

**GRAMMAIRE COMPARÉE DES LANGUES
AUSTRONÉSIENNES: QUE FAUT-IL PENSER
DU «GRAND TOURNANT» DE L'APRÈS-GUERRE?¹
BILAN, NOUVELLES HYPOTHÈSES ET PROGRAMMES**

Abstract

When a linguist trained in the comparative grammar of classical languages approaches the Austronesian languages, he may well be surprised. The language type is different: here are agglutinative languages only recently documented, some even unknown a century ago, with a phonology that, for most of them, seems to have been so little evolving for 6000 years that the reconstructed etymons are, more often than not, nearly identical with their reflexes.

*The methods in use are quite different too – which is still more surprising. Since WW2, in the wake of the works by Isidore Dyen and others, lexicostatistics and associated disciplines (glottochronology, migration theory), surveys and general anthropology as well have been in possession of the field, and, if we look more closely, even historical phonetics is based on a new seemingly rigorous conception of the protophonemes and their reconstruction. We'll offer a critical appreciation of it. The theories of Swadesh have been from the start much criticized but, strongly advocated by Dyen, they are still to be found everywhere. If the irrational multiplication of certain classes of protophonemes has been discarded, there remains however the idea that a reconstructed element is nothing more than the class of equivalence between its reflexes. Lastly, we'll show, from the observation of the article **(k)u*, that the migration theory imposes an unjustified filter to reconstruction hypotheses.*

We'll deal with some odd, and even intractable, data and we'll propose a few suggestions about historical phonetics (we've already dealt elsewhere with grammatical markers and the systems they form – the first step in any comparative grammar) and about the protolexicon, in which borrowing – the worm in the lexicostatistic fruit – according to us plays a far more important part than usually realized. We'll pay attention to the problem raised by the numberless lexical doublets, often reconstructed from two or three reflexes, the significans of which varies for the same

1. Cet article reprend la communication faite à la journée d'étude du 28 janvier 2006 pour laquelle les participants avaient à leur disposition la totalité des tableaux, carte, listes d'étymons utilisés ici.

Nous tenons à remercier ici Alexandre François qui a bien voulu nous communiquer voici quelques années la version électronique 2001 du *Dictionnaire étymologique* de R. A. Blust, mise en forme par ses soins; sans elle, rien du présent travail n'aurait été possible.

Nous emploierons pour les marques de voix ou «focus» les abréviations suivantes: AF pour «actor focus», PF pour «patient» ou «goal focus», RF pour «referent focus» et LF pour «locative focus», IF pour «instrumental focus», BF pour «benefactive focus»; pour les abréviations de noms de langue, voir la liste à la fin du texte.

*significatum only by the initial syllable. We'll propose to see, in this syllable the trace of older affixes (prefixes of nouns of action, 'focus' marker, etc.) integrated into bases and turned opaque. Even the *(N) in the many *C₁V₁(N)C₂V₂C₃ etymons reconstructed since Dempwolff should be, according to us, considered as the trace of an older infix *-aN- of 'plural action' occasionally standing in opposition to a 'specific' *-aR-.*

Introduction

Quand on aborde l'étude comparative des langues austronésiennes et que l'on a suivi une formation plus ou moins complète de grammaire comparée d'autres familles de langues comme les langues indo-européennes, ou de branches d'autres familles comme les langues romanes ou les langues bantoues, on est surpris par 1) un certain nombre de caractéristiques propres aux langues austronésiennes et au proto-austronésien tel qu'on le reconstitue, 2) par les principes suivis pour ces reconstructions.

On découvre qu'on ne fait pas du tout la même chose pour les langues austronésiennes que ce qu'on fait pour les autres familles ou sous-familles citées, et qu'on y admet souvent ce qu'on rejette ailleurs, ou bien encore qu'on place au centre des préoccupations ce qui reste tout à fait périphérique ailleurs; on y retrouve aussi des sujets devenus pratiquement tabous, mais anciens, comme le rapport entre géographie, archéologie, anthropologie et grammaire comparée-reconstruction. Certes, avec le développement de certaines macro-comparaisons, on sera peut-être moins surpris aujourd'hui qu'il y a une vingtaine d'années: en effet, des pratiques comme la glottochronologie qui est souvent le seul fondement de certaines «grandes classifications» ou de certaines macro-comparaisons, est une des origines du renversement qu'a connu la grammaire comparée des langues austronésiennes au lendemain de la deuxième guerre mondiale; l'intervention des comparatistes américains formés plus ou moins dans la ligne des américanistes n'est pas sans lien non plus avec la rupture épistémologique par rapport aux méthodes héritées des néo-grammairiens allemands utilisées jusque là.

1. De quelques particularités de la famille austronésienne et du proto-austronésien

Mais la première source d'étonnement pour le linguiste est peut-être la différence même de situation entre familles indo-européenne ou chamito-sémitique et famille austronésienne.

1.1. *Etat des données*

Cette dernière famille, qui a une grande extension géographique, de Taiwan au nord à la Nouvelle Zélande au sud, de Madagascar à l'ouest à

l'Ile de Pâques à l'est, comprend un très grand nombre de langues: 400 selon Dyen dans ses écrits des années 50, plus de 1100 dans les travaux actuels; sauf les langues de la péninsule malaise et de l'Asie du sud-est, elles sont parlées dans des îles. La connaissance qu'on a de ces langues repose sur des matériaux de qualité très diverse: pour une large majorité d'entre elles, on n'a disposé longtemps que de listes de mots — ce qui correspond donc au premier stade de «grammatisation» des langues, pour reprendre le terme de Sylvain Auroux². Cette situation globale n'a pas été sans influence sur les méthodes de reconstruction employées en grammaire comparée des langues austronésiennes et, en particulier, sur l'importance accordée à la comparaison lexicale, qui donne l'illusion qu'on peut se contenter de listes de vocabulaire; cela n'est donc pas sans lien non plus avec le succès de la lexicostatistique et de la glottochronologie.

Par ailleurs, sauf quelques exceptions pour lesquelles on a des attestations d'environ 1000 ans (vieux malais, vieux javanais, vieux bugis, etc.), les matériaux ne remontent guère au-delà du début du 19^{ème} siècle; dans d'autres cas qui pourraient paraître assez favorables (comme celui de certaines langues des Philippines), pour lesquels on a un recul de quatre ou cinq siècles, il s'agit malheureusement de langues ayant relativement peu évolué sur ce laps de temps. On n'en sera pas moins surpris du manque total d'intérêt manifesté par les comparatistes des langues austronésiennes du dernier demi-siècle à l'égard des états de langue les plus anciens et les plus divergents par rapport aux états actuels (vieux javanais, vieux bugis, vieux malais, etc.), au point qu'état actuel d'une langue et état ancien sont mis sur le même plan comme deux langues distinctes (voir le vieux javanais et le javanais moderne chez R. A. Blust³).

Par ailleurs, on a pu soutenir que l'absence de recul chronologique est une contre-indication majeure à toute tentative de grammaire comparée et de reconstruction. Nous ne le pensons pas: hormis, sans doute, l'antiquité des attestations védiques et homériques, le fait que les premières attestations de l'«altindischen, alteranischen, altgriechischen, altitalischen, altkeltischen, altslawischen, litauischen und altdeutschen» (la liste de langues figurant dans le titre de Schleicher 1861 parallèle à celle de Bopp) s'étalent sur plus d'un millénaire et demi n'a pas une réelle importance et elles ont été en fait traitées à peu près comme si elles étaient contemporaines; on peut dire que seule la découverte du hittite et de son ancienneté a pesé d'un réel poids sur le développement des études comparatives. Sinon, cette contemporanéité fictive a été prise comme point de départ de la reconstruction rétrospective: on a fait «comme si» les données étaient alignées et, pour une part, comme s'il

2. *La révolution technologique de la grammatisation*, Mardaga, 1994, plus particulièrement, les chap. 3 et 4.

3. Peut-être s'agit-il, dans la pensée de cet auteur, d'indices sur des faits hérités occultés en vieux-javanais et resurgis en javanais moderne, ou bien de traces d'emprunts du javanais à d'autres langues.

n'y avait pas d'après⁴. Le fait que l'on n'ait pas de recul chronologique pour certaines familles de langues ne constitue pas à notre avis un véritable handicap⁵ au moins jusqu'à un certain point.

La situation même des études a aussi de quoi surprendre. En fait, le regroupement des langues en sous-branches, qui, pour les langues indo-européennes, est en réalité déjà implicitement acquis dès Bopp, albanais et arménien mis à part, est loin d'être assuré pour les langues austronésiennes: seule est acquise la distinction entre les trois groupes, non pas mis plus ou moins sur le même plan, mais décalés en diachronie, constitués par les langues de Formose, les langues malayo-polynésiennes, les langues océaniques; la proximité géographique et la lexicostatistique sont le plus souvent la seule justification avouée ou non des classifications proposées à l'intérieur du groupe malayo-polynésien⁶, c'est-à-dire l'ensemble de langues qui fournit la base même de la reconstruction.

1.2. *Un protosystème phonologique étrangement proche des systèmes phonologiques des langues actuelles*

La surprise que peut éprouver le linguiste venu d'ailleurs quand il aborde le domaine de la linguistique comparative des langues austronésiennes est peut-être due en premier lieu au faciès même des langues de cette famille. Bien que l'on avance actuellement une date de 6000 ans avant le présent pour le proto-austronésien⁷, on est surpris de voir la très faible évolution aussi bien entre les protophonèmes et leurs reflets qu'entre les étymons et les reflets de ces étymons, c'est-à-dire entre le proto-austronésien et la majorité des langues qui servent précisément à l'établir.

4. Sauf pour un type de raisonnement/explication au moyen du concept de «tendance», mais on peut précisément mettre en doute que l'«idée de grandes tendances millénaires» de tel ou tel groupe puisse constituer plus qu'un simple constat, sinon une simple piste, et être érigée en causalité et en explication.

5. Il n'en va pas de même pour les difficultés signalées par Meillet (*Introduction...*, p. 479) dues au type de langues, entre autres pour les familles à morphologie restreinte ou trop simple; il conviendra donc de se demander si le caractère agglutinant de langues comme celles de la famille austronésienne est un handicap. On pourrait par exemple arguer de ce caractère agglutinant pour «expliquer» le peu d'évolution entre protolangue et langues actuelles, c'est-à-dire sur 6000 ans, si les propositions de datation de la protolangue qui ont cours aujourd'hui se trouvent confirmées.

6. Cf., par exemple, le jugement de Tryon (1995, I, p. 21): «The lower order subgroups (...) are, with a couple of notable exceptions, still problematic at all but the lowest levels (...) The higher order subgrouping of the Philippines languages is still the subject of ongoing debate (...) with important implications for the higher level subgrouping of Austronesian languages. At the same time, Ruhlen's Sulawasi, Borneo and Sundic subgroups of Western Malayo-Polynesian are convenient geographic groupings only, rather than lower order subgroups based on substantial linguistic evidence».

7. On émettra toutes les réserves qui s'imposent sur cet âge adamique que l'on retrouve à chaque fois qu'il est question de protolangues de familles à attestations orales récentes, datées à coups de glottochronologie.

Les seules ruptures que l'on constate dans un tableau comme le suivant, (emprunté à Ross, in Tryon 1995, I, p. 71) se situent entre le proto-austro-nésien et le proto-malayopolynésien (c'est-à-dire compte tenu ou non des langues de Formose) et entre le proto-malayopolynésien et le proto-océanien:

PAN	PMP	tagalog	javanais	toba barak	malais	POC
p	p	p	p	p	p	p,b
b	b	b	b,w- -b,w- -b	b- -b- -p	b- -w- -p	p,b
m	m	m	m	m	m	m
w	w	w	w	∅	h- -w-	w
t	> t	t	t	t	t	t
C		s	s	s	s	s,j
s		Z	d- -d-	d- -d-	J- -J-	s,j
d ₁	> d	d- -l-	d,r- -d,r- -d	d- -d-	d- -d-	r,dr
d ₂		d- -r- -d	d,r- -d,r- -d	d- -d- -t	d- -d- -t	r,dr
d ₃		(z)	J- -J,r-	J- -g-	J- -J-	r,dr
	(D)	d- -l- -l	d,r- -d,r- -d	d- -d- -r	d- -d- -r	r,dr
	(T)	t	t	t	t	t
r	r	l	r	r	r	r,dr
n	> n	n	n	n	n	n
L		l	l	l	l	l
l		(c)	c	s	c	s,j
...	h	s	∅	∅	∅	∅
S	j	-l- -d	r	-g- -k	-d- -t	c,j
j	ñ	n	ñ	n	ñ	ñ
...	y	y	y	∅	y	y
y	k	k	k- -k- -?	h	k- -k- -?	k,g
k	g	g	g	g- -g- -k	g- -g- -?	k,g
g	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
ŋ	q	?	h	∅	h	q
q	> -?- -H	h	∅	∅	∅	∅
h		g	∅	r	r	R
?		i	i,e	i	i,e,ə	i
R	u	u,o	u,o	u	u,o,ə	u
i	a	a	a	a	a,ə	a
u	e	i,a,u	e		ə,a	o
a	-aw	-aw	-o	-o	-aw	-o
e	-ay	-ay	-e	-e	-ay	-e
-aw	-uy	-uy	-i	-i	-i	-i
-ay						
-uy						

Les différences entre PAN et PMP sont dues à:

- 1) des confusions d'oppositions reconstruites à partir des langues de Formose ou d'une partie d'entre elles⁸:

PAN $*t + *C > \text{PMP } *t$
 PAN $*d_1 + *d_2 + *d_3 > \text{PMP } *d$
 PAN $*n + *L > \text{PMP } *n$
 PAN $*h + *? > \text{PMP } *h$

- 2) des protophonèmes PMP sans antécédent:

$(*z), (*D), (*T), (*c), *n$

et à la découverte d'un nouveau protophonème $*S^9$, représenté par une sifflante ou une chuintante dans une partie des langues de Formose, par /h/ dans une partie des langues MP et par $\emptyset$ dans les autres, protophonème dont l'existence constitue un des arguments forts dans l'établissement d'un proto-austronésien constitué d'une part des langues de Formose, considérées comme un sous-groupement unitaire ou comme un ensemble de sous-groupes tôt séparé, et d'autre part, des langues MP, là où pour Brandstetter le «formosan» n'était qu'une langue à côté des autres.

Les différences restent donc faibles et portent sur peu de phonèmes, dont certains comme $*q, *h$ et $*?$ peuvent être fragiles¹⁰, surtout si l'on prend en considération le matériel lexical concret sur lequel ils sont fondés où abondent les doublets avec ou sans C_3 , en $*h$ vs $*?$ et le faible nombre de reflets permettant effectivement de choisir. La situation est simple: les confusions de protophonèmes quand on passe de PAN à PMP semblent solidement établies vu la méthode rigoureuse employée, mais les différenciations de protophonèmes entre ces deux protolangues ne reçoivent pas plus d'explication que n'en recevraient celles ($*h > *h + *S, *n > *l + *n, *t > *t + *C$ et $*d > *d_1 + *d_2 + *d_3$) qu'on serait amené à poser si on considérait que les langues de Formose n'étaient qu'un sous-groupe divergent du PMP ou d'une

8. On relèvera dès maintenant le principe de méthode, contestable, consistant à ajouter en PAN un phonème symbolisé par $*C$ tout en gardant $*t$ à la fois pour le PAN et le PMP alors que, dans un étymon limité au PMP, un $*t$ peut aussi bien être issu d'un $*C$ PAN que d'un $*t$ PAN; ainsi, les effectifs des étymons en $*t$ se trouvent renforcés, ce qui en soi n'est pas grave tant qu'on n'oublie pas qu'il s'agit d'une protolangue distincte, mais dans un dictionnaire étymologique conçu de manière «descendante» comme celui de Blust et à plus forte dans des listes onomasiologiques comme celles de Tryon (1995, tomes III à V), tout se passe comme si $*C$ était une spécification secondaire de $*t$ et dans la chronologie comme si $*C$ donnait $*t$. Les symboles uniques de l'ordre alphabétique des entrées recouvrent des réalités distinctes du point de vue phonétique et phonologique et du point de vue historique.

9. Cf. plus loin, parag. III 2, pour une présentation plus détaillée de l'histoire de la découverte de ce nouveau protophonème.

10. Entre autres, ils peuvent correspondre à un stade intermédiaire entre une consonne et sa disparition (cf. $*k > /h/ > \emptyset$ en malgache, aussi bien que $*s > /h/$ en grec ancien, et, de là, $> \emptyset$, dans les dialectes à psilose).

branche philippine du PMP comme le suggérait la façon de faire intervenir le «formosan» dans les rapprochements de Brandstetter.

1.3. Des reflets souvent presque identiques à leur étymon

Une des conséquences de la faible évolution que nous avons relevée entre protosystème phonologique PAN et systèmes phonologiques attestés est que, comme on peut s'y attendre, pour un grand nombre de langues, la majorité des reflets d'un étymon est pratiquement identique à l'étymon lui-même.

Prenons l'exemple d'un étymon non analysable en l'état actuel de nos connaissances: le nom de l'«étoile», PAN **bituqen*, très largement attesté à travers toute la famille, des langues de Formose aux langues océaniques en passant par les langues des Philippines et des îles de la Sonde — nous reproduisons ici tel quel l'article du dictionnaire étymologique de R. A. Blust dans sa version électronique de 2001¹¹:

(1) <i>*/bituqen/</i> star		
F:	SAI	/bint<oe>?<ae>n/ star
	PAZ	/bintun/ star
	PUY	/vituHen/ star
WMP:	ITB	/vitoen/, /vitwen/ star
	DGT	/biton/ star; for the stars to come out after a rain shower
	ISG	/bittuwa'n/ star
	ITW	/bi'ttwan/ star
	DGT	/bito'n/ star; for the stars to come out after a rain shower
	IFG	/bitu'won/ star; also planet since the Ifugaw do not distinguish between the two... All the stars are believed to be the daughters of the moon deity
	IFGBt	/bittu_an/ be starry, of the sky; a star
	ILK	/bitue'n/ star
	"	/bitue'n bayba'y/ starfish
	PGS	/bite'wen/ star
	TAG	/bitu'?in/ star
	BKL	/bitu'?on/ star in the sky)
	AKL	/bitu'?on/ star (general term)
	HLG	/bitu'?un/ star
	TBW(K)	/bitukun/ star
	CEB	/bitu'?un/ star; kind of tree of strand, the star-shaped fruits of which are used as fish poison: <i>Barringtonia asiatica</i>
	"	/bitu'?un sa dagat/ starfish
	MSK	/bitoon/ be starry (of the sky)

11. s.v. Sont réservés pour les «notes» de chaque article les commentaires particuliers, en fin de compte assez rares, et les reflets aberrants par leur signifiant ou leur signifié. Pour les abréviations des noms de langues, voir la liste en fin d'article.

	BKD	/bituen/ star
	WBM	/bitu'en/ star
	MAR	/bito'on/ star
	LgT	/tekebin/ (< M) star
	SIM	/bintun/ star; name of a carving motif
	SAN	/bituin/ star
	«	/ma-mituin/ throw directly above
	BGG	/bituon/ moon; month
	«	/bituon pauno/ star
	MDR	/bittun/ star
	BGS	/wittun/ star
	MAK	/bintun/ star
	PAL	/btuch/ star; starfish; military star (on hat or clothing)
	CHM	/puti'on/ (< M) star
CMP:	TET	/fitun/ star
	«	/fitun loro teen/ evening star (Venus)
	«	/fitun rai naroma/ morning star (Venus)
	«	/fitun suar/ comet
	WETAN	/witiona/ star
SHWNG:	MAYA	/tuin/, /tue'n/ star
	(8) */pituqon/ star	
OC:	ARS	/hi'u/ star
	BAURO	/higu/ star
	PMC	*/fituu/ star
	GLB	/itoi/ GLOSS
	MSH	/yijiw/ GLOSS
	KUS	/itu/ GLOSS
	PON	/usu/ GLOSS
	TRUK	/fu'u'/ GLOSS
	PUL	/fu'u'/ star, point of the compass, canoe course plotted by the stars
	WOLEAI	/fius(iu)/ (< A) star
	MOTA	/vitu/ star
	SEA	/hitu/ star
	ROT	/hefu/ star
	TON	/fetu'u/ (< A) star; daisy
	NIUE	/fetu_/ star
	SAM	/fetu_/ star
	NUK	/heduu/ star, planet
	REN	/hetu'u/ star, constellation
	NAN	/fetuu/ star; shine, of stars

(2) */bituqen pitu/ the Pleiades (= 'seven stars')

WMP:	[MAL	/bintan tujoh/ a constellation, the Pleiades]
	[ACH	/bintan tujo_h/ a constellation, the Pleiades]
CMP:	[WET	/witiora woiti/ a constellation, <i>Ursa minor</i>]

NOTE: Also PAZ /bintul/ 'star', PAI /vitjuqan/ 'star', PGS /bitu'en/ 'star', KAP /batwi'n/ 'star', KEL /getu'en/ 'star', KEN /betuan/ 'star', MOKEN /bituek/ 'star', TAE' /binto'en/ 'star', TET /fitu/ 'star', WET /witiona/ 'star', HWN

/ho_ku_/ 'star'. The Rotuman and Polynesian reflexes of */bituqen/ show a sporadic lowering of the first vowel and what appears to be regressive assimilation of the last vowel (expected */o/), almost certainly as the result of a single innovation. Other types of irregularities appear in MOTU /hisiu/ 'star', and in the Micronesian reflexes cited here, and it is possible that the POC reflex of PMP */bituqen/ was */pituqun/ rather than */pituqon/. Dahl (1979) interprets the unexpected last vowel of PAI /vitjuqan/ as evidence that */bituqen/ contained a base morpheme */bituq/ which could be suffixed either with */-an/ or */-en/. This argument would be considerably enhanced if at least some comparative evidence were presented for a simple base */bituq/, but such evidence has not been forthcoming. Moreover, the proposal that a base meaning 'star' could be suffixed with */-an/ or */-en/ is itself open to question. In view of these observations I prefer to treat the Paiwan form as exhibiting an unexplained irregularity.

On n'est pas peu surpris de constater une pareille immobilité (hors Océanie) sur six millénaires, hormis quelques passages des plus triviaux de *b à /w/, /p/ ou /f/, ou comme l'amuïssement de C₂ intervocalique dans quelques langues, et enfin comme le passage de trois à deux syllabes (le *q¹² faisant surtout office de «séparateur syllabique» entre *bitu et *-en)¹³.

1.4. Des étymons comportant de nombreux doublets dont une partie est fondée sur un nombre très réduit d'attestations

Cette proximité entre les étymons et leurs reflets dans une partie des langues — celles précisément qui ont servi de point de départ aux reconstructions — a un certain nombre de conséquences.

On peut penser, entre autres, que cette proximité explique en partie le fait que les comparatistes récents spécialistes des langues austronésiennes, et Blust particulièrement, se soient cru autorisés à poser dans la protolange ou une protolange intermédiaire située assez haut dans la chronologie, des étymons fondés sur trois, sinon deux, reflets seulement, cela bien qu'il en résulte, comme on va le voir, un grand nombre de doublets à peu près de même sens dont il est peu probable a priori qu'ils aient cohabité sous cette forme dans une même protolange.

En effet, on a, à côté d'étymons isolés comme *bituqen, un grand nombre d'étymons formant famille, qui, pour un signifié identique, ont la propriété d'avoir en commun leur deuxième syllabe (/C₂V₂C₃/) et de ne se distinguer que par leur première syllabe (/C₁V₁-/)¹⁴. Or, dans la ligne de Dempwolff (et non de Brandstetter, ni de sa propre *Austronesian Root Theory*), Blust pose,

12. Notation d'une des 3 «postvélaïres», *q, *h et *ʔ, en l'occurrence une uvulaire, selon Ross in Tryon 1995, I, p. 57 et 61-62, à la suite de Zorc 1982.

13. Le passage dans quelques langues, ici sans conteste de nature phonétique, de C₂ *t à /nt/ n'est pas sans poser de problèmes, problèmes sur lesquels nous reviendrons plus loin.

