtenir des formes aussi semblables que possible dans ces deux colonnes.

Ce logiciel permet donc de tester à la fois la régularité des changements phonétiques mais aussi l'ordre d'application des règles.

Le choix du logiciel HyperCard a permis une réalisation rapide de cette application ; en contrepartie, sa faible vitesse d'exécution limite son utilisation à des corpus de faible volume.

**BIBLIOGRAPHIE**

- BANCEL P.(1986) *Etude comparée des langues du groupe fang pour la création d'une base de données bantu*; Mémoire de maîtrise, Université Lumière Lyon2, 65p , annexes
- BANCEL P.(1988) "A bon A.P.I. bon C.P.I.(pour un codage phonétique vraiment international)" *Pholia* 3, p 27--36
- BANCEL P., P. DUPONT, J.M. HOMBERT, A.M. MORTIER et G. PUECH (1988) "An expert system for comparative linguistics: Application to bantu languages" Communication présentée à la 19ème African Linguistics Conference, Boston University, 14-17 Avril
- EASTLACK C.L.(1977)"Iberochange: A program to simulate systematic sound change in Ibero-romance" *Computers and the humanities* 11 p81-88
- GRFENSTETTE G.(1987) "An expert system for philologists" Actes des 7ème journées internationales sur les systèmes experts et leurs applications, Avignon, 13-15 Mai, p 1251-1269,,
- HEWSON J.(1974)"Comparative reconstruction on the computer" in *Historical Linguistics. Proceedings of the First International Conference on Historical Linguistics*, Anderson J.M. and C. Jones (eds), North Holland, Amsterdam, p 191-197.
- HOMBERT J.M.(1988) "Mammals in the languages of Gabon: a step towards the reconstruction of proto-bantu fauna", Communication présentée à la 19ème African Linguistics Conference, Boston University, 14-17 Avril
- HEWSON J.(1977) "Reconstructing prehistoric languages on the computer: the triumph of the electronic neogrammarian" *Proceedings of the international conference on computational linguistics*, Pise, 27 Aout-1 Septembre 1973, p 263-273
- JOHNSON M. (1985)"Computer aids for comparative dictionaries" *Linguistics* 23 p 285-302.
- LOWE J.B. et M. MAZAUDON (1989)"Computerized tools for reconstruction for Tibeto-Burman" *Proceedings of the annual Meeting of the Berkeley Linguistics Society*, 15, p 367-378.
- MANIET A.(1980) "Recherche par ordinateur sur la phonologie diachronique du latin" Communication présentée au 16ème International Congress of Romance Linguistics, Palma de Mallorca.
- MAZAUDON M. et J.B. LOWE (à paraître) "Du bon usage de l'informatique en linguistique historique", *Bulletin de la Société Linguistique de Paris*

- NAIM C.M.(1962)"A program for partial automation of comparative reconstruction"  
Anthropological Linguistics,4 (9), p1-10.
- NICOLAI R. (1989) "Diffusion lexicale en zone sahel-saharienne" Projet de recherche  
présenté au 4th Nilo-Saharan Colloquium, Bayreuth.
- PUECH G.(1983) "Codage de l'Alphabet Phonétique International", Pholia 3, p 227-246
- PSHENICHNOVA N.N. et G.A.CHERKASOVA (1986) "An automated data base for  
dialectology and the history of Russian" Communication présentée au  
Symposium on Formalization in Historical Linguistics, Tallinn, 24-26 Novembre.
- REMMEL M.(1980) "Computers in the historical phonetics and phonology of Balto-  
Finnic languages: problems and perspectives" Communication présentée au 5th  
International Finno-Ugric Congress, Turku.
- SWADESH M (1963) "A punchcard system of cognate hunting" International Journal  
of American Linguistics 29 p 282-288
- WANG W.S.Y.(1970)"Project DOC: its methodological basis" Journal of the Oriental  
Society 90,1, p 57-66