14. Dans quelques cas, c'est C₁V₂C₃ qui se rencontre ailleurs (comme -C₂V₂C₃) et -V₁C₂- qui constitue l'élément variant sans entraîner de changement notable de signification: par exemple, on a *binit «to carry» à côté de divers *C₁V₁-bit «prendre entre les doigts», *beNuC «pull out, extract» à côté de divers *C₁V₁-buC «pluck, pull out, uproot, pluck, etc.».

dans son dictionnaire (version électronique 2001) et dans les articles d'*Oceanic linguistics* qui l'ont précédé, autant d'étymons doublets en $/C_1V_1C_2V_2C_3/$, dont les reflets — une fois laissés de côté ceux où C_1V_1- est constitué par une séquence en $*bV(N)$ -¹⁵ (lesquels occupent un tiers du dictionnaire de Blust dans sa version électronique de 2001¹⁶) — sont en nombre souvent limité à 2 ou 3!

L'exemple de l'ensemble des étymons en $*-tas$ (racine glosée «cut through, sever; rip» par Blust) figurant dans ce dictionnaire est tout à fait représentatif de cette situation¹⁷:

(2) $*/bentas/$ hack a passage through vegetation, blaze a trail

WMP:	MAL	/bentas/ tearing up and dashing down; e.g. of a Titan tearing out a hill by the roots and dashing it down on the divine warriors, his enemies
	SND	/bentas/ cut through a fence in order to make a passage to the other side
	SAS	/bentas/ press through something (as underbrush, or a crowd)
CMP:	BM	/bontat/ hack open, hack through, break to pieces
	ROTI	/beta/ top a tree, trim off branches
	SEL	/het/ chop, cut down

(3) $*/maŋ-bentas/$ hack a passage through

WMP:	SAS	/mentas/ press through something (as underbrush, or a crowd)
	BM	/mo-montat/ cut open a new path (through vegetation)

NOTE: With root $*/-tas/$ 'cut through, sever; rip'. Apparently distinct from $*/be(R)tas/$ 'tear, rip open'.

(1) $*/betas/$ tear, rip open (as cloth or stitches), snap (as string)

F:	AMIS	/ftas/ burst open; blow up with insides coming out; blow up
----	------	---

(2) $*/be(R)tas/$ tear, rip open (as cloth or stitches), snap (as string)

WMP:	KAN	/besa't/ torn, rent for the first time; it begins to tear, to be torn, to be rent, to wear out, to be worn out (applied to clothes)
	IFG	/boha't/ break by jerking, by a snapping pull; applied to strings, threads
	TAG	/bigta's/ unstitched
	BKL	/bagta's/ traverse; cut across in walking (as across a field)

15. Nous reviendrons (parag. IV 2) sur ce $*(N)$ qui dès Dempwolff précède la C_2 des étymons.

16. Seules les racines où C_2 est un $*b$ semblent exclure un renforcement au moyen de $*bV(N/R)-$: $*bebak$ — seule exception — est explicable par un redoublement en $*CV-$, simplification ou non d'une reduplication par ailleurs très largement attestée (pour presque toutes les racines): $*bakkak$, $*baŋbaŋ$, $*basbas$, $*bejbej$, $*bekbek$, $*beŋbeŋ$, $*betbet$, $*bidibid$ (?), $*bitbit$, $*buCbuC$, $*budbud$, $*bukbuk$, $*bunbun$, $*buRbuR$.

17. Cf. aussi Blust 1988, *Austronesian Root Theory*, où il prend comme unité de compte étymologique non plus l'étymon en $C_1V_1C_2V_2C_3$, mais la racine en $C_2V_2C_3$.

HAN	/bugta's/ a condition in which something is broken; ripped, unseamed, etc.
CEB	/bugta's/ for string or stitches to snap, break suddenly; rip stitches apart
MSK	/botas/ to leave behind, to separate (as from one's companions)
MAL	/betas/ split open; slit open; ripped open; come undone (of a seam); cracked (of an egg)
CMP:	MGG /wetas/ break, snap (rope, etc.); cut down; tear
NGA	/beta/ to tear, tear off, tear up
"	/beta reta/ tear off
"	/feta/ GLOSS
SIKA	/beta/ to break (as the restraining rope on a horse)
KMB	/botahu/ break off, interrupt, cut off

NOTE: Also KOM /bete'/ 'broken (of a cord), snapped', MGG /wete'/ 'break, snap off (cordage, roots)', ARS /hoa/ 'cut with a bamboo knife'. With root */-tas/ 'cut through, sever, rip'.

(2) */bitas/ tear, rip

WMP:	KAN	/bissa't/ to crack, as when tearing up cloth
	AKL	/bi'tas/ slit open, split up; pull apart
	CEB	/bi'tas/ lacerate, rip (as a nail ripping a dress that catches on it, laceration of the vagina to facilitate birth); rip, slash, tear; unmarried young woman who has had sexual experience
	MSK	/bitas/ to clear (as land), make a clearing; to separate, cut off
	TB	/bintas/ small cut, incision
CMP:	MGG	/wintas/ GLOSS

NOTE: Also TAG /bigta's/ 'unstitched', CEB /bigta's/ 'for stitches or bonds which hold things together to snap; cause them to do so'. With root */-tas/ 'cut through, sever; rip'.

(3) */butas/ √2 cut through, sever

WMP:	TAG	/bu'tas/ perforation, hole; loophole
	"	/buta's/ perforated
	"	/buta's-buta's/ full of holes
	"	/pa-mutas/ perforator
	IBAN	/buntas/ cut through, cut open, lance (as a boil)
	SND	/buntas/ reclaim, as a cultivated field; cut or burrow through (as a mountain); come to an end, of a discussion
	PS	*/butas/ behead, cut off head (as of fish)

NOTE: With root */-tas/ 'cut through, sever; rip'.

(3) */etas/ slash, chop away

dsj. */utas/

WMP:	AKL	/uta's/ slash, cut off (with a blade)
	SAS	/entas/ chop through (forest, hedge)

NOTE: With root */-tas/ 'sever, rip, cut through; short cut'.

(3) **/qantas/* cut through

WMP:	KAD	/antas/ take a shortcut
	BJR	/hantas/ shortcut; take a shortcut
	SMB	/antas/ fell trees, clear brush

NOTE: Also JAV /ancas/ 'take a shortcut', RMB /antat/ 'take a shortcut'. With root **/-tas/* 'cut through, sever; rip'.

(3) **/qutas/* cut through, sever, divide by cutting

WMP:	DGT	/uta's/ finish, complete, terminate
	BON	/o'sat/ (<M) to sever, as the roots of a plant or the strips of wood holding together two sections of an incompletely split log
	ILK	/u'sat/ (<M) open a road, clear a path (in the jungle)
	TAG	/uta's/ completely terminated; dead; completely crazy (about something)
	BKL	/uta's/ detach, remove; kill
	AKL	/uta's/ slash, cut off (with a blade)
	MAR	/otas/ sever completely, cut off completely
	WBM	/utas/ of falling a tree or cutting a log in two, to chop through it
	TIR	/utoh/ to pick, to harvest
	TAO	/utas/ cut something apart, sever something
	IBAN	/utas/ division, lot or parcel of land; divide land with a boundary
	NIAS	/huta/ cut off, cut down
	BAR	/untas-i/ burst open so that the contents spill out
	MDR	/sa-?utas/ a piece (as of rope)
	MAK	/utasa?/ strand, as of silk; piece of a fishnet

NOTE: Also TAO /untas/ 'to cross, go across a road, street, sea, bridge, etc.', SND /utas/ 'chop off, cut, sever'. With root **/-tas/* 'cut through, sever; rip'.

(2) **/ra(n)tas/* cut through, tear

dbl. **/Ratas/*

WMP:	TIR	/rantas/ cut away wrapped lashings
	SND	/rantas/ snap, break off a piece (as of a chain), break (as a snare)
	MAR	/rantas/ cut
	MAL (Ked.)	/rantas/ trace-cutting; cutting a track through the forest

CMP: MGG /ratas/ torn, tattered; to tear

NOTE: With root **/-tas/* 'cut through, sever, tear'. Zorc (n.d.) gives PPH **/raN-tas/* 'cut, sever'.

(3) **/ri(n)tas/* cut or tear off

dsj. **/Ri(n)tas/*

WMP:	MAR	/ritas/ separate, cut off
	IBAN	/rintas/ torn open, broken through
	MAL	/rintas/ take the shortest way, cut across

NOTE: With root **/-tas/* 'cut through, sever; rip'.

(2) **/rutas/* sever, tear

WMP:	HLG	/luta's/ wean; detach affection for something longed for
	MAL	/runtas/ sever with a jerk
	CMP:	MGG /rotas/ tear, make an incision

NOTE: With root **/-tas/* 'cut through, sever, tear'.

(2) **/Ratas/* sever, tear

dbl. **/ra(n)tas/*

dsj. **/Retas/*

WMP:	MIN	/ratas/ severing with a slit
CMP:	MGG	/ratas/ sever, tear

NOTE: With root **/-tas/* 'cut through, sever; rip'.

(2) **/tetas/* rip or tear open

WMP:	MAL	/tetas/ slitting open, ripping up
CMP:	MGG	/tetas/ tear, break open
	BURU	/teta-h/ (of a line) snapped

NOTE: With root **/-tas/* 'cut through, sever; rip'.

Nous y ajouterons le second **butas* répertorié par Blust, qui a une valeur quelque peu différente mais, à nos yeux, certainement apparentée, pour lequel il ne fait pas référence à la racine **tas*:

(3) **/butas/* \1 to separate people, disperse a gathering

WMP:	BKL	/mag-buta's/ to free, set free, release, set loose, emancipate, liberate; extricate; to wean
	"	/maka-buta's/ get or break away, escape
	"	/maka-butas-a'n/ lose control of; fumble; allow to escape
	MSK	/botas/ leave behind, separate (as from a companion)
	MAR	/botas/ to separate, disassociate
	BAL	/buntas/ disperse (a gathering)
	"	/buntas-an/ be made to disperse

Ainsi, à part **qutas* avec 15 reflets¹⁸ et **be(R)tas* avec 13, même les deux **butas* réunis ne totalisent que 8 reflets (avec 4 chacun), **bentas* et **bitas* n'ont que 6 reflets chacun, **rantas* 5, **qantas*, **rutas* et **tetas* 3, **etas* et **Ratas* 2, ce qui est fort peu surtout si l'on considère le nombre de langues. Tous ont un signifié glosé par «tear» et «sever» — sauf **butas*₂ qui a un sens «separate» qui en reste très proche¹⁹.

18. Nous ne comptons qu'un reflet par langue, les formes en **maŋ-*, **mag-*, *maka-*, *-aŋ-* ne peuvent être comptées séparément dans la mesure où ce sont des formes fléchies dont l'inventaire pour chaque langue est très loin d'être complet dans le dictionnaire, ce qui n'a rien d'étonnant pour des langues agglutinantes à la morphologie abondante.

19. On ne peut pas raisonnablement séparer de la racine **tas*, la racine **tes* «tear, rip» (Blust 1988, p. 155) sous laquelle on trouve: **ge(n)tes* «snap, break», **re(n)tes* «sever, cut», **tes* «tear», **testes* «tear», SAS *botes* «broken; piece (of rope)», DPB *etes* «back a path by felling trees and bushes», MAR *ketes* «cutting», SND *pites* «snap, break; snap or break to pieces», SAS *putes* «break (of a rope)». La valeur «tear» n'est pas attachée à tous les étymons

La situation de la famille d'étymons en **tas* n'est pas du tout isolée: un grand nombre de racines en $-C_2V_2C_3$ à valeur plus ou moins verbale, donne lieu à un ensemble d'étymons en $C_1V_1C_2V_2C_3$ à signifiés identiques ou proches et où seul $C_2V_2C_3$ a un signifiant constant. Citons, par exemple, les deux racines **kas*:

**kas* (1) «loosen, undo, untie» avec les étymons: **rika/es* «loosen, undo, untie», **lekas* «open, undress, remove, release», **luṅkas* «remove, loosen, untie», **taṅkas* «loosen, untie», **tukas* «open», **hekas* «release, loosen, untie», **hukas* «loosen, untie, undress; separate», **kaskas* «loosen, untie», **ba(ṅ)kas* «come loose, untied», **bekas* «to spring a trap», **beṅkas* «untie, undo», **biṅkas* «undo, untie, unspring a trap» (sous lequel on trouve: *bigkas* CEB, *igkas* TAG), **bukas* «to open, uncover», **bu(ṅ)kas* «expose, unveil», **pukas* «nudity, exposure of the body», **pukas* «release, loosen, untie», **wakas* «loosen, undo, uncover», **bekas* «trace, track, print (as footprint)», et, peut-être **belekas* «begin»;

**kas* (2) «swift» peut-être < *kas* (1) avec les étymons: **rikas/*rikes* «quick, fast», **la(ṅ)kas* «spirited, energetic», **ta(ṅ)kas* «agile, quick», **cekas* «quick, swift», **Ri(ṅ)kas* «trim», **ba(ṅ)kas* «swift, strong, energetic, fast», **beṅkas* «swift, fast», **bikas* «strong, vigorous, energetic».

Si l'on s'en tient à la lettre **b*, on constate que les racines **bak* «pound, hit» donnent lieu²⁰ à 11 étymons CVCVC, **bak* «split off, bark (of tree)» à 5, **baṅ* «spread out, broad» à 8, **baṅ* «float» à 3, **bas* «demolish» à 2, «cut away» à 4 (+ 2 moins sûrs), **bej* «wind around» à 5, **bek* «strike, break» (cf. **bak?*) à 7, **beṅ* «obstruct» à 6 (+ 1 sur **baṅ*), **beR* «sow, scatter» à 4 (cf. **bud*, **buR?*), **bet* «bind, strap» à 3, **bit* «hook» à 7, **but/C* «pluck, pull out, to weed» à 8, **bud/D* «strew, scatter, sprinkle» à 5, **buk* «pound, knock» à 10, «dust» à 5, **bun* «heap» à 9, **bur* «mix» à 3 (cf. **buR*), **buR* «sprinkle, scatter» à 7, «mix foods» à 3 (cf. **bur* «id.»), «dust» à 2 (cf. **buk* «dust»).

1.5. La multiplication des doublets et la notion de classe d'équivalence

Seul $C_2V_2C_3$ a le statut de signe saussurien associant un signifiant et un signifié; $(C_1)V_1(N/R)-$ a un signifiant variable, et aucun signifié récurrent à travers leurs emplois avec d'autres $C_2V_2C_3$ ne peut leur être attribué, sans qu'on puisse, non plus, assigner un C_1V_1- particulier à un sous-groupe de langues récurrent. Les différents C_1V_1- ²¹ fonctionnent comme des sortes

dérivés de **tes* et encore moins à tous les reflets: cette valeur ne peut donc être considérée comme spécifique de **tes* opposé à **tas* qui présente aussi une valeur «tear» dans une partie de ses reflets. Sur le problème que pose la variation de la voyelle de la racine, voir parag. IV 4.

20. Etant donné que le rattachement de certains étymons à telle ou telle racine peut se discuter, les chiffres fournis ici d'après le dictionnaire (version 2001) le sont uniquement pour donner une idée de l'extension du phénomène; Blust y consacre l'appendice 2 de son *Austronesian Root Theory* qui occupe 82 pages sur 190.

21. Egalement C_1V_1N- et, dans quelques cas, $-V_1C_2-$.

d'«élargissements radicaux préfixés» (et non suffixés comme dans les langues indoeuropéennes ou dans les langues bantoues²²), et les différents «étymons» en $*C_1V_1-tas$ apparaissent comme des doublets²³. La même protolangue — à savoir celle de niveau étiqueté (2) par Blust, sauf pour les deux $*butas$ limités au niveau (3) — ne peut guère les avoir possédés en tant que tels en même temps.

Du fait de la faible divergence attendue entre étymons et reflets, l'étymon semble à portée de main dès qu'on a quelques formes concordantes dans un petit nombre de langues plus ou moins dispersées. C'est sans doute ce qui explique qu'on ait pu considérer aussi facilement $C_1V_1C_2V_2C_3$ comme unité de compte étymologique, alors qu'il est évident que seuls les $C_2V_2C_3$ peuvent être reportés dans la protolangue. Effectivement, on constate que les spécialistes de la grammaire comparée des langues austronésiennes n'ont cessé d'hésiter entre considérer la «racine» en $C_2V_2C_3$ comme unité de compte étymologique (position de Brandstetter²⁴, et de Blust dans sa *Austronesian Root Theory*) et l'«étymon» en $C_1V_1C_2V_2C_3$ (position de Dempwolff et de Blust dans son «dictionnaire» et ses articles d'*Oceanic Linguistics*).

Le fait de poser autant de formes sources faisant double, triple, etc., emploi dans la protolangue a sans doute été favorisé aussi par la tendance générale apparue au lendemain de la seconde guerre mondiale consistant à envisager une protolangue reconstruite comme un ensemble de simples classes d'équivalence entre reflets attestés. On retrouvera la même attitude en phonologie, avec les mêmes effets: poser autant de protophonèmes que de classes d'équivalence entre reflets, contre toute vraisemblance.

2. Le «grand tournant» de l'après-guerre: «subgrouping» et migration

La grammaire comparée des langues austronésiennes a connu, dans l'immédiat après-guerre, essentiellement avec les travaux d'I. Dyen, un tournant et, dans une certaine mesure, une rupture épistémologique (qui est aussi un

22. On a beaucoup critiqué la notion d'«élargissement radical» en ce qui concerne le proto-indoeuropéen; pourtant, les langues indoeuropéennes ne sont pas les seules où, affixé à une base associant signifiant et signifié, on trouve un élément récurrent mais variable et sans signifié distinct restituable; les langues bantoues présentent la même situation, où, à côté de suffixes constituant de véritables morphèmes (de diathèse, etc.), on a des «suffixes» sans signifié.

23. Au sens où l'on peut supposer qu'ils sont issus d'un même signe saussurien, sur lequel ils sont formés par deux voix différentes, qui divergent non plus ici du fait de lois phonétiques différentes, celles de l'héritage vs de l'emprunt comme dans le cas de ce qu'il est convenu d'appeler d'ordinaire des doublets, tels que *grêle* et *gracile* en français, mais du fait de phénomènes qui doivent relever d'une manière ou d'une autre de la morphologie.

24. Brandstetter 1910, p. 4 et compte-rendu de G. Ferrand dans *BSLP* 16, p. CCVIII sqq. («La conclusion très nette est qu'un grand nombre de thèmes radicaux dissyllabiques peuvent être ramenés à une racine monosyllabique»).

changement de tradition) sans précédent. Jusque là, elle avait été dominée par les travaux des écoles néerlandaise (avec Van der Tuuk, Brandes, H. Kern), puis allemande (avec Brandstetter, et surtout Dempwolff à qui l'on doit la démonstration définitive de la parenté entre langues «indonésiennes» et océaniques²⁵, ces deux derniers auteurs clairement dans la ligne des néo-grammairiens).

Ce tournant se caractérise par la rencontre d'un ensemble de faits qui ne sont qu'en partie liés: 1) la prise en compte des langues de Formose, 2) l'apparition de la lexicostatistique et de la glottochronologie, et — corollaire héritée de la tradition américaniste — l'intérêt pour les «migrations», et enfin 3) une méthode mécaniste de reconstruction des protosystèmes phonologiques.

2.1. «Découverte» des langues de Formose et/ou changement de paradigme?

La découverte des langues de Formose, de leur variété et de leurs singularités, a amené à les considérer comme formant un sous-ensemble à part par rapport aux langues «malayopolynésiennes», quels que soient les regroupements que l'on opère dans cet ensemble et la façon dont on situe les langues malayopolynésiennes par rapport aux langues de Formose — «high order subgrouping» si les langues malayopolynésiennes forment un groupe parallèle à celui (ou ceux) formé(s) par les langues de Formose ou «low order subgrouping» si l'ancêtre commun des langues malayopolynésiennes est considéré comme un élément d'un sous-(sous-)groupe des langues de Formose²⁶. En fait, il ne s'agit pas d'une véritable découverte; dès Brandstetter, on trouve le «formosan» cité parmi les langues austronésiennes fournissant des reflets aux étymons proposés; par ailleurs, les anthropologues japonais avaient, dès avant la première guerre mondiale, mais surtout dans l'entre-deux-guerres, accumulé de nouvelles données qui ont, entre autres, révélé la diversité des langues de Formose²⁷.

Ce sont essentiellement les travaux de Dyen²⁸, ceux d'avant sa conversion à la lexicostatistique et encore inscrits dans la tradition néo-grammairienne, qui déclenchent ce processus de retournement — les langues de Formose seraient «archaïques» et Formose finira par être considéré comme le berceau de la famille —, retournement comparable, par certains aspects, à ce qui

25. Le samoan est déjà compté parmi les langues malayopolynésiennes par Meillet dans son compte rendu de Finck en 1909.

26. Pour une présentation plus détaillée du problème des «subgroupings», voir Lemaréchal 2001, p. 424 sqq.

27. Cf. Ogawa et Asai, 1935 «Les Mythes et traditions des tribus indigènes de Formose», en japonais, p.239-327, Université Impériale Taihoaku à Taipei; Abe, 1930, qui étudie les structures syntaxiques d'un dialecte du puyuma; Asai, 1936, «A study of the Yami language: an indonesian language spoken on Botel Tobago Island» publié à Leyde.

28. Dyen 1947, 1951, 1953.

s'est passé avec l'«expansion bantoue» telle qu'elle a été revisitée au début des années 70²⁹.

2.2. Une drôle de liste

La prise en compte des langues de Formose a joué un rôle essentiel, avec les premiers travaux de Dyen, mais le tournant de l'après-guerre consiste d'abord et surtout en un changement de paradigme. Un de ses aspects les plus lourds de conséquence a certainement été l'irruption de la lexicostatistique et de la glottochronologie (avec une importance primordiale accordée, dans la comparaison, au vocabulaire sous forme de listes de mots, et au problème des «subgroupings»): la centralité des études lexicales est déjà présente dans la littérature antérieure, et s'explique aussi en partie par l'état de la documentation, souvent limitée à des listes de mots et à des vocabulaires.

Dyen a été un des premiers à appliquer systématiquement les procédures de Swadesh. D'abord resté dans la droite ligne des néogrammairiens, il en est devenu un des plus fervents partisans et les langues austronésiennes un des domaines d'application privilégiée. La lexicostatistique est fondée explicitement sur la ressemblance directement perceptible entre item de langues supposées apparentées et sur leur comparaison de masse³⁰, qui donnerait la possibilité d'un traitement statistique.

Dyen travaille sur la «200-word basic vocabulary list» suivante (Dyen 1975, p. 230-233):

1. all	68. head	136. to sit
2. and	69. to hear	137. skin (of person)
3. animal	70. heart	138. sky
4. ashes	71. heavy	139. to sleep
5. at	72. here	140. small
6. back	73. to hit	141. to smell
7. bad	74. hold (in hand)	(perceive odor)
8. bark (tree)	75. how	142. smoke
9. because	76. to hunt	143. smooth
10. belly	77. husband	144. snake
11. big	78. I	145. snow
12. bird	79. ice	146. some
13. to bite	80. if	147. to spit

29. Cf. Hyman et Vorhoeve (éds.), *L'expansion bantoue*, I, 1980. Dans les deux cas, on constate un changement du lieu d'origine de l'expansion de la famille — de l'Asie du Sud-est à Formose pour les langues austronésiennes, de la région des grands lacs à celle des Grass-fields pour les langues bantoues — solidaire d'un changement de perspective sur des langues marginalisées jusque là. Dans les deux cas, ce sont des considérations néogrammairiennes et non la lexicostatistique qui conduisent à ce retournement: $\pm m$ - de certaines marques de classe pour les langues bantoues, *S, forme de la marque de 2sg, etc. pour les langues austronésiennes.

30. Cf. Fodor et Rapal, *BSLP* 101/I, p. 437.

- | | | |
|---------------------|----------------------|----------------------|
| 14. black | 81. in | 148. to split |
| 15. blood | 82. to kill | 149. to squeeze |
| 16. to blow (wind) | 83. know (facts) | 150. to stab |
| 17. bone | 84. lake | (or stick) |
| 18. to breathe | 85. to laugh | 151. to stand |
| 19. to burn (intr.) | 86. leaf | 152. star |
| 20. child (young) | 87. left (hand) | 153. stick (of wood) |
| 21. cloud | 88. leg | 154. stone |
| 22. cold (weather) | 89. to lie (on side) | 155. straight |
| 23. to come | 90. to live | 156. to stuck |
| 24. to count | 91. liver | 157. sun |
| 25. to cut | 92. long | 158. to swell |
| (with knife) | 93. louse | 159. to swim |
| 26. day (not night) | 94. man (male) | 160. tail |
| 27. to die | 95. many | 161. that |
| 28. to dig | 96. meat (flesh) | 162. there |
| 29. dirty | 97. mother | 163. they |
| 30. dog | 98. mountain | 164. thick |
| 31. to drink | 99. mouth | 165. thin |
| 32. dry (substance) | 100. name | 166. to think |
| 33. dull (knife) | 101. narrow | 167. this |
| 34. dust | 102. near | 168. thou |
| 35. ear | 103. neck | 169. three |
| 36. earth (soil) | 104. new | 170. to throw |
| 37. to eat | 105. night | 171. to tie |
| 38. egg | 106. nose | 172. tongue |
| 39. eye | 107. not | 173. tooth (front) |
| 40. to fall (drop) | 108. old | 174. tree |
| 41. far | 109. one | 175. to turn (veer) |
| 42. fat (substance) | 110. other | 176. two |
| 43. father | 111. person | 177. to vomit |
| 44. to fear | 112. to play | 178. to walk |
| 45. feather (large) | 113. to pull | 179. warm (weather) |
| 46. few | 114. to push | 180. to wash |
| 47. to fight | 115. to rain | 181. water |
| 48. fire | 116. red | 182. we |
| 49. fish | 117. right (correct) | 183. wet |
| 50. five | 118. right (hand) | 184. what? |
| 51. to float | 119. river | 185. when? |
| 52. to flow | 120. road | 186. where? |
| 53. flower | 121. root | 187. white |
| 54. to fly | 122. rope | 188. who? |
| 55. fog | 123. rotten (log) | 189. wide |
| 56. foot | 124. rub | 190. wife |
| 57. four | 125. salt | 191. wind (breeze) |
| 58. freeze | 126. sand | 192. wing |
| 59. fruit | 127. to say | 193. wipe |
| 60. to give | 128. scratch (itch) | 194. with |
| 61. good | 129. sea (ocean) | (accompanying) |

s'est passé avec l'«expansion bantoue» telle qu'elle a été revisitée au début des années 70²⁹.

2.2. Une drôle de liste

La prise en compte des langues de Formose a joué un rôle essentiel, avec les premiers travaux de Dyen, mais le tournant de l'après-guerre consiste d'abord et surtout en un changement de paradigme. Un de ses aspects les plus lourds de conséquence a certainement été l'irruption de la lexicostatistique et de la glottochronologie (avec une importance primordiale accordée, dans la comparaison, au vocabulaire sous forme de listes de mots, et au problème des «subgroupings»): la centralité des études lexicales est déjà présente dans la littérature antérieure, et s'explique aussi en partie par l'état de la documentation, souvent limitée à des listes de mots et à des vocabulaires.

Dyen a été un des premiers à appliquer systématiquement les procédures de Swadesh. D'abord resté dans la droite ligne des néogrammairiens, il en est devenu un des plus fervents partisans et les langues austronésiennes un des domaines d'application privilégiée. La lexicostatistique est fondée explicitement sur la ressemblance directement perceptible entre item de langues supposées apparentées et sur leur comparaison de masse³⁰, qui donnerait la possibilité d'un traitement statistique.

Dyen travaille sur la «200-word basic vocabulary list» suivante (Dyen 1975, p. 230-233):

1. all	68. head	136. to sit
2. and	69. to hear	137. skin (of person)
3. animal	70. heart	138. sky
4. ashes	71. heavy	139. to sleep
5. at	72. here	140. small
6. back	73. to hit	141. to smell
7. bad	74. hold (in hand)	(perceive odor)
8. bark (tree)	75. how	142. smoke
9. because	76. to hunt	143. smooth
10. belly	77. husband	144. snake
11. big	78. I	145. snow
12. bird	79. ice	146. some
13. to bite	80. if	147. to spit

29. Cf. Hyman et Vorhoeve (éds.), *L'expansion bantoue*, I, 1980. Dans les deux cas, on constate un changement du lieu d'origine de l'expansion de la famille — de l'Asie du Sud-est à Formose pour les langues austronésiennes, de la région des grands lacs à celle des Grass-fields pour les langues bantoues — solidaire d'un changement de perspective sur des langues marginalisées jusque là. Dans les deux cas, ce sont des considérations néogrammairiennes et non la lexicostatistique qui conduisent à ce retournement: $\pm m$ - de certaines marques de classe pour les langues bantoues, *S, forme de la marque de 2sg, etc. pour les langues austronésiennes.

30. Cf. Fodor et Rapal, *BSLP* 101/I, p. 437.

- | | | |
|---------------------|----------------------|----------------------|
| 14. black | 81. in | 148. to split |
| 15. blood | 82. to kill | 149. to squeeze |
| 16. to blow (wind) | 83. know (facts) | 150. to stab |
| 17. bone | 84. lake | (or stick) |
| 18. to breathe | 85. to laugh | 151. to stand |
| 19. to burn (intr.) | 86. leaf | 152. star |
| 20. child (young) | 87. left (hand) | 153. stick (of wood) |
| 21. cloud | 88. leg | 154. stone |
| 22. cold (weather) | 89. to lie (on side) | 155. straight |
| 23. to come | 90. to live | 156. to stuck |
| 24. to count | 91. liver | 157. sun |
| 25. to cut | 92. long | 158. to swell |
| (with knife) | 93. louse | 159. to swim |
| 26. day (not night) | 94. man (male) | 160. tail |
| 27. to die | 95. many | 161. that |
| 28. to dig | 96. meat (flesh) | 162. there |
| 29. dirty | 97. mother | 163. they |
| 30. dog | 98. mountain | 164. thick |
| 31. to drink | 99. mouth | 165. thin |
| 32. dry (substance) | 100. name | 166. to think |
| 33. dull (knife) | 101. narrow | 167. this |
| 34. dust | 102. near | 168. thou |
| 35. ear | 103. neck | 169. three |
| 36. earth (soil) | 104. new | 170. to throw |
| 37. to eat | 105. night | 171. to tie |
| 38. egg | 106. nose | 172. tongue |
| 39. eye | 107. not | 173. tooth (front) |
| 40. to fall (drop) | 108. old | 174. tree |
| 41. far | 109. one | 175. to turn (veer) |
| 42. fat (substance) | 110. other | 176. two |
| 43. father | 111. person | 177. to vomit |
| 44. to fear | 112. to play | 178. to walk |
| 45. feather (large) | 113. to pull | 179. warm (weather) |
| 46. few | 114. to push | 180. to wash |
| 47. to fight | 115. to rain | 181. water |
| 48. fire | 116. red | 182. we |
| 49. fish | 117. right (correct) | 183. wet |
| 50. five | 118. right (hand) | 184. what? |
| 51. to float | 119. river | 185. when? |
| 52. to flow | 120. road | 186. where? |
| 53. flower | 121. root | 187. white |
| 54. to fly | 122. rope | 188. who? |
| 55. fog | 123. rotten (log) | 189. wide |
| 56. foot | 124. rub | 190. wife |
| 57. four | 125. salt | 191. wind (breeze) |
| 58. freeze | 126. sand | 192. wing |
| 59. fruit | 127. to say | 193. wipe |
| 60. to give | 128. scratch (itch) | 194. with |
| 61. good | 129. sea (ocean) | (accompanying) |

62. grass	130. to see	195. woman
63. green	131. seed	196. woods
64. guts	132. to sew	197. worm
65. hair	133. sharp	198. ye
66. hand	134. short	199. year
67. he	135. to sing	200. yellow

200 mots, c'est peu; or le nombre de «notions» (?) inopérantes au pôle ou à l'équateur (cf. «freeze», «snow»), ou 6000 ans avant le présent (n° 132 «to sew»), etc., a pour effet de réduire la liste des «mots» utilisables à un nombre inférieur à 200: Dyen a travaillé sur 150 mots («150 or more», p. 91)! On pourrait donc croire qu'avec les progrès des moyens de calcul, il est souhaitable, afin d'en améliorer la pertinence et la valeur, d'augmenter encore ce chiffre. C'est ce que Tsuchida a fait pour les langues de Formose avec une liste de plus de 700 items. Reid, dans ses *Philippine Minor Languages* (1971), utilise la liste de 304 «meanings» établie par Elkin pour les membres du Summer Institute of Linguistics, qui comprend 170 éléments de la liste de Dyen additionnés de certains éléments «gleaned from the TRIPP questionnaire, others from Conklin's Test Vocabulary for Philippine languages». Mais, en augmentant la liste, on s'enferme dans une contradiction: dans un sens, le mot clé est «basic» et, plus on augmente la liste, plus on s'éloigne du caractère basique des «meanings» retenus, prétendue garantie (comme on va le voir) contre le risque de prendre des emprunts pour des termes hérités.

Le terme *basic* garde aussi, implicitement, son autre sens de «non dérivé» — ce qui a à voir avec le renouvellement des signifiants pour un même signifié —: sur 6000 ans (date supposée d'arrivée à Formose, plus ou moins proche, selon les hypothèses, du début de différenciation entre langues) ou même 3000 ans (date supposée des premiers départs de Formose), un «vent» non dérivé a largement le temps d'être remplacé par un «souffl-ant», une «dent» par un «mord-ant» ou un «mange-ant», ce type de renouvellement a encore plus de chances de résulter de développements parallèles que les autres types.

Par ailleurs, on contestera le bien-fondé du parti choisi par Reid (op. cit., p. viii) consistant à éliminer «one member of frequently occurring doublets, such as husband/wife, foot/leg, meat (flesh)/fruit, wood/stick, sun/day», sans doute pour ne pas fausser le comptage, mais cela ne fait que poser de façon plus voyante le traitement infligé aux systèmes sémantiques des langues et aux taxinomies en général. Il semble en outre qu'un tel parti contrevienne de toutes manières au prétendu caractère «en aveugle» de la procédure de Swadesh, caractère défendu par Dyen comme gage de scientificité. On peut enfin s'interroger sur le caractère universel de ces «doublets» (comme «foot/leg») et, par contre-coup, sur la validité des critères du «basic» selon Swadesh-Dyen ou, inversement, sur ce qui soustend l'idiosyncrasie de doublets comme «meat(flesh)/fruit».

Un autre trait qui nous paraît suspect dans la liste de «meanings» de Dyen (à la suite de Swadesh) est le nombre excessif de noms — qui plus

est, de noms concrets³¹ — dans la liste des étymons, surtout comparée à celles reconstruites pour des familles ou branches de familles comme les langues indoeuropéennes, sémitiques, bantoues. Sur les 1650 étymons (dont 359 PAN et 665 PMP, les autres étant limités à des sous-groupes plus réduits) répertoriés par Zorc in Tryon 1995, on ne trouve que 309 bases verbales, dont beaucoup de bases manifestement onomatopéiques — donc peu utilisables à des fins de reconstruction du fait de leur remotivation constante (avec de nombreux synonymes signifiant «suck», «squeeze», etc.) —, à côté de très nombreux noms d'espèces (végétaux, poissons) par définition facilement empruntables, vu que ces noms comme leur référent sont par définition plus attachés à une aire qu'à un héritage.

On est aussi, évidemment, étonné de rencontrer dans la liste des «mots grammaticaux» comme *and*, *at*, *because*, *in*, *if*, *not*, *he*, *how*, etc., dont la polysémie ou la complexité dans les langues sont telles que les risques de confusion, de fantômes, etc., sont considérables, sans parler du fait que leurs équivalents sont souvent des affixes ou des flexions et non des «mots»; d'une manière générale, le traitement de marques grammaticales comme des items lexicaux séparés pèse lourd dans les reconstructions proposées par la linguistique comparative des langues austronésiennes telle qu'elle est pratiquée actuellement.

L'établissement d'une telle liste repose sur un ensemble de naïvetés proclamées par Dyen (p. 93-94):

«We can construct a diagnostic list of meanings which we will call a vocabulary. If for some set of languages we collect the words for these meanings, we can proceed to compare the words in the different lists to see whether they are cognate».

Il s'agit donc de listes de traductions et non de sens, du genre des listes de mots obtenus par les premiers missionnaires et explorateurs, et qui constitue le premier pas du processus d'équipement d'une langue en outils descriptifs-normatifs que S. Auroux a appelé «grammatisation»³². Plus grave

31. Cette prééminence du concret et l'importance accordée à la source déonomatopéique d'une partie des étymons nous paraît renvoyer à une certaine image, assez douteuse il faut le dire, de ce qui est «primitif». Dans le dictionnaire de R. A. Blust, le nombre de noms d'espèces, de mots techniques est également frappant, de même que le nombre de bases signifiant «thub», «pound», «hit», etc.

32. Déjà cité en note 4. On rappellera à ce propos qu'on ne dispose de dictionnaires étymologiques au sens habituel du terme pour aucune des langues de la famille. Le dictionnaire de Blust est un dictionnaire «descendant», qui part de l'étymon, il ressemble en cela à un dictionnaire modèle comme le FEW de Warburg; mais il ne faut pas oublier que, dans le cas de ce dernier, la protolangue, ou une langue qui en est proche, est connue sous la forme du latin, sans compter que le FEW venait après une longue série de travaux et que la reconstruction d'étymons ne posait plus de problèmes majeurs; ce n'est évidemment pas le cas de celui de Blust. En fait, pour beaucoup de langues, une simple recension montre qu'il s'agit seulement de la compilation électronique des dictionnaires bilingues (vers l'anglais) disponibles: c'est le cas pour l'ilocano par exemple.

encore: pour Dyen, «mots qui se ressemblent» égale *cognate*. Dyen enchaîne:

«Cognate with the same meaning are homosemantic cognates»

ce genre de définition préalable, faussement précise, est un des ingrédients récurrents de la rhétorique scientifique de Dyen; on notera que l'existence de la polysémie, de l'homonymie et de la synonymie est oubliée: la moindre différence de catégorisation ou de niveau de langue, sinon le simple hasard, occultera l'existence de «cognates». La seule incertitude de traductions qui s'effectuent dans le sens du «thème» et non de la «version» suffirait à en anéantir toute valeur.

Malgré des ajustements, un bon nombre de regroupements proposés pour les langues austronésiennes reposent encore aujourd'hui entièrement sur ces «données» et ce genre de statistiques constitue une sorte de passage obligé au début de bien des descriptions, soit comme hypothèse à vérifier, soit «en attendant mieux». Cette dernière attitude, «prudente», n'en est pas moins contestable: on a perdu et on continue de perdre beaucoup de temps à faire des listes de mots et des calculs, au lieu d'élaborer, «sans arrière-pensée», des grammaires et des lexiques de langues en danger, et de procéder, quand il s'agit de comparaison, selon les méthodes de la véritable grammaire comparée; une mauvaise méthode ne peut faire gagner du temps. De mauvais «subgroupings» ne peuvent que multiplier étymons et protophonèmes fantômes, qui constituent de véritables écrans.

2.3. La lexicostatistique et le problème des emprunts

Dans toute reconstruction d'un protovocabulaire, quels qu'en soient le but et la méthode, le danger est de prendre des emprunts plus ou moins anciens pour des termes hérités. La lexicostatistique y échappe moins que toute autre méthode. Dyen en a évidemment conscience:

«The determination that we are dealing with cognates rather than borrowings is in some cases difficult to make (...) To keep the number of indeterminate³³ borrowings small we choose meanings for which the number of borrowings will probably be small in any case. It has become clear over years of investigation that the number of borrowings in meanings that we call *basic* is usually very small³⁴. The words for "ear", "nose", "river", "sand", "die", and the like seem relatively rarely to be borrowed»³⁵

33. On ne s'en donne pas les moyens par un examen détaillé des mots utilisés pour une langue donnée; on ne cherche même pas du tout à le faire: les listes sont en fait utilisées telles quelles.

34. Il est net que l'idée de partir de «basic meanings» est première et non le fruit d'expérimentations; ce sont simplement les premiers mots disponibles — car les plus «utiles» — dans les listes de mots.

35. On remarquera qu'une partie de ces mots sont des emprunts en anglais même (*animal, because, count, fruit, person, they, vomit*); on peut d'ailleurs douter que les conjonctions de

La liste des 200 mots est édifiante à cet égard, et, parmi les mots qui ont des «cognates» dans un nombre satisfaisant — nombre à définir — de branches de la famille indoeuropéenne, un bon nombre n'apparaîtrait pas dans les listes à cause des changements de sens: ce dernier point ne serait pas grave si les listes de départ contenaient beaucoup plus que les 200 mots et étaient établies par de bons connaisseurs de la langue considérée dans sa globalité, et non des enquêteurs avant tout en chasse de listes de mots. Dyen enchaîne:

«Our confidence that borrowings play an unimportant role except locally in the determinations is also in part based on the wide distribution of the languages»

Certes, mais vu l'échelle de temps considérée, si dans certaines régions comme l'Ile de Pâques ou Madagascar, la possibilité de contacts avec des langues apparentées ou de contacts partagés avec des langues non apparentées semble improbable, ce n'est pas le cas pour de nombreuses autres; de plus, on ne sait pas à quelles dates certaines populations de l'intérieur ont été soustraites au contact des populations fréquentant les côtes; effectivement, si l'on examine concrètement les reflets des étymons, on s'aperçoit que bon nombre d'entre eux n'est attesté que dans des régions où des contacts réitérés sur un laps de temps long ne peuvent pas être écartés.

2.3. *L'intérêt pour les migrations*

Corollaire hérité de la tradition américaniste (Dyen se réclame dans ce domaine explicitement de Sapir³⁹): l'intérêt pour les «migrations»³⁷ et

cause soient très stables dans les langues, d'où elles sont souvent tout simplement absentes! malgré le prétendu primitif sémantique CAUSE, primitif cognitif peut-être mais certainement pas linguistique (cf. Lemaréchal 1998a, p. 229 sqq.).

36. Voici l'exemple même de raisonnement utilisé par Sapir (1916, éd. 1949, p. 456) cité par Dyen à l'appui de sa méthode: «The Great Lakes can be considered as the historical centre of distribution of only the Central-Eastern tribes, while the linguistic equivalence with this group of the Blackfoot, Arapaho, and Cheyenne, each of which lie to the west of the former, pushes the historical centre of distribution of the Algonkin tribes proper considerably to the west. (footnote: This naturally has its significance in view of the presence of Yurok and Wiyot still farther west. It is hardly an accident that the greatest linguistic differentiation of Algonkin proper is found in the west, not in the Atlantic region.) We can hardly avoid the inference that in the remote past the general movement of Algonkin tribes were from west to east (Footnote: This in no way contradicts the fact that at a much later period there was clearly a westward drift of certain Algonkin tribes (Western Cree, Plains Ojibwa, Arapaho, Cheyenne)». Cf. Lemaréchal 2001, p. 420 note 3.

37. Dyen présente même cet intérêt comme premier (1975, p. 91 = *Language* 38, 1962): «The stimulus for such a classification came from the interest of ethnologists and archeologists working in the Pacific area. It was their hope that the implications of the distribution of Malayopolynesian languages for the migrations of their speakers might suggest to them probable fruitful areas of investigations». On ne voit pas pourquoi l'intérêt pour l'histoire et la préhistoire des langues doit être subordonné à celui pour les mouvements de population; comme nous l'écrivions dans le *BSLP* 2001 (p. 402, note 4), nous revendiquons l'autonomie de la grammaire comparée.

l'extrapolation de la lexicostatistique dans le domaine géographique sous la forme de théories des migrations, après son extrapolation historique sous la forme de la glottochronologie. Le style se veut de nouveau rigoureux et scientifique:

«The postulates of migration theory are two: (1) *The area of origin of related languages is continuous*, and (2) *The probabilities of different reconstructed migrations are in inverse relation to the number of reconstructed language movements that each requires.*» (Dyen 1975, p. 54)

«For any distribution of two units or more³⁸ there are a number of different available hypotheses. Each such hypothesis consists of two parts: (1) a posited homeland or center of distribution, and (2) the movements minimally necessary for each language not in the homeland to have reached its observed location» (p. 60)

«The unit of comparison or measurement of different possible migrations is the MOVE³⁹. A move as a quantity is the movement of one language necessitated by a possible migration (...) By counting the moves implied in each, we reach the hypothesis with the least number of moves; this is considered to have the highest probability» (p. 62-63)

Qu'est-ce qu'un «language movement»: changement linguistique ou migration de population? Quant au terme «probability», il a, on s'en doute, bon dos: sens technique ou sens usuel? Il cache beaucoup de pétitions de principe dont on remet à plus tard tout début de démonstration. Pour chaque déduction, Dyen donne en illustration des exemples de distribution de tel ou tel groupe de langues indoeuropéennes; le problème, c'est de savoir si le raisonnement permet d'établir ces faits en l'absence de ce qui a permis effectivement de les établir.

Dans le cas des langues austronésiennes, il s'agit d'îles entourées d'eau:

«Any two units of distribution determine an INTERVAL (...) The boundaries of the units which face each other across the interval are the FACING BOUNDARIES of the interval (...) Intervals are either OCCUPIED (by a language) or UNOCCUPIED (...) There is undoubtedly a difficult borderline in the size and nature of unoccupied intervals; so that in some instances (for example if the interval is a wide river) it would be arbitrary to choose between saying that two languages shared a boundary or an unoccupied interval. In such instances the distribution is ambiguous.» (p. 58-59)

Après quelques exemples, dont celui du français de France et celui du Vietnam, Dyen, comme dans tous les cas où il y a une «difficulté» qu'il

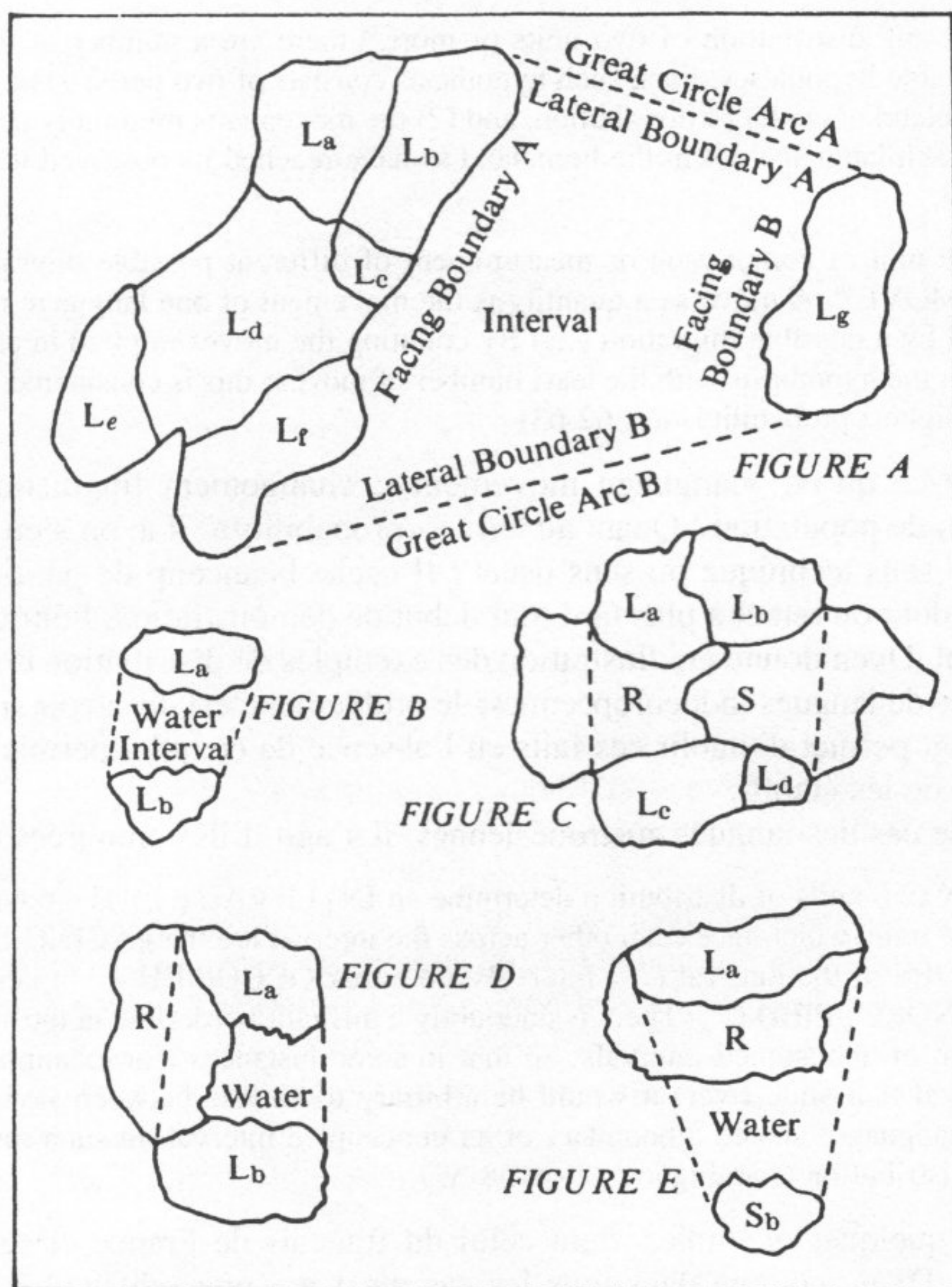
38. Si «units» désignent les langues; il s'agit donc de la distribution géographique des différentes langues.