# ATLAS Linguistique FANG

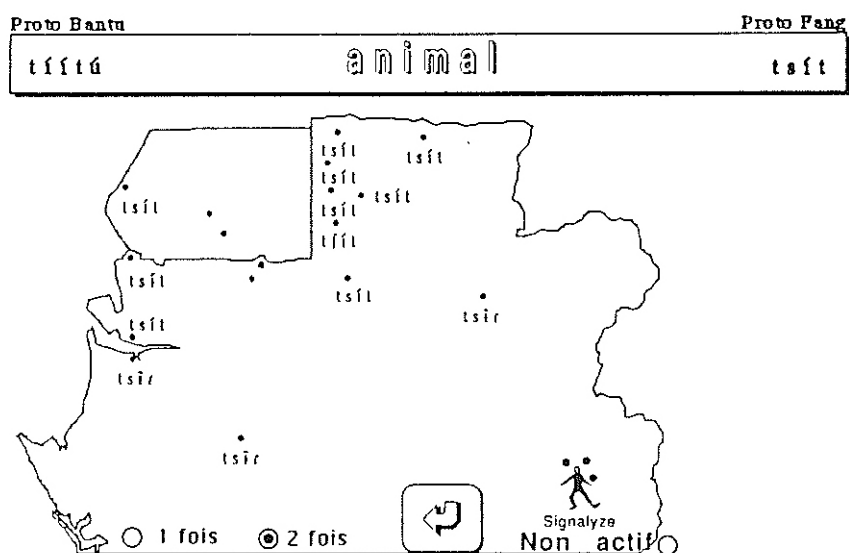


animal	↑
barbe	
buffle	
chute	
cicatrice	
cuisse	
être cuit	
froid	
noix de cola	↓

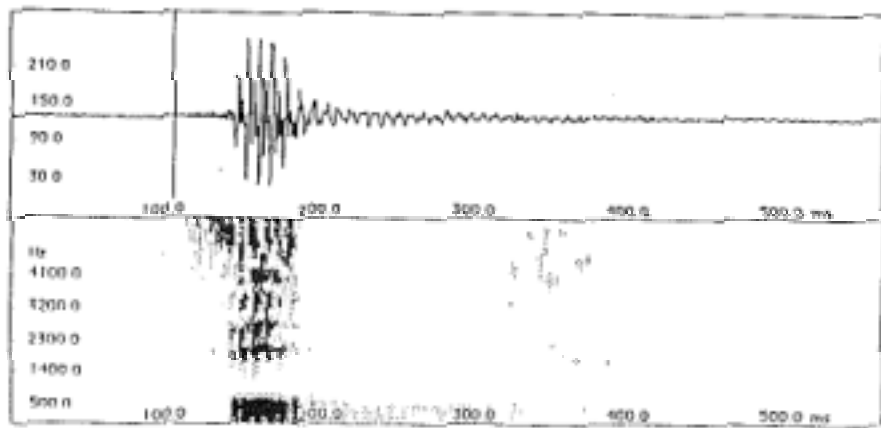
Quit HyperCard

Carte 1. Liste de mots cartographiés





Carte 2. Présentation de la carte linguistique du mot « animal ».

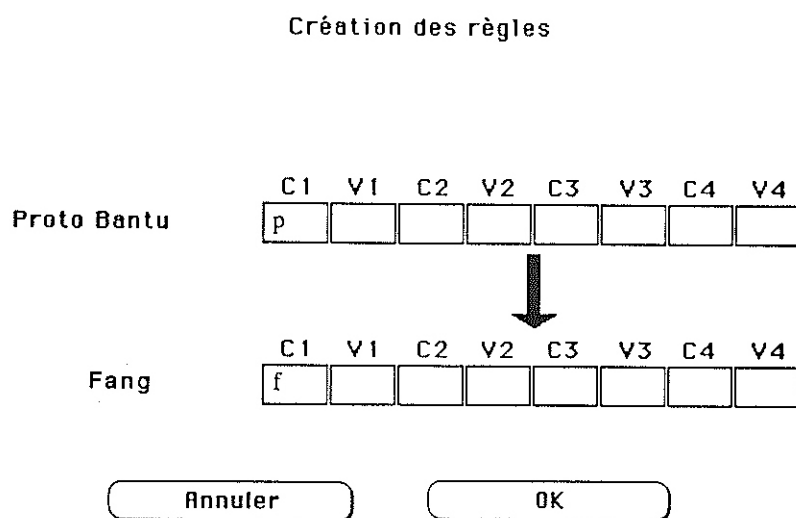


Carte 3. Représentation du signal et du sonagramme du mot « animal » : tsít (parler de Medouneu)

Compilation
Regles
Ranger

Traduc.	P Bantu	Règle	T Fang	Fang	
aille	pàpá	*	*	fāp	↑
froid	pépè	*	*	vóp	□
tique	kópá	*	*	kóp	
animal	tító	*	*	tsít	
bosse	túútú	*	*	tút	
lourd	dítò	*	*	dzīt	
plaie	pótá	*	*	fóó	
rapide	pót	*	*	vóó	
trois	tátò	*	*	láà	↓

Carte 4. Corpus sans règle.



Carte 5. Illustration de la création des règles.

Ranger

Compilation      Regles      Compilation partielle

Traduc.	P Bantu	Règle	T Fang	Fang
aille	pàpá	f1	fapa	fãp
froid	pépè	f1	fepo	vóp
tique	kópá	*	*	kóp
animal	tító	*	*	tsít
bosse	túútú	*	*	tút
lourd	dîtò	*	*	dzīt
plaie	pótá	f1	fota	fóó
rapide	pót	f1	fot	vóó
trois	tátò	*	*	láà

○ 1

Carte 6. Corpus avec application de la règle f1.