39. Ce concept est la clé d'un certain nombre de raisonnements faux puisqu'il permet de confondre nombre de langues migrantes (comment les compter? et que fait-on de celles qui ont disparu, puisque la séparation est ancienne et le comptage actuel?) et nombre de déplacements d'une langue.

signale explicitement, tranche par une définition/simplification (un axiome?):

«Since such intervals are probably rare⁴⁰, it is convenient to limit the type of interval considered; the term interval is used in the following in the meaning 'interval with a route whose passability is not less than that of any other route between the units'⁴¹» (p. 59)

On aboutit, p. 61, à la carte suivante:

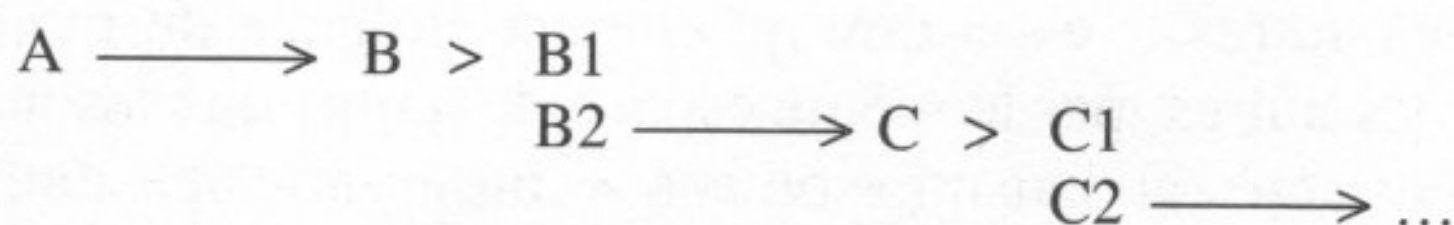


40. Au nom de quoi Dyen peut-il affirmer cela? En exagérant à peine, même si les distances sont moindres qu'entre le Vietnam et la France, cela pourrait parfaitement être le cas de presque toutes les langues austronésiennes!

41. Voilà bien ce dont il est impossible de juger, et d'ailleurs Dyen se garde bien, dans la suite, de poser concrètement le problème et de proposer une quelconque hypothèse sur les routes précises suivies par les différentes migrations.

On ne peut admettre cette extension du modèle des américanistes et de Sapir en particulier, traitant de langues parlées par des populations dans un milieu à peu près homogène: les mouvements de populations en milieu terrestre — jouant sur des surfaces et peut-être sur un intervalle de temps relativement réduits — ne sont pas directement transposables au cas des langues austronésiennes. La mer n'est pas une steppe; les points y sont disjoints, l'espace entre ces points est vide; au moins sur les derniers 3000 des 6000 ans qui séparent le PAN des langues actuelles, on a pu aller d'un de ces points à un autre en contournant un point; sur un assez long laps de temps, il existe plusieurs domaines dans l'espace géographique où l'on parle des langues austronésiennes, où les migrations — aussi bien d'ailleurs que les emprunts — ont pu avoir lieu de n'importe quel point à n'importe quel point⁴².

D'une manière générale, ces modèles migratoires privilégient implicitement un mode de migration en sauts de puce successifs, avec division de la population entre partie restant sur place, dont la langue est censée cesser d'évoluer, ou évoluer moins vite, et partie quittant ce lieu pour le point suivant:



Or, ce mode d'expansion, correspondant à une sorte de *Ver sacrum*, n'en est qu'un parmi d'autres. Les langues peuvent s'étendre selon des processus beaucoup plus diversifiés. D'abord une langue peut s'étendre ou se déplacer parce que ses locuteurs se déplacent, ou bien, si une population change de langue, alors c'est seulement la langue qui s'étend sans qu'il y ait nécessairement déplacement massif de population; dans les études austronésiennes, populations et langues sont trop souvent confondues⁴³: l'intérêt pour les migrations n'a pu qu'encourager cette tendance. De plus comme il s'agit, pour un grand nombre d'entre elles, de langues parlées par de petits nombres de locuteurs, le déplacement de population à l'origine de leur séparation a pu être limité à une seule famille⁴⁴. Cette expansion peut être également concentrique, par vagues successives: d'où l'idée que des archaïsmes peuvent être partagés par des langues parlées sur des confins opposés du domaine; ce mode de progression n'est nullement exclu pour la famille austronésienne, langues de Formose comprises. Une population — et, dans ce

42. Comme l'a bien souligné Jeff Mark dans sa communication sur «Austronesian languages, matrilinearity in Proto-Oceanic Society and the human genetics of the Polynesians and Micronesians» à la journée d'études organisée par L. Sagart le 5 mai 2004 sur le thème *Les premiers austronésiens: langues, gènes, systèmes de parenté*. D'une manière générale, nous trouvons que, jusqu'ici, la démographie historique n'est que d'un mince secours pour formuler et même départager les hypothèses (Lemaréchal 2001, p. 421 note 6).

43. Cf. Lemaréchal 2001, p. 419-424.

44. Jeff Mark a également insisté sur ce point dans la communication citée ci-dessus.

cas, par conséquent, une langue — peut en pousser une autre devant elle qui peut se retrouver aux confins, ou refoulée à l'intérieur des terres: ce dernier mode de progression est manifestement attesté dans les îles les plus grandes. Une langue peut se répandre à travers une hiérarchie de centres directeurs: cela a été le cas pour les langues romanes, cela peut aussi l'être pour certaines époques et certaines aires du domaine austronésien.

Le schéma en saut de puce semble associé, chez certains auteurs, à l'idée que la langue des locuteurs les plus mobiles évolue plus vite que celle des locuteurs qui le sont moins: il n'y a aucune raison pour qu'il en soit ainsi, sauf contact avec d'autres langues, sauf changement de langues, etc. La langue de locuteurs restés un moment sur place peut se voir fortement affectée par l'arrivée d'autres populations qui les soumettent, etc. L'idée que les langues restées sur place sont un reflet plus fidèle de la protolangue est absurde, mais l'idée même que certaines langues seraient «archaïques» ou «primitives» par rapport au reste de leur famille ne l'est guère moins: même si on est amené à juger, après s'être fait une idée assez précise de la protolangue, qu'une langue présente plus d'«archaïsmes» qu'une autre et se révèle plus «conservatrice» sur certains points, on ne doit pas oublier qu'elle a le même âge que les autres, c'est-à-dire qu'elle est éloignée du moment où elle s'est séparée des autres, par le même espace de temps que les autres se sont séparées d'elle; son «archaïsme» ou son «conservatisme» dans un domaine ne peut faire préjuger en rien de ce qu'elle est dans d'autres domaines où elle peut tout aussi bien avoir été innovante; elle ne peut être revêtue d'aucun cachet global d'«archaïsme» qui justifierait par exemple qu'on considère comme archaïsme tout trait qui la distingue des autres langues⁴⁵.

Enfin, même s'il se vérifiait que les langues de Formose se sont débranchées de bonne heure du tronc commun, cela ne prouve pas qu'à cette époque elles aient été parlées à Formose, et que, par exemple, leurs locuteurs n'aient pas été repoussés sur ce laps de temps fort long, progressivement, jusqu'aux confins où l'on en constate la présence aujourd'hui.

45. On a déjà rencontré un effet pervers de cette idée fausse dans la structuration même de l'article **betas* du dictionnaire de Blust. En effet, l'entrée **betas* est notée sous cette forme parce que l'amis a un reflet de **betas* et non de **beRtas*; cette entrée est étiquetée (1) à l'intérieur de laquelle **be(R)tas* constitue une sous-entrée étiquetée (2); or, une fois que l'on a choisi d'écrire **be(R)tas* et de ne pas séparer **betas* d'un côté et **beRtas* de l'autre, il n'y avait aucune raison de ne pas choisir **be(R)tas* comme entrée principale, l'amis n'étant que l'illustration de la possibilité **betas* subsumée sous **be(R)tas*; si, au contraire, on refuse les écritures à parenthèses, ce qui, à nos yeux, est en tout point préférable — et évite, dans le cas présent, de supposer que l'amis a connu **-R-* —, sous **betas* on devrait avoir non seulement l'amis, mais toutes les langues dont les reflets excluent **beRtas*. Au passage, cet exemple fournit une nouvelle illustration du caractère ambigu de parenthèses qui couvrent à la fois les reflets ne permettant pas d'opter — seul cas auquel elles devraient être réservées — et un ensemble indifférencié de langues à reflet de **beRtas* et de langues à reflet de **betas*; cela supposerait évidemment qu'on soit au clair, dans le cas présent, sur les langues où l'on affirme que **Rt* > /t/ et non /rt/ ou /gt/, c'est-à-dire qu'on perçoive l'enjeu de ce **-R-*.

2.4. *La théorie des migrations introduit des contraintes indues sur les hypothèses de reconstruction: l'exemple de l'article *(k)u*

La théorie des migrations a pour effet de formater d'avance ou d'imposer un filtre sélectionnant a priori les hypothèses de reconstruction possibles, au lieu que des faits possiblement apparentés attestés dans un ensemble de langues devraient amener à proposer des arbres généalogiques (ou d'autres dispositifs) fondés uniquement sur des faits linguistiques.

Ainsi l'existence d'un article *(k)u* dans quelques langues de Formose et d'un article *k(o)* dans des langues de Fidji ne donne lieu, chez Geraghty⁴⁶ par exemple, à aucune proposition d'hypothèse tendant à les rapprocher, alors que **k* > /k/ aussi bien dans les premières que dans les secondes et que PAN **u* > /u/ dans les unes et /o/ dans les autres, et que, par ailleurs, des articles en *o* sont attestés dans des langues extérieures aussi bien à Formose qu'à Fidji et que l'élément *k-* est également largement attesté à travers toute la famille.

Le tableau suivant présente quelques unes de ces formes; il se divise en trois parties. Dans la première (A), nous rassemblons les formes de l'article **u* ou **ku*, aux différents cas⁴⁷, dans un échantillon de langues qui en présentent des reflets: on a *u* ou *o* en amis, yami, ivatan, *ku* en atayal mayrinax; le paiwan montre les effets de la concurrence de l'article *a*: *u* n'y est plus séparable des marques de cas *n-* et *t-*; pour une partie d'entre elles nous donnons aussi les formes correspondantes de l'article des noms propres, et, pour les autres, où il n'y a pas d'amalgame entre article et marque de cas, il s'agit soit de la forme unique (mori, wolio), soit de formes qui sont isolées au milieu du paradigme et qui semblent résiduelles (*nu* à interpréter comme un ancien **n-u* selon nous, en tukang besi, karo batac, chamorro)⁴⁸; dans toutes ces langues, les formes issues de **u* fournissent les articles des noms communs. Dans la seconde partie de ce tableau (B), nous passons en revue une série de langues où un article a servi à renforcer les marques personnelles pour former des personnels lourds (emphatiques ou personnels indépendants): dans la plupart des langues, c'est l'article des noms propres qui joue ce rôle; dans d'autres, c'est l'article des noms communs; dans d'autres encore, c'est une marque manifestement identique à une forme d'article attestée dans des langues proches, mais qui ne fonctionne plus comme telle dans la langue considérée (cette situation est signalée par un * dans le tableau). Dans la troisième partie du tableau (C), nous donnons, pour deux langues de Fidji, le paradigme des marques

46. Geraghty a consacré une monographie à *The history of the Fijian Languages* (1983) où il traite abondamment des faits dont il va être question. Nous lui empruntons tous nos exemples.

47. Nous utiliserons les abréviations suivantes: Préd = forme spéciale de l'article pour les prédicats nominaux, Suj = cas sujet, Gén = cas génitif-complément d'agent, Obj = cas objet, Loc = cas locatif; NP = article des noms propres, NC = article des noms communs; incl = (1ère pers. duel ou triel ou pluriel) inclusif, du = duel, tri = triel ou paucal.

48. Sur le problème des articles-marques de cas de l'austro-nésien, cf. Lemaréchal 2004.

personnelles où intervient la forme de l'article (k)o, par ailleurs article des noms propres:

A

		mayrinax		saisiyat	paiwan		amis	yami	ivatan
		déf	/indéf						
NC	Préd						<i>u</i> ⁴⁹		
	Suj	<i>ku'</i>	<i>a'</i>	(<i>ka</i>)	<i>a</i>		<i>k-u</i>	<i>u</i>	' <i>o</i>
	Gén	<i>n-ku'</i>	<i>n-a'</i>	<i>no(ka)</i>	<i>nu-a</i>		<i>n-u</i>	<i>n-u</i>	<i>n-o</i>
	Obj	<i>c-ku'</i>	<i>cu'</i>	<i>ka</i>	<i>tu-a</i>		<i>t-u</i>	<i>s-u</i>	<i>s-o</i>
	Loc	<i>c-ku'</i>	<i>i'</i>	<i>ray (V?)</i>	<i>(i) tu-a</i>		<i>i t-u</i>	<i>d-u</i>	<i>d-o</i>
NP	Suj	' <i>i</i> '		(<i>hi</i>)	sg	pl	sg	pl	sg /pl
	Gén	<i>n-i'</i>		<i>n-i</i>	<i>ti</i>	<i>ti-a</i>	<i>ci</i>	<i>c-a</i>	<i>si si sa</i>
	Obj	' <i>i</i> '		<i>hi</i>	<i>n-i</i>	<i>ni-a</i>	<i>n-i</i>	<i>n-a</i>	<i>n-i da</i>
	Loc	<i>k-i'</i>		<i>kan</i>	<i>tjai</i>	<i>tjai-a</i>	<i>ci X-an</i>	<i>ca X-an</i>	<i>d-i da</i>
							<i>i ci-X-an</i>	<i>i ca X-an</i>	<i>d-i da</i>

		mori	wolio	tukang besi	karo batak	toba batak	chamorro
NC	Suj	<i>o/∅</i>	<i>o</i>	<i>na</i> (nom.)			<i>i</i>
	Gén		<i>-na + Poss</i>	<i>nu/-u</i> (gén.)	<i>ni/nu/u/∅</i>	<i>ni</i>	<i>ni/nu</i>
	Obj			<i>te</i> (accus.)		<i>∅/di</i>	<i>ni</i>
	Loc	<i>aN-/i-/ndi-</i>	<i>i</i>	<i>i/di</i>	<i>i, ku</i>	<i>di, tu, tijan</i>	<i>gi</i>

B

		rukai	tsou	mayrinax	paiwan	kavalan	seediq	amis
sg	1	<i>ku-n-aku</i>	<i>na a'o</i>	' <i>ikuing</i>	<i>ti-aken</i>	<i>a-iku</i>	<i>yaku</i>	<i>k-aku</i>
	2	<i>ku-su</i>	<i>na suu</i>	(' <i>i</i> ') <i>'isu'</i>	<i>ti-sun</i>	<i>a-isu</i>	<i>isu</i>	<i>k-isu</i>
	3	<i>ku-a-ni</i>	<i>na ic'o</i>	' <i>ihiya</i>	<i>ti-madju</i>	<i>a-izipna</i>	<i>heya</i>	<i>cingra</i>
	ind	<i>ku-a-Da</i>						
incl		<i>ku-ta</i>	<i>na a'ati</i>	(' <i>i</i> ') <i>'ita'</i>	<i>ti-tjen</i>	<i>a-ita</i>	??	<i>k-ita</i>
pl	1	<i>ku-nai</i>	<i>na a'ami</i>	' <i>icami</i>	<i>ti-amen</i>	<i>a-imi</i>	<i>yami</i>	<i>k-ami</i>
	2	<i>ku-numi</i>	<i>na muu</i>	' <i>icimu</i>	<i>ti-mun</i>	<i>a-imu</i>	<i>yamu</i>	<i>k-amu</i>
	3	<i>ku-l-i-ni</i>	<i>na xin'i</i>	' <i>inha'</i>	<i>ti-a-madju</i>	<i>qaniau</i>	<i>deheya</i>	<i>cangra</i>
		<i>ku-l-i-Da</i>	<i>na xe</i>					
article		<i>ku</i> NP	<i>na</i> indéf.	' <i>i</i> ' NP	<i>ti</i> NP	<i>a</i> NC (sf 3pl)	* <i>i</i> NP (sf 3)	<i>k-</i> NC (sf 3)

49. Seul l'amis a une forme spéciale pour les noms prédicats («neutral» chez Wu 1995, ici ArtPréd) distincte de celle du cas sujet:
u amis kaku «I am Amis» (p. 11)
ArtPréd Amis 1sgSuj
ma-ala ni aki ku paysu nu singsi
PF prendre ArtNPGén/agt NP ArtSuj argent ArtNCGén/Agt professeur
«Aki took the teacher's money» (p. 16)

	ilocano	pangasinan	limos k.	mori	wolio	tukang	vx bugis
sg 1	<i>siak</i>	<i>siak</i>	<i>sakon</i>	<i>ongkue</i>	<i>iaku</i>	<i>iaku</i>	<i>ia'</i>
2	<i>sika</i>	<i>sika</i>	<i>sika</i>	<i>omue</i>	<i>ingkoo</i>	<i>iko'o</i>	<i>iko, io</i>
3	<i>isu(na)</i>	<i>sikato</i>	<i>siya</i>	<i>onae</i>	<i>incia</i>	<i>ia</i>	<i>ia</i>
incl du	<i>data</i>	<i>sikata</i>	<i>dita</i>				
pl	<i>datayo</i>	<i>sikatayo</i>	<i>ditayu</i>	<i>ontae</i>	<i>ingkita</i>	<i>ikita</i>	<i>idi'</i>
pl 1	<i>dakami</i>	<i>sikami</i>	<i>dikami</i>	<i>omami</i>	<i>ingkami</i>	<i>ikami</i>	<i>(id)ikk n</i>
2	<i>dakayo</i>	<i>sikayo</i>	<i>dikayu</i>	<i>omiu</i>	<i>ingkomiu</i>	<i>ikomiu</i>	<i>iko, io</i>
3	<i>isuda</i>	<i>sikara</i>	<i>dida</i>	<i>onda</i>	<i>(manga)</i> <i>+ incia</i>	<i>amai</i>	<i>ia</i>
article	<i>ni</i> NP <i>ti</i> NC	<i>si</i> NP	<i>si</i> NP	<i>o</i> NC	<i>o</i> Art <i>*ing?</i>	<i>na</i> Art <i>*i</i>	<i>-e(ng)</i> <i>*i</i>

C

	tokaimalo ⁵⁰		nadrau ⁵¹	
sg 1	<i>(k)o yau</i>	<i>(*k-)u i-au)</i>	<i>(k)o yau</i>	<i>(*k-)u i-au < *i-aku)</i>
2	<i>(k)o iko</i>	<i>(*k-)u ikau)</i>	<i>(k)o keko</i>	<i>(*k-)u k-ikaw: /e/?)</i>
3	<i>(k)o koya</i>	<i>(*k-)u k-u -ia)</i>	<i>i kwaya</i>	<i>(*i k-u -a??-ia)</i>
du incl	<i>kita-ru</i>	<i>(*kita-+«2»)</i>	<i>kitaru</i>	<i>(*kita-+«2»)</i>
1	<i>o kei-rau</i>	<i>(*u kai-+«2»)</i>	<i>kai-rau</i>	<i>(*kai-+«2»)</i>
2	<i>(k)o drau</i>	<i>(*k-)u +«2»)</i>	<i>(k)o drau</i>	<i>(*k-)u +«2»)</i>
3	<i>ko rau</i>	<i>(*k-u +«2»)</i>	<i>i kirau</i>	<i>(*i k-i-+«2»)</i>
tri incl	<i>kitou</i>	<i>(*kita-+«3» + hapl)</i>	<i>i kita-tou</i>	<i>(*i kita-+«3»)</i>
1	<i>o kei-tou</i>	<i>(*u kai-+«3»)</i>	<i>kai-tou</i>	<i>(*kai-+«3»)</i>
2	<i>(k)o dou</i>	<i>(*k-)u +«3»)</i>	<i>(k)o dou</i>	<i>(*k-)u +«3»)</i>
3	<i>ko ra-tou</i>	<i>(*k-u +«3»)</i>	<i>i kira-tou</i>	<i>(*i k-i-da-+«3»)</i>
pl incl	<i>kikita</i>	<i>(*k-u kita)</i>	<i>i kita</i>	<i>(*i kita)</i>
pl 1	<i>o kei-mamu</i>	<i>(*u kai-mami)</i>	<i>kai-mamu</i>	<i>(*kai-mami)</i>
2	<i>ko-ni</i>	<i>(*k-u niu < n-i(S)u)</i>	<i>(k)o niu</i>	<i>(*k-)u n-i(S)u)</i>
3	<i>kera</i>	<i>(*k-[i[da]: /e/?)</i>	<i>(i) kira</i>	<i>(*k-[i-[da])</i>

50. Les étymologies proposées entre parenthèses demanderaient des démonstrations qui ne peuvent être développées ici; nous nous limiterons aux remarques suivantes.

yau est partagé par un grand nombre de langues où la disparition du **k* ne peut être phonétique; ce *yau* provient pourtant certainement de **i-aku*, forme renforcée par **i*. Ce n'est pas le seul cas où l'on est devant deux âges successifs de renforcement: par **i* pour *yau*, *iko* (< **ikau*), *ia* (et, en fait, *isu* en mayrinax, kavalan, seediq, cf. Lemaréchal 2003), puis par **k-* pour *kita*; *ikau* est très largement attesté (*ikami*, *ikamu*, etc. existent aussi ailleurs); *k-* est présent dans **(i)kaw* 2sg, *k-ami*, *k-amu*, *k-ita* qui ont les attestations les plus larges, mais rentrent dans un système transparent en amis — réfection standardisante ou ancien système? —: cela pose le problème de la valeur de *k-*, marque d'émphatisation du sujet vraisemblablement.

Dans les deux langues de Fidji, les formes de duel et triel sont des composés où entrent les mots «2» et «3» moyennant une variation morphophonologique *drau* et *dou* vs *rau* et *tou*.

51. Mêmes remarques que pour les étymologies des formes du tokaimalo.

On note à côté des formes renforcées au moyen de *(k)o*, des formes renforcées par *i*, à rapprocher de l'article des NP; la distribution entre *(k)o* et *i* est la suivante: *i* est réservé aux 3ème pers. (la plus «nominale» des personnes), mais apparaît aussi devant *kita* (influence du /i/ de *kita*?).

L'amis, langue de Formose, et deux langues hors Formose, le yami⁵² et l'ivatan, présentent un article des NCs *u* ou *o* (< **u*) qui se combine avec les marques de cas. L'atayal mayrinax, langue de Formose, présente un article *ku*, sans nul doute comparable à la forme sujet *k-u* de l'amis, langue où *k-* est la marque du cas sujet⁵³ (commutant avec *n-* et *t-*); *ku* est réinterprété en mayrinax comme base⁵⁴, qui se combine elle-même avec les marques de cas (*n-* et *c-* dans cette langue), ceci dans une langue où la forme sujet est caractérisée par l'absence de marque de cas. C'est *o* (< **u*) que l'on trouve en mori et wolio, tandis que *nu* génitif-agent opaque se trouve inséré dans des paradigmes sans *-u* en chamorro et dans les langues de Sumatra prises en exemple⁵⁵. L'emploi de **u* ou **ku* comme renforcement de personnels sujets lourds n'est pas inconnu, à Formose et hors Formose, selon un procédé d'engendrement de formes lourdes mettant en œuvre aussi bien d'autres articles, de noms propres (*i* ou *ti/si*), mais aussi de noms communs (*a*, *o*, *e(ng)*, *na*) et même d'articles indéfinis dans les langues qui présentent cette catégorie (tsou *na*). Enfin, dans des langues de Fidji comme le tokaimalo et le nadrau, *o* et *ko*, présentés comme des variantes plus ou moins libres, fonctionnent comme articles des noms propres, mais apparaissent aussi dans une partie des personnels.

Comme on le voit, on trouve un (*k*)*o* devant les noms, mais aussi devant les personnels indépendants lourds dans un certain nombre de langues

Dans les deux langues, *kei-* est à rapprocher du *kai* (à côté de *kami*) attesté en uma et dans d'autres langues proches (sans doute aussi dans le tondano *key*); *mami* est la forme de possessif-agent dans un groupe de langues comprenant le vieux javanais, le mori, le wolio, le tukang besi, mais aussi en chamorro, woleai, palau, etc. (sur le *m-* de *mami* et d'autres marques personnelles, cf. Lemaréchal 2003). L'étude du développement du système des marques personnelles dans un échantillon de langues (semblable à celui utilisé dans Lemaréchal 2004 pour les articles) a fait l'objet d'un travail non encore publié dans sa totalité, d'où a été extraite la notule de 2003.

Outre ces articles et ces formes particulières de marques personnelles, un certain nombre de rencontres poussent à voir dans le POC une langue fort proche du groupe des langues qui viennent d'être citées: voir, en particulier, le couple de marques de «focus» *-i* à valeur de RF-LF et de PF de l'objet partiellement affecté et **-ak^on* à valeur d'IF-BF et de PF de l'objet déplacé, anciennes marques de ces focus à l'IRREALIS qui ont renouvelé les marques REALIS *-an* et *i-* (sur ce dernier point, cf. Lemaréchal 2001).

52. Abusivement recruté du côté des langues de Formose par Starosta par exemple et dès les auteurs japonais de l'entre-deux-guerres, le yami parlé dans l'île de Botel Tobago est, du point de vue du signifiant des morphèmes grammaticaux et des systèmes qu'ils forment, beaucoup plus proche des langues des Philippines que de celles de Formose. Dans la classification de Grimes et al. (in Tryon 1995, I, p. 278), il est d'ailleurs classé: «Northern Philippine, Bashiic-Central Luzon-Northern Mindanao, Bashiic».

53. Fonctionnant aussi avec les prédicats «définis» des propositions équatives.

54. Sauf cas où cela est imposé par une évidence précise, nous privilégions toujours, au moins en première hypothèse, amalgame et univerbation, et excluons les fausses coupes qu'on se trouve contraint de multiplier dès qu'on se laisse enfermer par une conception géographique de l'«archaïsme» des langues de Formose; cf. Lemaréchal 1998b.

55. Une fois de plus, ces langues gardent trace de la forme primitive simple, en l'occurrence **u*, et non de la réfection, en l'occurrence **ku*.

d'Océanie et un *ku* est, en amis, langue de Formose, le nominatif de l'article qui figure non seulement devant les noms mais aussi devant les personnels indépendants. Il est évident que, s'il n'y avait pas de théorie des migrations, la première idée qui viendrait à l'esprit serait d'établir un rapprochement, qu'il en tire des conclusions sur le «subgrouping» des langues si ce rapprochement s'accompagne d'autres plus ou moins parallèles. Si, au contraire, il y a eu une migration Formose > Philippines > Sonde > Océanie, c'est-à-dire si l'«archaïsme» des langues de Formose doit être associé à leur localisation géographique et à une origine géographique de la famille, il est difficile d'expliquer comment *o* et *na* ont pu survivre (avec *i*, **t'i*, *a(ng)*) comme article, sur un aussi long trajet entre Formose et Fidji, trajet qui n'est plus seulement une histoire. D'ailleurs, quelle protolange a-t-elle eu tous ces articles en même temps et avec quelles valeurs et pendant combien de temps tout au long de ces longs voyages? Sans cesse, l'arrière-pensée géographique introduit un biais dans les hypothèses historiques que l'on peut formuler.

L'extrapolation qui prétend faire passer d'hypothèses portant sur la chronologie relative à des hypothèses portant sur des migrations a un effet pervers sur les reconstructions elles-mêmes, par le fait qu'elle traduit du temps en espace: en effet, selon cette conception, les reconstructions ne peuvent aboutir à une histoire possible que si elles sont conformes à la géométrie imposée par des migrations; les dendrogrammes ne sont pas des représentations de l'histoire du développement de tel ou tel phénomène linguistique, mais un itinéraire sur une carte.

2.5. Conclusion

Au fond, le pire qui puisse arriver à des reconstructions fondées sur la lexicostatistique, c'est-à-dire à une comparaison de masse portant avant tout sur ce qui se ressemble, serait de prendre les emprunts pour les héritages et que les véritables héritages passent inaperçus, étant donné que les emprunts ont plus de chance de se ressembler que les héritages⁵⁶. Pour qu'une telle mésaventure survienne, il faut que les emprunts soient très nombreux, que les lois phonétiques historiques soient mal connues ou que la phonologie elle-même ait peu évolué sur de longues périodes. Il est clair que ces conditions sont réunies dans le cas des langues austronésiennes. Des emprunts ont pu se produire à date moyenne (entre 3000 et 500 avant le présent, 2500 ans sur les 6000 qu'on prête habituellement à la famille) sur de très larges portions de son domaine, entre autres à l'intérieur de cette espèce de Méditerranée⁵⁷.

56. Cf. l'exemple bien connu de v. irl. *pian*, gall. *poen*, vha *pina* (all. *Pein*), v. angl. *pin* (angl. *pine*, *pain*) tous empruntés à gr. *poina*, via lat. *poena* (Meillet, *Introduction...*, p. 378), qui se ressemblent beaucoup plus que s'ils étaient hérités directement dans chaque langue d'un dérivé de PIE **k^wei-* (en lat. on aurait **quoina* > **cuna*).

57. Pour reprendre cette expression du regretté Denys Lombard.

qui s'étend de Formose à la Malaisie entre l'Insulinde et le continent — tout autant d'ailleurs que des migrations, c'est-à-dire tout autant que des transmissions par héritage s'il n'y a pas eu de changement de langue —, à peu près à n'importe quel moment de n'importe quel point à n'importe quel autre, sans compter ce qui a pu avoir eu lieu autour de Fidji et dans de nombreuses sous-domaines plus restreints.

3. Le «grand tournant» de l'après-guerre: phonétique historique et «classes d'équivalence»

3.1. *Une méthode mécaniste de reconstruction des protosystèmes phonologiques*

D'une façon peut-être moins visible, le concept de lois phonétiques a également changé profondément de physionomie et cela constitue un ingrédient essentiel du changement de paradigme qu'a connu la linguistique comparative des langues austronésiennes au lendemain de la seconde guerre mondiale. Dans le nouveau modèle, les protophonèmes ne sont plus conçus comme des objets phonétiques ayant une substance et une véritable histoire déployée dans une diachronie sur laquelle il faut faire des hypothèses, mais comme de simples classes d'équivalence entre reflets; cela conduit à poser autant de protophonèmes distincts qu'il y a de reflets non réductibles⁵⁸.

On a ainsi posé plusieurs protophonèmes là où l'on n'en posait qu'un, dès qu'une langue nouvellement prise en compte présentait un reflet divergent, en négligeant le poids des emprunts anciens éventuels ou des variantes conditionnées apparues à date ancienne et dont les facteurs conditionnants se sont trouvés occultés entre temps. Ce qui a conduit à proposer un système phonologique typologiquement fort improvable pour le PAN⁵⁹. Certes, la typologie ne permet ni de prédire, ni d'exclure — sinon on ne risquerait guère de découvrir quelque système nouveau que ce soit —, mais elle constitue, comme on sait, un garde-fou.

Pour donner un aperçu de cette multiplication et de l'«antinaturalité» qui en résulte, mais aussi des tentatives de retour à la normale, nous reproduisons ici le tableau d'équivalences, proposé par Ross (in Tryon 1995, I,

58. On sait que cela a réussi à Saussure avec ses coefficients sonantiques.

59. C'est une conséquence mathématique du simple fait d'ajouter de nouvelles langues au corpus de départ — du moment qu'il est improbable que tous les items vérifient les correspondances déjà dégagées au moyen des autres langues —, si l'on ne se donne pas d'outils explicatifs rendant compte des divergences, autres qu'une différence d'héritage. On s'attend même à une augmentation exponentielle du nombre de tout proto-objet reconstruit (proto-phonème, étymon, marque grammaticale, etc.) au fur et à mesure qu'augmente la base documentaire. Si cela n'a pas été le cas auparavant, c'est qu'on a en fait une attitude différente à l'égard des nouveautés apportées par les langues de Formose et à l'égard de celles apportées par les autres langues.

p. 54), entre les reconstructions de Dempwolff⁶⁰, Dyen, Tsuchida, Dahl et Ross:

Dempwolff	Dyen	Tsuchida	Dahl	Ross
p	p	p	p	p
b	b	b	b	b
m	m	m	m	m
v	w	w, W	u	w
t	t	t	t ₁	t
t̥	T	T	...	...
...	C	C	t ₂	C
d̥	d	D ₃ , d-	d ₁	d ₁
		D ₂ , -d	d ₂	d ₂
		D ₁	d ₃	d ₃
d	D	D ...	...	
n	n	n	n	n
...	s ₂	θ	θ	...
t'	s ₁	s	t'	s
d'	z	z	d' ...	
...	Z	Z, d-	Z	Z
n'	ñ	ñ	n' ...	
l	l	l	l	l
...	L/N	N	l̥	L
l̥	r	r	r	r
...	S,X,x	S,X,x	S	S
j	y	y	i	y
k'	c	...	k'	...
k	k	k	k	k
g	g	g	g	g
g'	j	j	g'	j
ŋ	ŋ	ŋ	ŋ	ŋ
γ	R	R	γ	R
h	q	q,Q	q	q
...	h	h	H	h
...	?	?	...	?
i	i	i	i	i
u	u	u	u	u
ə	e	ə	ə	e
a	a	a	a	a

60. Le système de Dempwolff n'est pas exempt de faiblesses. On lui a beaucoup reproché sa base documentaire réduite, mais Dahl a bien montré que, si dans sa présentation des faits Dempwolff se limitait à trois langues, ses reconstructions se fondent sur plus de 100 langues (cf. Dahl 1976, p. 5 sqq.). Son tableau des sons de la protolanguage accuse aussi un goût de la symétrie discutable et il reconnaît lui-même que, de ce beau tableau, il est bien possible que des cases aient été vides. Mais on ne doit pas lui faire grief de ce que ses reconstructions datent d'avant l'invention de la phonologie. Enfin, il est certain qu'un poids excessif est accordé à l'«indonésien» et que cela introduit un biais considérable, mais qui était inévitable vu le rôle que cette branche a joué dans la naissance de la grammaire comparée de cette famille.

Ce tableau montre la multiplication — due à la «découverte» des langues de Formose et surtout à la méthode de Dyen, suivie et même amplifiée par Tsuchida — des protophonèmes chez ces derniers, puis la simplification progressive des protosystèmes proposée par Dahl, Ross, et d’autres.

On remarque, à partir de Dyen, 1) l’addition d’un *C distinct du *t de Dempwolff (*T chez Dyen et Tsuchida), éliminé par Dahl et Ross en même temps que le *d de Dempwolff (*D chez Dyen et Tsuchida), ce qui est revenu à éliminer les «cérébrales»⁶¹, 2) l’apparition d’un *s2 (retenu par Tsuchida et Dahl, éliminé par Ross), de *S, *X, *x (tous trois retenus et démultipliés par Tsuchida, réunis en un seul protophonème par Dahl et Ross), de *Z (associé à *d- par Tsuchida sans doute comme variante phonologisée ou non de /*d/, conservé par Dahl et Ross). Chez Tsuchida, on remarque, par rapport à Dyen, la démultiplication des symboles «d», conservée par Dahl et Ross.

3.2. Des effets pervers de la conception des protophonèmes comme simples classes d’équivalence: l’exemple de l’histoire du protophonème *S

La démultiplication des protophonèmes quand on passe du PMP au PAN est fondée sur des divergences entre reflets dans un sous-ensemble de langues de Formose, conformément à une méthode néogrammairienne apparemment irréprochable. Prenons par exemple la distinction *C vs *t (telle qu’elle est présentée par Ross, in Tryon 1995, I, p. 52):

	«ear»	«eye»	«head louse	«three»	«fresh-water eel»	«seven»
atayal (atayalic)	čañia?		kuču?	tu-yał	tuła-kiy	ma-pitu?
tsou (rukai-tsouic)		mɕō	kɕū	turu	tuŋ-roza	pitu
rukai (rukai-tsouic)	ɕaliŋa	maɕa	koɕo	toło	tola	pito
paiwan (paiwanic)	ɕaliŋa	maɕa	ɕəɕilu	cəlu	culʼa	picu
amis (paiwanic)	taŋiŋa?	mata?	kutu?	tulu	tuła	pitu
bunun (paiwanic)	taiŋa	mata?	kutu?	tau?		pitu?
tagalog (WMP)	tēŋa	mata	kūto	ta-tlo		pito
toba batak (WMP)		mata	hutu	tolu		pitu
uma (WMP)	tiliŋa	mata	kutu	tolo		pitu
manggarai (CMP)		mata	hutu	təlu	tuna	pitu
kairiru (oceanic)	tiliŋe	mata	ɕut	tuol	tun	
samoan (oceanic)	taliŋa	mata	?utu	tolu	tuna	fitu
ce qui amène à poser une opposition au niveau						
PAN	*Calíŋa	*maCá	*kúCuh	*telúh	*tuLá	*pitú
là où il n’y en a pas (c’est-à-dire plus) au niveau						
PMP	*talíŋa	*matá	*kútuh	*telúh	*tuná	*pitú

61. L’influence du sanskrit est sans doute la cause de l’apparition des «cérébrales» en javanais (avec répartition subséquente entre registres).

La seule objection qu'on puisse soulever relèverait de la naturalité plus grande d'un passage de **t* à une affriquée, type de variation qui a pu être conditionnée par la voyelle adjacente par exemple, avant d'être phonologisée à la suite de processus analogiques ou bien du fait de la présence de frontières entre morphèmes ou bien en termes d'extension progressive du changement à travers une aire, ou d'un contexte vocalique de départ (/i/) à d'autres contextes: autant de phénomènes largement attestés dans l'histoire des langues romanes, des langues slaves ou des langues bantoues⁶². De même, à propos de PAN **n* + **L* > PMP **n*, on peut noter qu'un /l/ est le produit normal de la dénasalisation de /n/ et que **n* > /l/, dans une langue comme le palau: il pourrait en être allé de même dans certaines langues de Formose⁶³. Des ensembles comme **d/*r*, **d/*l* évoquent des allophonies bien connues; de même d/ɟ, g/ɟ — c'est-à-dire les reflets de Z, d₁, d₂, d₃, (z), (D), j —; /n/ < [nj] réalisation de **n* + **i* est possible.

La reconstruction du nouveau protophonème **S*, et surtout l'histoire de cette reconstruction, montrent cependant les risques d'une application trop mécanique de la méthode. Les protophonèmes étant considérés par Dyen et, encore plus, par Tsuchida, comme des classes d'équivalence entre reflets, et la distribution des reflets de **S* étant loin d'être homogène d'un étymon à l'autre pour une même langue, on a abouti non pas à un seul protophonème **S*, mais à un nombre grandissant de protophonèmes parallèles, nombre vite devenu invraisemblable pour des sifflantes ou chuintantes, d'autant que, plus ce nombre augmentait, plus le nombre d'étymons sur lesquels étaient reconstruits les protophonèmes supplémentaires diminuait, au point d'aboutir à un protophonème différent par étymon pour 5 d'entre eux.

Ainsi Tsuchida (1976) a été conduit à poser jusqu'à neuf classes d'équivalence (nous indiquons entre parenthèses le nombre d'étymons représentatifs de chaque classe): **S₁* (19 étymons), **S₂* (1), **x₂* (1), **X* (1), **S₆* (6), **S₂* (6), **x₁* (3), **S₅* (1) et **S₄* (1), c'est-à-dire 39 étymons, cela à travers 20 langues ou dialectes (5 dialectes pour le rukai, 2 pour l'atayal). Soit:

Tableau des «Formosan **S* Dyen's and Tsuchida's symbols» (Dahl 1982, p. 28):

	<i>*S₁</i>	<i>*S₂</i>	<i>*x₂</i>	<i>*X</i>	<i>*S₆</i>	<i>*S₂</i>	<i>*x₁</i>	<i>*S₅</i>	<i>*S₄</i> ⁶⁴
Nb d'étymons	19	1	1	1	6	6	3	1	1
kanakanabu	s	-s-	s-	s	s	∅	∅		
saaroa	∅	-∅-	∅-	∅	s	∅	∅		
tsou	s	-s-	s-	s	s	∅	∅		

62. Cf. Hyman 1997, p. 163-175.
63. Un indice en est peut-être, par exemple, l'existence en paiwan des morphèmes *-al-*, *-aɬ-* (et *-ar-*), de valeur identique qui paraissent être de simples variantes anciennes (avec peut-être pour le troisième une trace de confusion, morphologique cette fois, entre **aN-* et **aR-*).
64. Il s'agit des étymons suivants, longuement discutés par Dahl (dans la notation de ce dernier):
— pour les étymons avec **S₁*, en position C₁: **S₁ag'ək* «to smell», **S₁apa* «to lay mats», **S₁ap(ə)S₁ap* «to grope in the dark», **S₁inau* «to wash clothes», **S₁ənau* «id.», **S₁ipi* «to

		<i>*S₁</i>	<i>*S₂</i>	<i>*x₂</i>	<i>*X</i>	<i>*S₆</i>	<i>*S₂</i>	<i>*x₁</i>	<i>*S₅</i>	<i>*S₄</i>
Nb d'étymons		19	1	1	1	6	6	3	1	1
rukai:	Tan ⁶⁵	s	-s-	s-	s	s		∅		
	Bud	s	-s-	s-	s	s		∅		
	Mag	s	-s-	∅-	s	s	-∅-	∅		
	Ton	s	-s-	∅-	s	s		∅		
	Man	?	-?-	∅-	?	?	-∅-	∅		
puyuma		∅	-∅-	∅-	∅	∅	-∅-	∅-	∅-	s-
thao		s	-s-	∅-	h	s	∅	∅-	∅-	s-
bunun		s	-s-	∅-	s	-s	-∅-s	s-s	∅/?	s-
paiwan		s	-s-	s-	s	s	s	s-		s-
kavalan		s	-s-	s-	s	-s	-s	∅-∅		
amis		s	-s-	s-	s	-s	s	h-	?/∅	h-
pazeh		s	-s-	s-		-s	h-h-x	h-∅-		h-
saisiyat		s	-s-	s-	s	-s	h	h	h-	
siraya		-h-x	-h-	s-	h	-h-	-∅-	∅-	∅	
atayal:	Sq	s	-s-	∅-s-	s	-s	-h-	-∅-	∅-	
	Ci	s	-s-	∅-s-	s	-s	-h-	h-		
seediq		s	-x-	s-	s	-s	-h-h	∅/x-		

Il est évident qu'on ne pouvait se permettre une telle «rigueur» que dans la mesure où l'on traitait d'un petit nombre d'attestations dans un petit nombre de langues; si on avait procédé de cette manière pour tous les reflets, éventuellement divergents, de tous les protophonèmes du PAN à travers la totalité des langues austronésiennes, la procédure aurait été intenable: c'est par centaines que l'on aurait compté les phonèmes du protosystème⁶⁶.

- dream», **S₁uyaθ* «to rinse off», avec **S₁* en position de C₂: **buS₁ək/bukəS₁* «hair of the head», **iS₁əq* «urine», **luS₁əq₁* «tears», **niS₁awa* «breath», **Q₂aS₁əluH₂* «pestle», **t₁aS₁aN* «village», et avec **S₁* en position de C₃: **t₂aliS₁* «rope», **t₂iŋaS₁* «food particles caught between the teeth», **(dD)aqeiS₁* «forehead» (Dyen 1965, 1971), **t₂aS₁iq/t₂aq₃iS₁* «to sew» (Dyen), **t₁apəS₁* «to winnow» (Dyen), **t₁apuS₁* «id.» (Dyen), **t₂əbuS₁* «sugar-cane» (Dyen), **t₁uqaS₁* «elder sibling» (Dyen);
- pour les étymons avec **S₃*: **DewS₂a* «2» (Dyen) remplacé par **d₃uS₃a* (Dahl);
 - pour les étymons en **x₂*: **x₂epat* «4»;
 - pour les étymons en **X*: **iXu* «thy, thee» remplacé par **iS₁u*⁶⁶
 - pour les étymons en **S₆* (Tsuchida⁶⁶), en C₁: **S₆i(m)pa(rR)* «to walk across the river», **S₆ikam* «sort of mat», en C3: **t₂aq₃iS₆/t₂aS₆iq₃* «to sew», **paliS₆i* «taboo», **q_aamiS₆* «winter», **Raus₆* «scoop with a ladle»;
 - pour les étymons en **S₂*: **kaS₂iw* «tree, wood» (Dyen), **kuS₂kuS₂* «fingernail, claw» (Dyen), **S₂uni* «to chirp» (Tsuchida), **guS₂am* «sickness» (Tsuchida), **q₄uS₂u* «mushroom» (Tsuchida), **t₂uS₂u* «to thread a needle» (Tsuchida);
 - pour les étymons en **x₁*: **x₁apuy* «fire», **Nix₁əpis* «thin» (Tsuchida), **x₁əθi* «content», «meat, flesh» (Tsuchida);
 - pour les étymons en **S₅*: **S₅eyup* «to blow»;
 - pour les étymons en **S₄*: **S₄iRup* «to sip»;

65. Ces abréviations représentent différents dialectes.
66. Comme on le verra, c'est un peu ce qu'on fait pour le lexique, avec la même conséquence: la multiplication des doublets phonologiques (opposant des **k* vs **q*, des voyelles

A la suite, entre autres, de Dahl, on a réduit progressivement ce nombre tout à fait irréaliste de protophonèmes jusqu'à n'en retenir qu'un seul, en éliminant surtout ceux pour lesquels une seule langue présentait un traitement aberrant par rapport aux autres variantes de **S*, et/ou en exploitant l'idée que le passage **S* > **h* > $\emptyset$ pouvait ne s'être étendu que progressivement à tout le lexique concerné. Il n'en demeure pas moins que, même si on réduit les neuf protophonèmes de Dyen-Tsuchida à un plus petit nombre et, en fin de compte, à un seul **S*, **S* n'a de reflet dans une langue comme le puyuma que dans un seul mot (d'où le protophonème **S₄* de Tsushida) où ce /s/ reflet de **S* fait figure d'exception; en saaroa, **S* n'a de reflet que pour six étymons sur 39; et les étymons contenant **S₂*, **x₁* et **S₅* chez Dyen-Tsuchida, soit dix étymons, n'ont de reflet de ce **S* que dans 8 ou 10 des vingt langues et dialectes. Seuls les étymons ayant un **S₁*, un **S₃* ou un **X* chez Dyen-Tsuchida (c'est-à-dire 27 étymons sur les 39) ont des reflets dans presque toutes les langues (sauf saaroa et puyuma)⁶⁷.

**S* pose encore un autre problème quand on passe des langues de Formose aux langues malayopolynésiennes: en effet, parmi les langues malayopolynésiennes, certaines présentent un reflet /h/ de **S*, mais il se trouve que, même dans ces langues, une partie des étymons reconstruits avec un **S* en PAN ont des reflets sans /h/; or, il s'agit d'étymons très largement attestés comme **Sepat* «quatre» ou **Si-* la marque d'IF.

On peut se demander si le **S* de tous les étymons où l'on en reconstruit un a bien une origine unique, sans pour autant accepter la solution d'en reconstruire plusieurs. En effet, il n'est pas impossible que des problèmes de morphologie interfèrent avec les problèmes de phonétique historique que nous venons de voir, puisque **S* apparaît dans des marques grammaticales comme **Su* marque de 2ème pers. ou comme **Si-/*Sa-* marque d'IF-BF et PF de l'objet déplacé, ou dans des élargissements radicaux *C₁V₁-* qui sont peut-être d'anciens préfixes⁶⁸.

4. Hypothèses sur le protolexique et programmes de recherche

Le problème majeur de la reconstruction d'un lexique PAN ou PMP, ou de toute protolangue intermédiaire, reste pour nous celui posé par le grand nombre de doublets morphologiques.

différentes) aussi bien au niveau des étymons en *C₁V₁C₂V₂C₃* qu'à celui des racines en *C₂V₂C₃* (opposant des **k* vs **q*, des voyelles différentes, comme **tas* et **tes*).

67. Un certain nombre des étymons invoqués par Dyen, Tsuchida et surtout Blust ne sont guère solides, n'étant appuyés que sur une seule attestation à Formose, quand ce n'est pas une à Formose et une seule également hors Formose.

68. En attendant la parution de notre étude en préparation, on se reportera, pour **Su*, à Lemaréchal 2003, où nous défendons l'idée qu'un **u* (forme lourde **i-u*) de valeur «2ème pers.» est à poser dans les **u*, **n-u*, **m-u*, **i-u*, **n-i-u* présents dans les marques de 2sg et de 2pl, sujet et possessif/ complément d'agent singulier ou pluriel de l'ensemble de la

4.1. *N'importe quelle séquence */CV(N)/ peut constituer un «élargissement radical»*

La racine **tas* est associée avec les quinze élargissements radicaux suivants: **be-*, **beN-*, **beR-*, **bi-*, **bu-*, **e-*, **qaN-*, **qu-*, **ra-*, **raN-*, **ri-*, **riN-*, **ru-*, **Ra-*, **te-*, auxquels il faut ajouter, si l'on pense qu'une partie des /N/ apparaissant dans les reflets mais non répercutés par Blust dans les étymons sont, malgré cela, possiblement hérités d'un **N*: **biN-* (TB, MAR), **buN-* (IBAN, SND), **eN-* (SAS), **quN-* (BAR), **ruN-* (MAL).

Combien existe-t-il de C₁V₁- différents à travers la totalité du lexique des étymons tel qu'il est reconstruit dans un dictionnaire étymologique comme celui de Blust? Comme on peut en juger par le tableau suivant, on s'aperçoit que presque toutes les combinaisons /CV/ possibles sont attestées:

Nombre d'étymons attestés avec les différents «élargissements radicaux» préfixés:

avec	/N/ = a			/N/ = e			V/ = i			/N/ = u			toutes /V/			MAL
	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	
*bV(N/R)-	non encore comptés															
*pV(N/R)-	non encore comptés															
*wV(N/R)-	11			0			3			0			14			
*V(N/R)- ⁶⁹	143			94			112			107			456			
*qV(N/R)-	88			49			45			49			232			
*hV(N/R)-	46			18			31			22			117			
*SV(N/R)-	15			12			12			12			51			
*RV(N/R)- ⁷⁰	20	7	7	10	7	2	7	4	0	4	3	1	42	21	10	10
*rV(N/R)- ⁷¹	20	8	2	27	2	5	13	6	5	13	6	4	73	31	16	34

famille, y compris dans **(i)ka(w) < *ka-u*: **Su* pourrait être issu d'un **/yu/* par fortition et dévoisement, lui-même issu d'un **i-u* réinterprété comme monomorphématique. Pour le reste, cf. Lemaréchal 2001, p. 471-472 et notes 106 à 108.

69. Les effectifs de **V-* sont sans doute à augmenter d'un certain nombre de reflets sans C₁ (en JAV, KAY, etc.) inventoriés, au nom de troncations arbitraires, sous des étymons en **RV(N)-*. Nous pensons avoir montré qu'une bonne morphématique amène à poser des préfixes issus de **aN-* et **aR-* dans les langues des Philippines elles-mêmes même s'ils y sont toujours précédés (mais non en ILK) des préfixes **m-*, **n-*, **p-*, et que ces mêmes préfixes issus de **aN-* et de **aR-* apparaissent sans **m-*, ni **n-*, ni **p-* initiaux dans des langues comme le bugis (et l'ilocano). Jusqu'à preuve du contraire, il faut supposer que ces langues n'ont pas connu de renforcement par **aR-* de leur forme en **aN-* ou en **V-* (***R-aN-* ou ***RV-*).

70. Dans un certain nombre d'étymons en **CVN-*, le **/N/* a entraîné, comme cela est bien attesté avec *paN-*, une assimilation totale de C₂ — sans doute faut-il par conséquent poser 2 **-VN-* distincts assimilant vs non assimilant comme on pose deux préfixes *paN-* —; cette assimilation peut passer inaperçue et amener à reconstruire des racines à nasale initiale fantômes: il s'agit, pour ce qui est des étymons en **RVN-*, de **Rames* (< **pes*), **remek* (< **pek*), etc.

71. On constate que MAL a des reflets pour un grand nombre d'étymons en **rV-*, (34), alors que l'on n'en a que 10 sous **RV-*; en fait, en MAL, où **R > /r/*, on a des reflets en */r/* dans les deux séries: toutes les conditions sont donc réunies pour que le fait de privilégier systématiquement l'étymologie des */r/* du MAL par **r* crée autant d'étymons fantômes. Le malais

avec	/V/ = a			/V/ = e			V/ = i			/V/ = u			toutes /V/			MAL
	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	Tot	<2	<3	
*gV(N/R)-	11	6	3	6	3	1	6	3	1	6	4	0	29	16	5	15
*kV(N/R)-	38	13	7	46	19	8	31	12	9	40	18	7	155	63	31	51
*tV(N/R)- ⁷²	43	21	4	41	18	11	24	10	6	39	12	9	147	61	30	20
*CV(N/R)- ⁷³	4	0	0	6	2	0	1	1	0	3	0	1	14	5	0	1
*cV(N/R)-	3	0	3	12	5	3	6	3	2	7	2	2	28	12	6	17
*sV(N/R)-	49	13	20	33	10	11	48	13	11	29	13	5	158	49	37	40
*lV(N/R)-	56	12	10	49	13	4	46	13	15	46	16	13	197	54	42	63
*dV(N/R)-	24	6	5	25	10	7	12	5	2	12	5	5	73	27	19	20
*zV(N/R)-	9	2	3	5	0	3	4	1	1	1	1	0	19	4	7	7

Note: Les élargissements radicaux sont classés dans ce tableau, en ordonnée, par leur C_1 et, en abscisse, par leur voyelle. Les colonnes sont réparties en 6 groupes: le premier ne contient qu'une colonne où sont listés les élargissements prefixaux compte non tenu de leur voyelle; les 4 groupes de colonnes suivants indiquent le nombre d'étymons recensés pour chaque voyelle; chacun de ces 4 groupes contient 3 colonnes: la 1ère indique le nombre total d'étymons, la 2ème le nombre d'étymons reconstruits à partir de deux langues seulement (343 sur 935, compte non tenu des étymons en *bV(N/R)- et *pV(N/R)-), la 3ème le nombre d'étymons reconstruits à partir de trois langues seulement (203 sur 935): on voit par là que les trois cinquièmes des étymons reconstruits par Blust le sont à partir de 2 ou 3 langues (546 sur 935); le 5ème groupe ne contient qu'une colonne: y figure le nombre d'étymons reflétés en malais pour la totalité d'une ligne, c'est-à-dire pour un *CV(N/R)- donné, quelle que soit *V: le malais est sans aucun doute un des donneurs-receveurs d'emprunts majeurs et se trouve ainsi au centre de la question.

Ces différents comptages — faits sous toute réserve, et qui comportent nécessairement des choix qui ne peuvent qu'être subjectifs — ont laissé systématiquement de côté les noms de parties du corps, d'espèces, de parenté — qui sont par définition sans étymologie sauf hypothèse sémantique supplémentaire — et les déonomatopéiques restés tels⁷⁴, et ne prennent en compte que les étymons où $C_1V_1(N/R)$ - est

a toute chance d'être la source, via l'emprunt, de tous ces reflets en /rV-/ dans des langues où *R > /g/ ou autre chose.

72. MAL est particulièrement peu représenté pour les reflets d'étymons en *tV- — avec 2 reflets de *ta-, 6 de *te-, 4 de *ti- et 8 de tu-, c'est-à-dire 20 sur 147 —, alors que cette langue est, comme on le sait et comme c'est attendu quand on connaît l'histoire de cette aire, un grand donneur, et receveur, d'emprunts, bien plus que ne pourraient le faire croire les rares cas évoqués par Blust dans ses notes, au point que nous pensons — même en dehors de toute hypothèse faisant de l'emprunt un facteur épistémologiquement «catastrophique» pour la grammaire comparée-reconstruction des langues austronésiennes —, que les emprunts au et du MAL (ou plutôt une protolangue intermédiaire dont le malais est issu et qui en était fort proche au moins sur ce point) constituent un trait essentiel du faciès du lexique reconstruit. L'explication est peut-être que *ter*- est vivant en malais, et que *ter- ne pouvait devenir un élargissement radical opaque que dans des langues où il n'est plus vivant, ou dans des emprunts où il n'était plus senti comme vivant.

73. Etant donné la méthode adoptée, on n'a par définition d'étymons en *C (*t^r) que s'il existe des reflets dans une des langues de Formose distinguant les reflets de *t et de *C.

74. Nous n'avons pas la place ici de traiter du «déonomatopéisme généralisé», très à la mode (voir Bohas et al. pour les langues sémitiques), qui nous ferait toucher les origines du

réellement isolable par commutation. Sont évidemment exclus les affixes et mots grammaticaux. Pour les $*C_1V_1-$ où $*C_1 = *b$ et $*p$, les comptages ne sont pas encore terminés, on se rappellera que les étymons avec $*b$ initial représentent un tiers de la totalité du dictionnaire.

Si presque toutes les combinaisons de C et V possibles selon le proto-système phonologique de l'austro-nésien commun sont attestées, elles sont loin d'être représentées en nombre égal⁷⁵. Il nous semble qu'un certain nombre d'entre elles ne sont que des variantes d'autres.

4.2. Les élargissements radicaux comme anciens affixes démotivés

Une des explications possibles de l'existence de ces «élargissements radicaux préfixés» (ou infixés), est qu'au moins une partie d'entre eux soit d'anciens affixes intégrés à la base, phénomène qui serait sinon la cause, du moins partie prenante d'une dissyllabisation générale des bases⁷⁶, avec opacification et figement.

Ce qui pousse le plus dans ce sens est la présence massive d'étymons en $*bV-$ (ou, moins fréquemment, $*pV-$), $*bVN-$ (ou $*pVN$) et, beaucoup plus rarement, $*bVR-$ (ou $*pVR-$) — on a déjà rencontré $*beRtas$ à propos de $*tas$. L'élargissement $*bVR-$ évoque immédiatement le préfixe *ber-* des langues malayo-javaniques, qui est une des issues de $*paR$ -⁷⁷, parallèle au *pag-* des langues des Philippines comme le tagalog; or, dans ces langues,

langage et des langues (il y a loin entre -110 000 et -6000!), et qui se manifeste dans Blust 1982, Zorc 1990, et un peu partout, selon lequel l'origine des racines serait déonomatopéique: cela n'a évidemment aucun intérêt pour une grammaire comparée-reconstruction historique, puisque, par définition, il est impossible de remonter au delà de l'instant où un étymon ou une racine déonomatopéique a cessé de l'être, la remotivation permanente des déonomatopéiques les empêchant d'évoluer selon les lois phonétiques reconnues dans le reste du lexique, excluant donc toute évolution, c'est-à-dire toute temporalité. Pour une fois que le lexique n'est pas ce qu'il y a de plus instable (Meillet), il est trop stable pour être utilisable. Aussi considérons-nous comme (dé)onomatopéique seulement un signe linguistique censé représenter le bruit qu'il mime, et cela seulement: un /tok/ qui signifie au plus «faire /tok/» et non pas un /tok/ signifiant «frapper qqch ou qqn en faisant *tok*» ou bien «tomber en faisant *tok*» ou bien «jouer d'un instrument en produisant un son représentable par *tok*», qui seront considérés comme des hyponymes de «frapper», «tomber» et «jouer d'un instrument».

75. Blust s'est longuement attaché au problème dans son ouvrage sur la théorie de la racine, en en tirant semble-t-il un argument supplémentaire en faveur de la thèse selon laquelle la racine n'a aucune réalité pour les locuteurs dans la synchronie actuelle — ce qui est sûr, mais la question est-elle là? —, mais que, de ce fait, c'est l'étymon qui est l'unité lexématique actuelle: les deux n'ont rien à voir et cela n'empêche pas, à notre avis, que seule la racine puisse être l'unité de compte étymologique.

76. Les autres moyens de cette dissyllabisation étant la reduplication CVCCVC (attestée pour pratiquement toutes les racines) ou le redoublement CVCVC.

77. Nous ne pensons pas que MAL moderne *ber-* vienne directement du *mar-* du vieux malais, comme le suggère Blust, ou, en tous cas, que malayo-javanique $*ber-$ puisse s'expliquer de la sorte (cf. Nothofer, sur PAN $*p >$ malayo-javanique $*p/*b$).

pag- fournit une sorte de nom d'action, ou d'infinitif si l'on préfère⁷⁸: il n'y aurait rien d'étonnant à ce que des bases, qui, dans les langues où elles peuvent fonctionner sans affixe, ont des emplois similaires à ceux des formes en *pag-*⁷⁹, soient renouvelées par un infinitif et qu'il en résulte des bases néoformées dissyllabiques. Ainsi, on a, en tagalog, des bases et des formes en *pag-* qui ont un sens très proche:

<i>kung Lunes</i>	<i>ang</i>	<i>alis</i>	<i>sa Maynila</i>	<i>ng (= n-ang)</i>	<i>eruplano</i>
Prép lundi	ArtSuj	partir		Art+Gén-Agt	avion
(lit.) «le départ de l'avion pour Manille (est) le lundi»					
(Schachter et Otones, p. 165)					

vs *ikinagulat ko ang pag-alis niya*
 Causal+surpris 1sg ArtSuj PAG-partir 3sgAgt
 «son départ m'a surpris» (ibidem, p. 163)

et Schachter et Otones vont jusqu'à traduire en partie de la même façon les exemples suivants (ibidem, p. 166):

<i>mabilis</i>	<i>ang</i>	<i>tugtog</i>	<i>niya</i>	<i>n-ang</i>	<i>piyesa</i>
rapide	ArtSuj	jouer	3sgAgt	ArtComplt	pièce
«his playing of the piece is fast»					

<i>mabilis</i>	<i>ang</i>	<i>pag-tugtog</i>	<i>niya</i>	<i>n-ang</i>	<i>piyesa</i>
rapide	ArtSuj	PAG-jouer	3sgAgt	ArtComplt	pièce
«his playing of the piece is fast»					
(«the way that he plays the piece is fast»)					

mais on notera que la différence d'ordre de l'entité réapparaît bien avec:

<i>masyado-ng</i>	<i>mabilis</i>	<i>ang</i>	<i>pag-tugtog</i>	<i>(*tugtog)</i>	<i>niya</i>	<i>ng</i>	<i>piyesa</i>
excessif	rapide	ArtSuj	PAG-jouer		3sgAgt	ArtComplt	pièce
«his playing of the piece is too fast» (ibidem, p. 163)							

Le fait que ces formes en **paR-* aient servi de bases à une extension des marques de focus (avec une valeur différente) n'a pu que favoriser la confusion entre bases radicales en **CVC* et formes en **paR/N-CVC*⁸⁰:

	<i>ang papel na</i>	<i>s-in-ulat-an ko</i>	«le papier sur lequel j'ai écrit»
vs	<i>ang desk na</i>	<i>p-in-ag-sulat-an ko</i>	«le bureau sur lequel j'ai écrit»

Avec **bVR-*, on a les étymons suivants:

**baRtiq* «parched, bonedry» (cf. **beRtiq*, **R/retiq*): CEB *bagti?*, MAL *beras berteh*, TB *borti* (ce dernier considéré comme un emprunt au malais par Blust);
**beRkis* «bind up; bundle» (sur **kis*): TAG *bigkis*, KB *berkis*, SAN *bekise?*

78. Une forme verbale non plus orientée vers un participant mais vers le procès lui-même (cf. Lemaréchal 1989, p. 153-154), c'est-à-dire un prédicat d'ordre supérieur à 1. Il s'agit des «aspectless gerunds» de Schachter et Otones.

79. A une différence d'ordre près (cf. Lemaréchal 1998a, p. 140-151); il s'agit des «nominalized verb bases» de Schachter et Otones.

80. De même, un *maX-* a renouvelé, sans doute indépendamment, les AF en *-um-* dans de nombreuses langues.

«manner of tying» et, sous **beRkes* «bundle» (**kes* sans doute = **kis*), le même TAG *bigkis*, BKL et AKL *bugkos*, AHN, HLG et CEB *bugkus*, TBW(K) *begkes*, IBAN *berekas*, MAL *berkas*, DPB *berkes*; **beRtak* «to hit» (cf. **tak* «split»?)⁸¹: KAN *beltak* (**R* > /l/ en KAN), BKL *bugtak*; **buRsuk* «drive or force into» (**suk* «insert, penetrate, enter»): BKL *bugsok*, CEB et HLG *bugsuk*;

mais aussi, sans racine isolable:

**beRnat* «relapse» (cf. **ñat* «stretched»?): reflets en ISG, DGT, KAN, ILK, IFG, TAG (avec un /y/); **beRsay* «canoe paddle, to paddle»: reflets de **R* en HAN, AKL, HLG, CEB et MSK *bugsay*, BKD et WBM *begsay*, CHM *poksai*; **beRtek* «thorn, sting»;

On trouve aussi au hasard des notes du dictionnaire de Blust (2001) de nombreuses formes isolées reflétant un **R* devant C2:

sous **biŋkas* «undo, untie»: CEB *bigkas* «for string, rope, to snap apart» et TAG *igkas* pour lequel on devra poser VR- en préfixe et non en infixé; sous **biŋkis* «tie»: AKL *bigkis* «sash (piece of cloth around the waist)»; sous **bitas* «tear up» (**tas*): TAG et CEB *bigtas* «unstitched»; sous **retiq/* **Retiq*: MAL *bertih* et TB *borti*.

L'analyse proposée vaut à plus forte raison pour les étymons en **pVR-*, comme:

pe(R)cit/*becit* «squeeze, squirt out» (cit*): MAL *percit* à côté de KEL *besit*, HAN *pucit*, NGA *bhesi*; **peRcik* «splash, spatter» (**cik*): DGT *pegsek*, MAL *percek* «speckle, small spot» et *recek* (à expliquer non par une troncation, mais comme un doublet sans **p/bV-* et avec **aR-* seulement); **pe(R)tak* «knock; pop, burst» (sans doute sur **tak* «split», à rapprocher de **beRtak*, ci-dessus): SAS *pertak* à côté de BON *petak*; **piReket* «sticky, adhesive» (sur **keC*): ILK *pigket*, MAL *perekat* à côté de WOL *pika*; **puRcit* «spurt or gush out» (sur **cit*; cf. **becit*): ILK *pugsit* à côté de TON *puhi*; et, sans racine isolable: **paRhid* «wipe off»: AKL *paghid*, CEB *paghid* à côté de HAN *payid*, et CEB *pahid*, *pahir-an*.

Les quelques étymons en **tV(R/N)-* évoquent le *ter-* du malais et renvoient donc de nouveau à cette langue ou à une protolangue proche — mais on prendra garde que même le tsou, et beaucoup d'autres langues, ont une marque de TAM en *tV* de valeur comparable —; les étymons en **ter-* sont les suivants:

ta(R)kes* «wrap around, encircle» (kes*): KAD *tagkos* à côté d'AMIS *takec* sans reflet de **R*; **taRpis* «thinness, fineness of texture» (**pes*): TAG *tagpis* à côté de OJ *tapis* sans reflet de **R*.

En fait, on trouve un certain nombre de reflets d'un **-R-* précédés ni de **bV-* ou **pV-*, ni de **te-*: ou bien dans des étymons en **CVR-* où C est une consonne autre que **b*, **p* (et **w*), comme:

ce(R)neb* «dive, immerse» (neb*): TB *sornop* à côté de MAR *seneb*, KB *cunem* (tous présents également sous **ceneb*); **ke(R)lap* «flash, shine» (**lap*): MAL *kerlap*; **keRteŋ* «striped, spotted fish»: ILK *kagteŋ*, MAL *ikan kertaŋ*;

81. Pourquoi les reflets n'ont-ils pas été reportés ici sous un **be(R)tak*?

su(R)ked* «prop, support; staff, walking stick» (ked*; cf. *su(η)kud*): WBM *sugked*, à côté d'AMIS *coker*.

ou bien dans des étymons en **VR-* comme:

iReteŋ* «stretched tightly» (ten*): ILK *irteŋ*, TAG *igtiŋ*, MAL *rentaŋ* (**R* > /r/ en ILK; à noter le problème de C₃ -ŋ au lieu de -n)⁸²; **aRsem* «sourness» (**sem*): PUY(T) *HaRsem*, BTK *agsem* (dans la note sous *qalesem*); mais également sous **biŋkas*: TAG *igkas*; sous **hukay* «dig up something buried, disinter»: CEB *ugkay* «dig out something buried» (de nouveau sans C1); sous **ised* «budge, shift, move aside» (cf. **sed*, **sit*, **sud*, **sut*, **sug*): CEB *ugsog* (sans trace de **p*/**b-*).

L'existence, dans des langues comme le bugis ou l'ilocano, de préfixes *ar-* ou *ag-* et *aN-* non précédés de /p/ (ou du *m-* d'AF ou *n-* accompli), aussi bien qu'une simple segmentation morphématique rigoureuse qui impose de segmenter **p-* + **aN-* et **p-* + **aR-* même dans des langues comme le tagalog où *-ag-* (< **aR-*) et *-aN-* sont toujours précédés d'une consonne incitent à reconstruire des «infixes» **aN-* et **aR-*⁸³ de même valeur que **paN-* ou **paR-*, et à proposer **aR-* comme source de la 1ère syllabe de cette dernière série d'étymons.

Dans un certain nombre de langues où les reflets de *(*p*)*aR-* fonctionnent encore comme marques grammaticales vivantes (verbes en **maR-*), on a, sur d'autres bases verbales, au lieu de formes présentant des reflets de *(*p*)*aR-*, des formes avec des reflets de *(*p*)*aN-* (verbes en **maN-*), soit sans différence de valeur sensible, soit avec une opposition entre action «délibérée» ou spécifique pour *(*p*)*aR-* vs «action plurielle» pour *(*p*)*aN-*. Enfin, dans un grand nombre de langues, on constate que tout se passe comme si la neutralisation de la différence au profit de l'un ou de l'autre, mais le plus souvent au profit de *(*p*)*aN-* (comme en palau ou en ivatan⁸⁴), avait eu pour conséquence le passage à une simple valeur de détransitivation (objet indéfini ou inanimé pluriel ou périphérisé ou absent). Les étymons qui présentent un **bVN-* pouvant être issus de ce **paN-*, comme ceux en **bVR-* peuvent être issus selon nous de **paR-*, sont légion, puisque cachés derrière tous les **bV(N)-* du lexique. Et, de même que l'infixe **-aR-* peut être la source des */R/ qui apparaissent devant C₂ dans certains étymons, de même **un **-aN-* peut être la source de tous les *(N) qui apparaissent devant C₂.** Une telle analyse a l'avantage de reconnaître une source morphologique⁸⁵ et

82. On a donc *ir-* ou *ig-* en face de *re-*; par ailleurs MAL *rentaŋ* associe *-aR-* et *-aN-* et cela dans cet ordre précis, ce qu'on retrouve largement au niveau des étymons en **RVNC₂V₂C₃*.

83. On y ajoutera les reflets de **RV(N/R)-* en JAV et quelques autres langues, et sous réserve ceux du BAL et quelques autres langues où il reste à prouver que */R/ > Ø phonétiquement à l'initiale.

84. Cf. Lemaréchal 1986 et 2001, p. 444-445, 450-453.

85. Que l'existence de doublets, figés et opaques, avec et sans **N*, ait fini par donner lieu dans une partie des langues — mais moins nombreuses qu'on ne le suppose d'ordinaire — à une apparence de mutation consonantique est indiscutable, mais cette analyse n'est recevable que là où **C₂* > *NC₂* sans exception.

non plus phonétique aux fameux $*(N)$ qui pullulent dans les reconstructions depuis Dempwolff: ce $*(N)$ est d'ordinaire expliqué par une mutation phonétique (prénasalisation des occlusives) qui se serait étendue progressivement à un nombre de plus en plus grand de mots dans un nombre de plus en plus grand de langues sans qu'on puisse bien souvent dégager une loi phonétique régulière, ce qui oblige à conserver de toutes façons un $*(N)$ dans les étymons et donne lieu nécessairement à un traitement laxiste du détail des reflets⁸⁶.

Si l'hypothèse selon laquelle les élargissements $*bVN/R-$ ou $*pVN/R-$, et sans doute les quelques $*wVR/N-$, sont issus des préfixes $*paR$ et $*paN-$ ne pose guère problème dans la mesure où il s'agit de marques de nom d'action ayant très bien pu servir à renouveler des bases verbales simples fonctionnant elles-mêmes souvent comme infinitif ou nom d'action, il n'en va pas de même pour les élargissements sans trace de $*R$ ou de $*N$ qui devraient être issus du préfixe $*pa-$ dont la valeur à travers toute la famille est celle de marque de causatif, car aucune trace de cette valeur ne semble en général décelable. Une possibilité est que, dans un certain nombre de langues ou de contextes, $/RC/ > /C/$ ou $/NC/ > /C/$, et qu'à côté de quelques $*bV- < *pV-$ provenant peut-être effectivement de causatifs, **une partie des $*bV- < *pV-$ proviennent de $*pVN-$ ou $*pVR-$ avec chute de $/R/$ ou $/N/$ devant $/C/$.**

Enfin, il existe un préfixe $*paN-$, mais aussi semble-t-il $*aN-$, de nom d'instrument et d'IF, dont on pourrait trouver trace dans des étymons à valeur de noms d'instrument ou de bases verbales exprimant la mise en œuvre d'un instrument.

Comme sources d'élargissements préfixés, on dispose donc non seulement de $*paR-$ et $*paN-$ (et peut-être $*pa-$), mais aussi de $*aN-$ et $*aR-$.

Nous ferons l'hypothèse que $*VN-$ est aussi la source de $*NV-$ et $*VR-$ la source de $*RV-$, par métathèse⁸⁷: des exemples de doublets comme $*beRtiq$, $*Retiq/*retiq$ y invitent. On peut penser que, ou bien $*RV-$ est issu de $*VR-$ par simple métathèse, ou bien que $*RV-$ est issu de $*VR-$ + $*V-$ avec chute de la voyelle initiale⁸⁸, sur le modèle des voix actives de l'ilokano en *ag-i-*, voix active (intransitive) de verbes de déplacement (en fait, antipassif formé sur le focus en *i-* «passif de l'objet déplacé»).

Les étymons en $*be(N/R)-$ (et $*teR-$) pointent vers le malais ou vers une protolangue intermédiaire proche sur ce point du malais. Qu'il s'agisse d'emprunt ou d'héritage, le fait que des formes issues de $*paR-$ ou $*paN-$ à valeur de noms d'action se soient figées en étymons en $*bV(V/N)-$ ou en $*pV(V/N)-$ ou (rarement) en $*wV(V/N)-$ renvoie à des regroupements de langues différents et à une chronologie échelonnée. A plus forte raison, il est évident qu'un figement à partir non plus de formes en $*paR-$ ou

86. Il n'est pas exclu que certains $*(N)$ ne proviennent pas de $*aN-$, mais de $*-um-$ et de $*-in-$ avec assimilation de la nasale et, peut-être, conservation de la voyelle ou d'un de ses traits comme l'aperture.

87. On constate que si $*RV-$ est largement attesté, $*NV-$ est pratiquement inexistant.

88. Due à la pression exercée par la dissyllabité des bases.

**paN-* mais de formes en **aR-* ou en **aN-*, avec ou sans métathèse, renvoie à un autre regroupement de langues et à un autre temps, antérieur, selon nous, à la généralisation, dans une partie des langues, des préfixes en *p-* qui en sont issus et qui en sont une réfection. De même, si, au lieu de métathèses, pour les formes en **RV-* (et comme on le verra en **rV-*, **gV-* et, sans doute, certaines formes en **IV-* peut-être en **kV-*), on part de formes du type d'ILK *ag-i-* AF de verbe de déplacement formé au moyen d'une issue de **aR-* sur une forme en **i-* (< **Si-* IF-BF dans son emploi de PF de l'objet déplacé ici), il est évident qu'on tient là un nouvel élément de chronologie relative⁸⁹, ce que nous ne développerons pas ici.

Les affixes **paR-* et **paN-*, ou **aR-* et **aN-*, ont, entre autres fonctions, celle de permettre de dériver des AF de verbes de déplacement (antipassif) à partir de formes verbales orientées vers l'objet mis en mouvement. Or, on sait que ces formes verbales orientées vers l'objet déplacé d'un verbe de déplacement (c'est-à-dire, en termes de voix ou «focus», un PF de l'objet déplacé) sont marquées par **Si-* ou **Sa-*; on peut s'attendre, si notre hypothèse est juste, à ce que ces préfixes d'IF-BF et PF de l'objet déplacé **Si-* et **Sa-* aient laissé des traces sous la forme d'élargissements radicaux en **SV-*, **hV-* et **V-* selon les langues, la simple ergativité de ces langues pouvant contribuer à la confusion entre forme orientée vers un objet déplacé et base. Le statut phonologique du **q* aussi bien que du /*ʔ*/ initial dans les reflets ainsi que la constance ou non de sa notation ne nous paraissant pas toujours clairement assurés (en dépit des arguments de Zorc 1982), c'est peut-être aussi dans ces *(*S*)*a-* et *(*S*)*i-* qu'il faut chercher l'origine d'une partie des élargissements en **qV-*⁹⁰. On dispose donc également, comme sources d'élargissements, de:

**SV-*, **hV-*, **V-* et, peut-être, certains **qV-* < **Si-* ou **Sa-*

Il existe aussi au moins un infixe *-*al-* à valeur entre autres d'action plurielle ou de répétition⁹¹; on dispose donc de:

*-*Vl-* et, par métathèse, **IV-*

89. L'absence de formes en **NV-* issues de **VN-* < **aN-* nous fait pencher pour cette deuxième solution, ce qui aura peut-être pour conséquence indirecte de placer bien haut le débranchement entre langues du type du tagalog et langues du type de l'ilocano. On aurait deux structures de paradigme des marques de diathèse dans ce domaine: **i-* vs **aR/N-* et **i-* vs **aR-i-*; il est net que le second est une réfection du premier.

90. Mais il est possible, étant donné l'existence de variantes d'étymons en **k*/*q* en position de C₂ et de C₃, que certains **qV-* proviennent de **kV-* ou inversement; cf., par exemple, sous les seules entrées en **bi-* du dictionnaire de Blust: **bikas*/**bikas*, **bikaŋ*/**bikaŋ*, **bikak*/**bikak*, **bikit*/**bikit*, **bikuŋ*/**bikuŋ*, **bikut*/**bikut*, et, en fin d'étymons, un certain nombre de **q* peuvent aussi être rapprochés d'autres consonnes, dont **k* (cf. *-*kuk*/*kuq*/*ku*); l'enquête devra être menée systématiquement.

91. Mais le paiwan possède une série d'infices *-*al-*, *-*aɬ-* (< **aN-*) et -*ar*₁ (< **aR-*?), synonymes, «having sound or quality of; involving use of» distincts d'un *-*ar*₂ «do indiscriminably» reflet possible d'une valeur d'«action plurielle»: on peut se demander si *-*al-* n'est pas issu de variantes de **aN-* ou de **aR-*.

Toutefois, comme nous allons le voir plus loin, d'autres sources sont envisageables pour les élargissements préfixés en /l/: dans certaines langues, $*R > /l/$, et, de là, par emprunt à ces langues, un /l/ peut refléter indirectement $*R$ dans des langues où $*R > /r/$ ou /g/ ou autre chose.

Il existe également des préfixes en $*kV-$, en particulier un ou plusieurs $*ka-$ ⁹², à valeur de causatif des verbes statifs (ou des adjectifs), de noms abstraits de qualité, mais aussi, en combinaison avec d'autres affixes, à valeur de réciproque, ou à valeur modale («ability» et «accident»); il existe aussi des préfixes en $*ki-$ (réfléchi et passif)⁹³. On dispose donc:

$*kV- < *ka-$ ou $< *ki-$

On a vu qu'il existait des préfixes en $*tV-$ (cf. MAL *ter-*), de valeur souvent assez proche des préfixes en $*ka-$ modaux potentiels, etc., futurs, etc.; on dispose donc de:

$*tV(N/R)-$

Des préfixes en $*mV-$ (AF ou moyen) et $*nV-$ (accompli) restent très vivants dans la plupart des langues et sont sans doute restructurables dans la protolangue, mais il n'existe guère d'étymons à élargissement préfixé en $*mV-$ ou $*nV-$ ⁹⁴. Comme nous l'avons dit plus haut, il n'est pas exclu toutefois qu'une partie des $*(N)$ devant C_2 proviennent non pas de $*-aN-$, mais de $*-um-$ et de $*-in-$ avec assimilation de la nasale.

Restent enfin des élargissements préfixaux pour lesquels aucun affixe vivant ne semble pouvoir fournir de sources dans la protolangue, sans hypothèse supplémentaire. Il s'agit de $*rV(N/R)-$, $*gV(N/R)-$, $*dV(N/R)-$, $*cV(N/R)-$, $*sV(N/R)-$, $*hV(N/R)-$ (si $*hV-$ est différent de $*SV-$), $*qV(N/R)-$ (si $*qV-$ est différent de $*\emptyset V-$).

Nous proposons d'introduire ici un second ensemble d'hypothèses, qui fait intervenir la phonétique historique et les phénomènes d'emprunt.

Un certain nombre de faits encourageant à supposer qu'une partie des $*r$ et des $*g$ sont en réalité des reflets indirects de $*R$, via l'emprunt, à savoir qu'un $*r$ peut être un fantôme dû à des emprunts faits par des langues où $*R > /g/$ (comme le tagalog) à une langue où $*R > /r/$ (comme le malais), et qu'un $*g$ peut, inversement, être un fantôme dû à des emprunts faits par des langues où $*R > /r/$ à une langue où $*R > /g/$: il suffit, en effet, du fait de la méthode employée par les comparatistes actuels, qu'il y ait de tels reflets indirects dans deux ou trois langues pour que des fantômes en $*r$ ou $*g$ soient projetés comme des étymons dans la protolangue. Ainsi, à côté des formes, la plupart du temps isolées dans les langues actuelles, en $*(C)VR-$,

92. Cf. Lemaréchal 1998a, p. 152-161.

93. Qui contient peut-être un $*-i-$ à valeur de moyen ou de réfléchi; $*ki-$ pourrait en être une réfection via le réciproque-comitativ $*ka- + *i-$ (cf., par ailleurs, $*paRi-$).

94. Dans quelques cas, dont certains signalés par Blust dans ses notes, $*N < *-um-$; il existe quelques étymons en $*nV-$.

il existe des formes, non moins isolées, en $*(C)Vr-$, ou en $*(C)Vg-$. Il est possible donc qu'une partie au moins des étymons en $*rV-$ et $*gV-$ proviennent des $*RV-$. Etant donné que $*R > /l/$, dans quelques langues (comme le cebuano) et, ailleurs, dans certains cas devant $/u/$, il est possible que quelques $*-Vl-$ et $*lV-$ (après métathèse) soient dûs également à l'emprunt, emprunt fait par une langue où $*R > /r/$ ou $*R > /g/$ à une langue où $*R > /l/$.

Par ailleurs, l'existence de formes en $*(C)Vk-$, à côté des formes en $*(C)VR-$, $*(C)Vr-$ et $*(C)Vg-$, en contexte dévoisant (c'est-à-dire devant C_2 sourde) suggère que quelques $*k$ sont des fantômes provenant de $*g$ par dévoisement, lui-même issu, selon le mécanisme décrit ci-dessus, de $*R$. Il est possible qu'une partie des $*kV-$ aient également cette origine.

La rencontre de quelques variantes d'étymons en $*V-$ + racine vs $*SV-$ + racine vs $*sV-$ + racine, mais aussi $*cV-$ + racine, $*tV-$ + racine, pourrait même suggérer, du fait uniquement de sa naturalité bien attestée ailleurs, qu'il aurait existé une chaîne d'évolution phonétique portant sur C_1 du type: $*t > *C/*c > *s > *h > \emptyset$, ou bien des fragments de cette chaîne reliant transversalement des étymons en: $*t > *C/*c > *s$, ou bien: $*s > *S > *h > \emptyset$. Il n'est pas impossible qu'on doive rendre compte ainsi d'une partie des étymons en $*cV-$, $*CV-$, $*sV-$, $*SV-$, $*hV-$. Cela reste évidemment, comme le reste, à démontrer.

4.3. Un programme: la recherche des traces de motivation

Une des façons d'apporter un début de vérification à nos hypothèses serait de dégager des traces, dans les différents étymons d'une racine donnée, de la valeur ancienne des affixes supposés être à l'origine de leurs élargissements radicaux. On verra que l'examen des étymons formés sur la racine $*tas$ donne quelques résultats allant dans ce sens.

Nous commencerons toutefois par un exemple plus simple destiné seulement à prouver le bien-fondé d'une telle recherche, une sorte d'étude de faisabilité. La racine $*keb$ «cover» (à rapprocher par ailleurs de $*kep$, $*kub$ et $*kup$ ⁹⁵) est au centre d'un ensemble de 9 étymons: $*a\eta keb$ «covering, lid», $*e keb$ «cover», $*le(\eta) keb$ «cover, shut in», $*lu(\eta) keb$ «cover», $*Ru\eta keb$ «cover over, cover up» — sous lequel on trouve, abusivement à notre avis, JAV $u\eta keb$ et $ukeb$, à côté de véritable reflets en $*R-$ —, $*se(\eta) keb$ «cover», $*si(\eta) keb$ «cover», $*ta(\eta) keb$ «cover, overlapping part» (+ «eyelid»), $*ti(\eta) keb$ «close up». Au sein de cet ensemble, $*a\eta keb$ a la particularité, aux côtés de $*ta(\eta) keb$, d'être glosé par un nom: les reflets en sont AKL $a\eta kob$ «seal, filling, joiner, gap-filler», CEB $a\eta kub$ «covering over the opening of a small boat, the analogue of a deck on large boats; fill something in to cover up a deficiency»⁹⁶, BAL $a\eta keb$ «cover(ing), lid, bed-spread, table-cloth»;

95. Nous évoquerons plus loin le problème posé par ces variations de voyelle ou de consonne.

96. En note, Blust signale que AKL et CEB sont ambigus et pourraient venir aussi bien de $*kub$ que de $*keb$ (le BM possédant un $a\eta kub$ «cover» où $/u/$ ne peut venir que de $*u$); on remarquera que: 1) $*kub$ pourrait n'être qu'un fantôme de $*keb$ via des emprunts à une ou

**aN-* y est bien associé à la valeur de «nom d'instrument» ou de «base verbale exprimant la mise en œuvre d'un instrument», une des valeurs que nous reconnaissons à **aN-*, refait dans une partie des langues en **paN-*.

Notre racine **tas*⁹⁷ nous fournira l'occasion d'un commencement de vérification de deux de nos prédictions. La première est que, si l'infixe **-aN-*, marque d'«action plurielle» (secondairement, marque de détransitivation, entrant éventuellement dans la formation d'antipassifs⁹⁸ et à distinguer du **aN-* > **paN-* «nom d'instrument dédié» que nous venons de voir⁹⁹), est bien la source du *(*N*) prétendument labile des étymons en **CV(N)CVC*, on peut s'attendre à ce que des reflets en /*N*/ s'opposent à des reflets sans /*N*/ en tant que présentant une valeur de répétition, etc. Loin de nous l'idée que formes avec et sans /*n*/ n'aient pas pu interférer, sinon se confondre dans certaines langues ou sous-groupes de langues, mais dans l'hypothèse où **N* serait la trace d'un ancien morphème, on ne peut poser de *(*n*) dans les étymons sans courir le risque d'occulter ces traces. La seconde prédiction est que, si au moins une partie des élargissements radicaux en **SV-*, **hV-*, **V-* et peut-être **qV-* sont bien issus de **Sa-* ou **Si-* marque d'IF-BF et de PF de l'objet déplacé, les étymons présentant de tels élargissements pourraient garder la trace de cette valeur.

plusieurs langues où PAN **e* > /*u*/ (ou l'inverse via des langues où PAN **u* > /*e*/), et 2) la stratégie, adoptée assez récemment par Blust et étendue à tout le dictionnaire, consistant à éliminer toute trace d'étymologie ascendante du style CEB *aŋkub* < **aŋk(e,u)b* — ce qu'il avait pratiqué antérieurement — a pour effet de multiplier les entrées (ici **aŋkeb* n'est fondée que sur BAL *aŋkeb* et **aŋkub* que sur BM *aŋkub*) mais comporte aussi le grave risque d'engendrer des fantômes.

Le doublet, à notre avis purement phonétique, **aŋkub* mérite qu'on s'y attarde. D'abord cette entrée réunit des reflets hétérogènes par leur signifiant: BON *akob* «large covered basket, generally used as a lunch basket for carrying rice to the fields» — il s'agit d'un nom (mais non d'un nom d'instrument) qui désigne un objet couvert, et non un couvercle ou une couverture, etc., et où, par conséquent, **a-* peut être issu de la marque de PF de l'objet déplacé —; ILK oppose *akkub* «cover, wrapper, shell (of brab)» — c'est donc un nom et, cette fois, un nom d'instrument (ce mot suggère que ILK /*kk*/ < **Nk*, mais /*kk*/ < **Rk* est également possible) — et *akob* «join together...» verbe de déplacement mettant en jeu un objet déplacé joint à un autre par l'action (*V-*, en l'occurrence *a-*, < **Sa-* marque d'IF-BF et PF de l'objet déplacé); JAV et BM ont *aŋkub*: le premier est traduit par «calyx, flower sheath (green cover that splits at blossoming)», c'est-à-dire un couvercle (NInstr), et le second par «to cover» un verbe, dont on ne peut savoir si le *aN-* est simplement une marque d'intransitivation ou de verbe de mise en œuvre d'un couvercle ou tout autre instrument destiné à couvrir; s'y ajoute ARS (donc une langue OC) *saku*, dont nous ne sommes pas en état, par pure ignorance, d'interpréter le /*s*/ (**S* > /*s*/ ? ou bien **s* > **S* dans certains cas comme nous le pensons, voir quelques doublets en **s* vs **S*?).

97. Voir plus haut la reproduction intégrale des différentes entrées du dictionnaire de Blust (version 2001).

98. Cf. Lemaréchal 2001, p. 442-453.

99. Un rapport peut exister entre «action plurielle» et instrument «dédié» (à usage réitéré), distinct d'un objet mis en œuvre épisodiquement comme instrument (IF **Si/a-*); de même, une des valeurs de la reduplication dans les langues austronésiennes est la répétition, mais elle sert aussi à former des noms d'«instrument dédié» du type français *coupe-coupe*, *pousse-pousse*.

Blust lui-même distingue deux entrées **bentas* et **betas* qui s'opposent comme «hack a passage through vegetation, blaze a trail» et «tear, rip open (as cloth or stitches), snap (as string)». Il ne faut pas voir dans ces deux valeurs une simple différence de signifié justifiant l'existence de deux entrées séparées, le *-n-* fonctionne ici pratiquement comme un véritable morphème («cut off» vs «cut through» = «couper» ± «action plurielle»): le chemin ouvert n'est pas à entendre métaphoriquement comme une déchirure du tissu végétal vue à échelle macroscopique, mais comme résultat de couper à coups répétés de multiples végétaux pour se frayer un chemin; tous les reflets de **bentas*¹⁰⁰ vérifient ce sème, sauf MAL¹⁰¹, tandis que ceux de **beRtas* et de **betas* évoquent plutôt une rupture unique (fil qui se rompt, tissu qui se déchire ou se découd, entre autres), sauf BKL *bagtas* «traverse; cut across in walking (as across a field)», qui semble fort proche de «(take a) shortcut»¹⁰². De même, les reflets sans /N/ de **bitas* ne dénotent pas d'action de couper à coups répétés, sauf MSK *bitas* dans un de ses sens¹⁰³. Cette valeur de *-n-* est en fait confirmée pour la plupart des reflets en /n/ issus d'autres étymons formés sur **tas* présentant un **(N)* entre parenthèses ou même d'étymons sans **n* du tout pour lesquels le /n/ de certains reflets n'est pas répercuté: on trouve un /n/ pour les reflets désignant le «raccourci» ou l'«action consistant à en ouvrir un, ou à en prendre un», tandis qu'il se vérifie que les reflets sans *-n-* expriment une action unique; c'est le cas pour **qantas*. L'étymon **etas*, qui est sans doute un fantôme, n'a que deux reflets, l'un sans *-n-*, AKL *utas* — repris aussi sous **qutas* où il est plus à sa place — qui dénote une action unique, et l'autre avec *-n-*, SAS *entas*, où l'on retrouve la valeur «cut through (forest, hedge)». Sous **butas*, traduit «cut through, sever», glose obtenue, comme cela arrive assez souvent, par addition des signifiés de reflets qui ont en fait des sens divergents, on trouve TAG *butas* «perforation, hole; loop-hole», etc. (qu'il faut entendre comme action unique¹⁰⁴) et PS **butas* «behead, cut off head (as of fish)», en face de SND *buntas* «reclaim, as a cultivated field; cut or burrow through (as a mountain); come to an end, of a discussion», sens issus d'un «clear vegetation, cut through»; en revanche, seul le premier sens donné pour IBAN *buntas* implique une action multiple. Pour **ri(n)tas*, l'opposition avec vs sans *-n-* paraît également pertinente: sur

100. **maŋ-bentas* sous **bentas* manifeste la tendance assez générale à travers les langues à réitérer un marquage qui tend à s'obscurcir — du fait qu'un affixe tende à être ressenti comme un infixe (ici **-N-*) en raison du figement de **be-* —, au moyen de la réitération de l'affixe en position de préfixe (ici dans le préfixe néoformé **maN-*), à comparer au cas des verbes en *-N-* + *-ano* du grec ancien: *la-m-b-án-ō* sur *lab-*.

101. MAL *bentas* «tearing up and dashing down; e.g. of a Titan tearing out a hill by the roots and dashing it down on the divine warriors, his enemies»

102. On dira peut-être qu'ouvrir un raccourci dans une végétation dense n'est pas la même chose que simplement «traverser»...

103. Et, inversement, TB *bintas* «small cut, incision» dans une langue où **t* ne donne pas systématiquement /nt/.

104. L'apparition d'un sème «multiple» dans TAG *butas-butas* «full of holes» est due à la reduplication.

les trois reflets, on a IBAN *rintas* «torn open, broken through»¹⁰⁵ et MAL *rintas* «take the shortest way, cut across» en face de MAR *ritas* «separate, cut off». Tous les reflets de **ra(n)tas* qui ont un /n/ renvoient à une coupe exécutée par action répétée, sauf MAD «cut» avec -n- et MGG «tattered» sans -n-, qui sont plutôt à mettre du côté des contre-exemples. Les reflets de **Ratas* et de **tetas*, dont aucun ne présente de /n/, ne renvoient jamais à des «actions multiples». Quant à **butas*₂ «to separate people, disperse a gathering», les reflets sans -n- expriment l'action consistant à détacher un élément d'un groupe ou, pour un élément, à se détacher d'un groupe (BKL, MSK), tandis que BAL¹⁰⁶ *buntas* «disperse (a gathering) et *buntas-an* «be made to disperse» (vraisemblablement avec -an < **-ak^on* IF à valeur de causatif) expriment une action multiple. Sous **qutas*, seuls BAR et TAO présentent un /n/, le premier signifie «burst open so that the contents spill out» (avec éparpillement du contenu ou du contenant en plusieurs morceaux qui se dispersent?) et le second «to cross, go across a road, street, sea, bridge, etc.» qui n'implique pas de patient ou d'action multiples, mais qui peut très bien être un ancien «traverser la forêt ou la végétation»; les autres reflets, sans /n/, ont une valeur d'«action unique», sauf sans doute ILK *usat* (avec méatèse) «open a road, clear a path (in the jungle)»¹⁰⁷, signifié régulièrement associé à des reflets en -n- et TIR *utoh* «to pick, to harvest», action dont on s'attend à ce qu'elle soit comprise comme «plurielle» (ou bien **NC* > TIR /C/?). Ainsi notre première prédiction se trouve largement vérifiée.

On aura relevé que la traduction d'un certain nombre de reflets (sans -n-) de **qutas* exprime la séparation d'une partie ou d'un morceau précis d'un ensemble: cette valeur de «séparer un objet précis d'un ensemble par déchirure ou coupure» se comprend mieux si ce **(q)u-* provient d'une marque d'IF-BF et de PF de l'objet déplacé (**Si-* ou **Sa-* > **i-* ou **a-* hors Formose, et, de là, en raison de l'opacification et du figement en élargissement, > **V-*) qui, dans ces langues ergatives, spécifie comme soumis à un déplacement l'objet précis désigné par le sujet (normalement toujours défini). Par rapport à ces étymons issus de PF de l'objet déplacé (défini), les étymons issus des formes en **aR-* et en **aN-* font encore figure d'antipassifs d'«action délibérée» vs «action multiples», non plus orientés vers l'objet séparé mais vers l'action elle-même, une ou plurielle. Ainsi notre seconde prédiction se trouverait aussi vérifiée. On aurait ainsi la trace d'un système **(p)aR-* vs **(p)aN-* vs **Si/a-*.

105. Il est possible, au vu des traductions par des participes passés, que /in/ garde en IBAN plutôt la trace de l'infixe **-in-* d'accompli. On en conclura que tous les /rintas/ ne sont pas nécessairement le reflet du même étymon; des étymons peuvent être des fantômes dûs à ce qu'on reconnaisse trop vite comme «cognates» ce qui se ressemble le plus immédiatement. Il est vraisemblable que d'autres **(N)* soient issus de **-in-*: le détail des faits reste à cerner de plus près, reflet par reflet.

106. Dont il est faux, à notre avis, de considérer que **t* > /nt/ et très risqué de prétendre que **t* produit indifféremment /t/ et /nt/.

107. ILK est une langue où il semble que **NC* > /C/ (et parfois /CC/? à moins que **NC* > /C/ vs **RC* > /CC/).

Les autres étymons¹⁰⁸ présentant la racine **tas* avec l'un des élargissements **Ra-*, **ra(n)-*, **ri(n)-* ou **ru-* (avec un reflet en *-n-*) contiennent un reflet **R* ou **r* du **aR-*, marque d'antipassif de l'action spécifique ou délibérée, qui s'oppose à **aN-* (> **-N-*) ou se combine avec lui. Seul subsiste le problème de la voyelle de ces **R/rV(n)-*, simple voyelle d'appui colorée ou écho d'un affixe (**i*-¹⁰⁹ de **(S)i-* représenté par une voyelle haute et **a-* de **(S)a-* ou de **aR-* représenté par **a* ou **e*)?

4.4. *Un second programme de recherche: résoudre les problèmes du signifiant; une phonétique historique parallèle?*

Quant au signifiant, on aura remarqué que nous n'avons étymologisé le plus souvent que **C₁* et qu'implicitement nous avons fait comme si **V₁* était indifférente; dans certains cas, la voyelle de la racine semble même l'être aussi: voir **tas* et **tes*, de même sens.

Nous avons donc fait comme si les voyelles étaient négligeables; cela nous a permis — peut-être de façon artificielle — de faire intervenir un nombre très réduit d'affixes — certainement beaucoup trop réduit par rapport à la réalité¹¹⁰ — censés rendre compte de très nombreux «élargissements radicaux» préfixés et infixés. L'hypothèse implicite est qu'une fois figé, et éventuellement sujet à l'emprunt, le couple formé par les deux voyelles de l'étymon devenu monomorphématique a été l'objet d'influences réciproques d'une voyelle sur l'autre; ce n'est plus chaque voyelle, prise séparément, qui est un trait d'une protoforme monosyllabique, mais **le couple des deux voyelles, prises solidairement, qui devient un trait des nouvelles bases dissyllabiques.**

Ici encore, le nombre de phénomènes — par ailleurs, bien connus — qu'il est nécessaire de faire intervenir est réduit:

- réduction d'une des deux voyelles en schwa,
- harmonie vocalique,
- copie de voyelle,
- effet de mise en contraste en termes d'aperture (couples V1-V2 en *a-i*, *a-u*, *i-a*, *u-a*) ou de lieu d'articulation (couples en *i-u* et *u-i*).

La variation de la voyelle de la racine elle-même pose un autre problème: on a **tes* et **tas*, mais aussi **keb* et **kub*, **kep* et **kup*, tous glosés «cover» et précédés de nombreux **CV(N/R)-* divergents, et bien d'autres racines. Or, entre **butes* et **butas*, entre **(q)etes* et **(q)utas*, entre **re(n)tes* et **R/rV(n)tas* (laquelle de ces variantes faut-il choisir comme pendant de **re(n)tes*? **ra(n)tas*?), il y a eu forcément développements parallèles, mais

108. **tetas* n'est reconstruit que sur trois reflets qui permettent difficilement de savoir si **te-* est un écho d'un des affixes connus en **te-* (avec quelle valeur?) ou d'un redoublement (**tetas* < **tatas*); on notera seulement qu'aucun reflet ne présente de /n/.

109. Comme dans ilocano *ag-i-*, dérivé de verbes de déplacement en *i-*.

110. On rappellera de nouveau la possibilité qu'une partie des /N/ soit issu de **-in-* ou de **-um-* et non de **-aN-* et que l'opposition de voyelles garde trace de l'opposition d'affixe.

lesquels? Le fait de poser des racines distinctes implique que c'est la production d'étymons à partir de **tes* et à partir de **tas* qui constitue des développements parallèles entre:

**tas* > **betas*, **beNtas*, **beRtas*, **bitas*, **butas*,
**etas*, **qa(n)tas*, **qutas*,
**ra(n)tas*, **ri(n)tas*, **rutas*, **Ratas*,
**tetas*
 et **tes* > **butes*,
**(q)etes*,
**re(n)tes*

Or, rien n'est moins sûr: dans le cas du couple **tas/*tes*, il paraît plus vraisemblable de poser des développements parallèles de la forme:

**tas* > **(q)utas* > **(q)utes*
 // > **butas* > **butes*
 // > **R/rV(n)tas* > **re(n)tes*

ou

// **tes* > **(q)utes* > **(q)utas*
 // > **butes* > **butas*
 // > **re(n)tes* > **R/rV(n)tas*

Si c'est bien le cas, une des deux racines est un artefact dû à la méthode employée.

Quelle que soit la solution, il semble que, pour les ensembles d'étymons contenant **-tas* et **-tes*, ce soit **/tes/* qui provienne de **/tas/* et non l'inverse. Si tel est le cas, **a* > **e* — et, en ce qui concerne **kub/p* et **keb/p*, **u* > **e*, ou l'inverse —, ce qui jette le doute sur la belle stabilité des voyelles qu'enregistre un tableau comme celui reproduit au parag. I 3 où **a* > /a/, **i* > /i/, **u* > /u/ (à côté, il est vrai, de /e/ et /o/ pour les deux dernières, mais, dans beaucoup de langues, il ne s'agit encore que de simples allophones). **Les voyelles auraient-elles une autre histoire, occultée par la masse de ce qui se ressemble? Emprunt ou héritage?** Il est nécessaire de dresser un inventaire complet des faits de ce genre.

Enfin le choix d'une solution dans le traitement du parallélisme entre **(q)utas*, **butas*, **R/rV(n)tas* et **(q)utes*, **butes*, **re(n)tes* n'est pas indifférent pour ce qui est du rapport à établir entre **Ra-* et **re(N)-*, **ra(n)-*, **ri(n)-* et **ru-* (peut-être **ru(n)-*), c'est-à-dire aussi entre **R* et **r*, et enfin, comme nous venons de le voir, de manière plus générale, entre les voyelles des élargissements et celles de la racine elle-même.

Conclusions

Il n'est pas question ici de minimiser le travail colossal de recueil des données, de synthèse et d'élaboration d'hypothèses accompli pendant le

dernier demi-siècle, sans lequel — sans les travaux de Blust en particulier — pas une ligne du présent article n'aurait pu être écrite; il est beaucoup plus difficile d'évaluer les reconstructions auxquelles ont mené des méthodes dont les présupposés apparaissent souvent extrêmement contestables.

La lexicostatistique, bien loin d'être une «solution d'attente» doit être déclarée «nulle et non avenue»; tout au plus, de trop grands écarts entre la quantité de vocabulaire partagé et ce qu'on attendrait entre deux langues ou trop proches ou trop éloignées pourront être considérés comme un signal d'alarme et l'indice d'une anomalie à élucider; mais on est loin de cette recette pour construire des arbres généalogiques en toute rigueur mathématique, qui aurait dû balayer les méthodes classiques d'appareillage et de regroupement des langues.

La théorie des migrations comme on l'a vu à propos de l'article **(k)u* exerce des contraintes illégitimes, à la fois a priori et externes, sur l'élaboration des reconstructions possibles; il faut redonner à la linguistique comparative son autonomie par rapport aux données ou aux suggestions de la linguistique externe; leurs résultats respectifs doivent se conforter les uns les autres et apporter par là un élément de vérification et non se compléter les uns les autres dans une fresque infalsifiable.

Un certain scientisme qui se pare d'une rigueur illusoire dans la mise en œuvre de concepts légitimes conduit à des hypothèses irréalistes; c'est le cas de l'idée selon laquelle un objet reconstruit, comme un protophonème ou un étymon, se définit comme une simple classe d'équivalence entre ses reflets, ce qui a conduit un temps à une multiplication irréaliste de certaines classes de protophonèmes, et aboutit encore aujourd'hui à une multiplication irréaliste d'étymons qui font doublets.

Le risque de prendre des emprunts pour des héritages est inscrit dans le projet même de comparaisons de masse — au nom de la loi des grands nombres —, où c'est la ressemblance immédiate qui prime. La faible évolution des systèmes phonétiques entre la protolangue et un bon nombre des langues actuelles en est peut-être un reflet direct; un grand nombre de divergences qui sont nécessairement le résultat d'évolutions par héritage ou par emprunt sont occultées par la reconstruction de doublets ou du fait qu'elles sont reléguées au rang d'exceptions, au point qu'on peut se demander si, au moins dans certains cas, ce n'est pas le rapport entre l'exception et la règle qui doit être inversé¹¹¹.

Nous avons suggéré qu'un grand nombre de doublets en CVCVC distincts uniquement par leur C₁V₁- sont les *membra disjecta* d'une ancienne morphologie¹¹², que les voyelles et une partie des consonnes y ont une autre

111. En grammaire comparée, on peut moins que jamais définir, en cas d'emprunts massifs, l'exception et la règle en termes quantitatifs.

112. Les affixes concernés sont **(-)aR-*, **(-)aN-*, **pa-*, **paR-*, **paN-*, **(a)R-aN-*, **V-* (*<*(S)ali-?*), **tV(R)-*, **ka-*, **-um-*, **-in-*; manquent **-an*, **-°n*. Une partie de ces affixes sont ceux que l'on trouve aussi dans les langues mon-khmer (cf. Lemaréchal 1998a, p. 161-164) comme khmer *prə-* «cause to V, keep on V-ing, V one another» (**pa-* + **-aR-*); un *-r* - semble

histoire — où interagissent emprunt et héritage — que celle qui apparaît dans les tableaux d'équivalence entre protophonèmes et phonèmes attestés. L'examen indispensable du détail des faits où les itinéraires des personnes mais aussi des mots n'ont cessé de se croiser sur de larges domaines et pendant de longues périodes et cela, éventuellement, en plusieurs strates successives, constitue un ou plutôt des programme(s) de recherche considérable(s)¹¹³.

Abréviations de noms de langues

ACH: aceh	ILK: ilocano	PAI: paiwan
AKL: aklanon	ISG: isneg	PAL: palau
ARS: arosi	ITB: itbayaten	PAZ: pazeh
ATY: atayal	ITW: itawit	PGS: pangasinan
BAL: balinais	JAV: javanais	PM: proto-minahasan
BAR: bare'e	KAD: kadazan	POC: proto-océanien
BGG: banggai	KAN: kankanay	PON: ponape
BGS: bugis	KAP: kapampangan	PS: proto-sangir
BJR: banjar	KAV: kavalan	PSS: Proto-South-
BKD: bukidnon	KAY: kayan	Sulawesi
BKL: bikol	KB: karo batak	PUL: puluwat
BM: bolaang mongondow	KEL: bario kelabit	PUY: puyuma
BON: bontok	KMB: kambara	REN: rennell
BUN: bunun	KUS: kusaie	ROT: rotuma
BUR: buru	LgT: long terawan	RUK: rukai
CEB: cebuano	MAK: makassar	SAI: saisiyat
CHM: chamorro	MAL: malais	SAM: samoan
CMP: Central Malayo-	MAR: maranao	SAR: saaroa
Polynesian	MDR: madurais	SAS: sasak
DGT: casiguran dumagat	MGG: manggarai	SED: seediq
DPB: dairi-pakpak batak	MIN: minangkabau	SHWNG: South Halmahera-
F: Formosan	MLG: malgache	West New Guinea
GED: gedaged	MNY: ma'anyan	SIK: sika
GLB: gilbert	MP: malayo-polynésien	SIM: simalur
GOR: gorontalo	NGA: ngadha	SIR: siraya
HAN: hanunoo	NgD: ngaju dayak	SML: samal
HLG: hiligaynon	NUK: nukuoro	SND: sundanais
HWN: hawaïien	OC: océanien	TAG: tagalog
IFG: ifugao	OJ: vieux javanais	

isolable dans *trə-*, *crə-*, *krə-* «keep on V-ing» à rapprocher de la valeur d'«action délibérée, insistance sur l'action» assignée à **aR-*; on trouve des infixes *-m-* d'«agent», *-n-* de «résultat» et *-rəN-* de nom d'instrument, à rapprocher de **-um-*, de **-in-* et du **(-)aN-* de nom d'instrument; une valeur de «causatif» est attribuée à *prə-* (cf. **paR-?*). On comprendra que, dans ces conditions, nous continuions de placer plus volontiers le «berceau» de la famille dans l'Asie du sud-est qu'à Formose. Mais une valeur de «causatif» est attribuée aussi à *bən-* (cf. **be(N-?)*): héritage commun ou emprunt?

113. Cf. nos travaux en cours à paraître sur **S* et sur le vocalisme.

TAO: taosug	TIR: tituray	WMP: West Malayo-
TB: tobaatak	TON: tongien	Polynesian
TBW: tagbanwa	WBM: Western	
TET: tetum	bukidnon manobo	
	WET: wetan	

Références bibliographiques

- ASAI Erin, 1934, «Some observations on the Sedik Language of Formosa», *Phologia Orientalis* I.
- ASAI Erin, 1953, *The Sedik language of Formosa*, Kanazawa.
- AUROUX Sylvain, 1994, «L'hypothèse de l'histoire et la sous-détermination grammaticale», *Langage* 114, pp. 25-40.
- BELLWOOD Peter, James J. FOX et Darrell TRYON (éds.), 1995, *The Austronesians. Historical and Comparative Perspectives*, Canberra, The Australian National University.
- BENEDICT Paul K., 1942, «Thai, Kadai, and Indonesian. A next alignment in Southeastern Asia», *American Anthropologist*, 44, p. 576-601.
- BENEDICT Paul K., 1966, «Austro-Thai», *Behavior Science Notes* 1, p. 227-261.
- BLUST Robert A., 1978, «The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary Report», *Working Papers in Linguistics, Honolulu* 9/2, p. 1-15.
- BLUST Robert A., 1980, «Austronesian Etymologies», *Oceanic Linguistics*, 19, p. 1-181.
- BLUST Robert A., 1984, «Austronesian Etymologies — II», *Oceanic Linguistics*, 22-23, p. 29-149.
- BLUST Robert A., 1985, «The Austronesian Homeland: a linguistic perspective», *Asian Perspectives* 26/1, p. 45-67.
- BLUST Robert A., 1986, «Austronesian Etymologies — III», *Oceanic Linguistics*, 25, p. 1-123.
- BLUST Robert A., 1988, *Austronesian Root Theory*, Amsterdam, Benjamins.
- BLUST Robert A., 2001, *Austronesian Comparative Dictionary* (version électronique 2001).
- BRANDES Jan L. A., 1884, *Bijdrage tot de vergelijkende klankleer der Westersche Afdeling van de Maleisch-Polynesische Taalfamilie*, Utrecht.
- BRANDSTETTER Renward, 1893-1908, *Malaio-polynesische Forschungen*, Lucerne.
- BRANDSTETTER Renward, 1910, *Wurzel und Wort in den Indonesischen Sprachen*, Lucerne, Haag.
- BRANDSTETTER Renward, 1915, *Die Lauterscheinungen*. Monographien zur Indonesischen Sprachforschung XII, Luzern.
- BRANDSTETTER Renward, 1916, *An Introduction to Indonesian Linguistics*, London (traductions abrégées d'une partie des précédents).
- CHANG Kwang-chih, G. W. GRACE et Wilhelm G. SOLHEIM, 1964, «Movement of the Malayo-Polynesians: 1500 B.C. to A.D.», *Current Anthropology* 5, p. 359-406.
- DAHL Otto Christian, 1971, *Proto-Austronesian* (2ème éd. 1976), London, Curzon Press.
- DAHL Otto Christian, 1976, «Semantics in lexicostatistics and the problem of borrowing», *Norwegian Journal of Linguistics* 30, p. 203-233.
- DAHL Otto Christian, 1981, *Early phonetic and phonemic changes in Austronesian*, Oslo, Universitetsforlaget.

- DEMPWOLFF Otto, 1920, *Die Lautentsprechungen der indonesischen Lippenlaute in einigen anderen austronesischen Südseesprachen*, 2. Beiheft zur Zeitschrift für Eingeborenen-Sprachen, Berlin.
- DEMPWOLFF Otto, 1923, «Fiktion und Hypothese in der Sprachwissenschaft», *Annalen der Philosophie* 3, p. 246-257.
- DEMPWOLFF Otto, 1925, «Die L-, R- und D-Laute in austronesischen Sprachen», *Zeitschrift für Eingeborenen-Sprachen* 15, p. 19-50, 116-138, 223-238, 273-319.
- DEMPWOLFF Otto, 1925-26, «Ivatan als 'Testsprache' für uraustronesisches l», *Zeitschrift für Eingeborenen-Sprachen* 16, p. 298-302.
- DEMPWOLFF Otto, 1927, «Das austronesische Sprachgut in den melanesischen Sprachen», *Folia Ethnoglologica* 3, p. 32-43.
- DEMPWOLFF Otto, 1928, «Das austronesische Sprachgut in den polynesischen Sprachen», *Feestbundel uitgegeven door het Koninklijk Bataviaasch Genootschap*, p. 62-86.
- DEMPWOLFF Otto, 1931, «Einigen Probleme der vergleichenden Erforschung der Südsee-Sprachen», *Anthropos* 26, p. 157-170.
- DEMPWOLFF Otto, 1934-1938, *Vergleichende Lautlehre des austronesischen Wortschatzes* (I-III), Beiheft zur Zeitschrift für Eingeborenen-Sprachen 15, 16, 17.
- DYEN Isidore, 1947, «The malayo-polynesian word for "two"», *Language* 23, p. 50-55.
- DYEN Isidore, 1951, «Proto-Malayo-Polynesian *Z», *Language* 27, p. 534-540.
- DYEN Isidore, 1953a, *The Proto-Malayo-Polynesian laryngals*, Baltimore.
- DYEN Isidore, 1953b, «Dempwolff's R», *Language* 29, p. 359-366.
- DYEN Isidore, 1962, «Some new Proto-Malayopolynesian initial phonemes», *Journal of the American Oriental Society* 82, p. 214-215.
- DYEN Isidore, 1963, «The position of the Malayopolynesian languages of Formosa», *Asian Perspectives* 7, p. 261-271.
- DYEN Isidore, 1965a, *A lexicostatistical classification of the Austronesian languages*, Baltimore.
- DYEN Isidore, 1965b, «Formosan evidence for some new Proto-Austronesian phonemes», *Lingua* 14, p. 285-305.
- DYEN Isidore, 1969, «Reconstruction, the comparative method, and the proto-language uniformity assumption», *Language* 45, p. 499-518.
- DYEN Isidore, 1971, «The Austronesian languages and proto-Austronesian, in *Linguistics in Oceania*, p. 5-54.
- DYEN Isidore, 1973, «The validity of the mathematical model of glottochronology», in *Lexicostatistics in genetic linguistics*, The Hague, p. 11-29.
- DYEN Isidore, 1975, *Linguistic Subgrouping and Lexicostatistics*, The Hague, Mouton.
- FERRAND Gabriel, 1905-1907, 1908, Comptes rendus de Brandstetter 1893-1908, 1910, 1915, *BSLP* XIV p. 266, XVI p. 408, XVII p. 159, XVIII p. 354.
- FERRELL Raleigh, 1969, «Taiwan aboriginal groups: Problems in cultural and linguistic classification», Taipei.
- FERRELL Raleigh, 1972, «Construction markers and subgrouping of Formosan languages», pp. 199-211.
- FERRELL Raleigh, 1982, *Paiwan dictionary* (= *Pacific Linguistics* C-73).
- GERAGHTY Paul A., 1983, *The History of the Fijian Languages*, Oceanic Linguistics Special Publ. n° 19, Honolulu.
- GONDA Jan, 1940, «Opmerkingen over de toepassing der comparatieve methode op de indonesische talen, voornamelijk in verband met hun woordstructuur», *Bijdrage tot de taal-, land- en volkenkunde van Nederlandsch-Indië*, 99, p. 397-466.

- GONDA Jan, 1942, «Inwendige nasaal- en liquida-verbindingen in indonesische talen», *Bijdrage...* 101, p. 141-206.
- GONDA Jan, 1949, «Prolegomena tot een theorie der woordoorten in indonesische talen», *Bijdrage...* 105, p. 275-332.
- GONDA Jan, 1975, *Selected studies V. Indonesian linguistics*, Leiden.
- GRACE George W., 1959, «The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) family», Baltimore.
- HARRISON S. P., 1981, *Proto-Oceanic *aki(ni) and the Proto-Oceanic Periphrastic Causative*, The 3rd International Conference on Austronesian Linguistics, Bali.
- HAUDRICOURT André-Georges, 1954a, «Les origines asiatiques des langues malayo-polynésiennes», *Journal de la Société des Océanistes* 10, p. 180-183.
- HAUDRICOURT André-Georges, 1954b, Compte rendu de I. Dyen, The Proto-Malayo-Polynesian laryngals, *BSLP* 50/2, p. 134.
- HAUDRICOURT André-Georges, 1955, Compte rendu de I. Dyen, The Proto-Malayo-Polynesian laryngals, *Word* 11, p. 349-350.
- HAUDRICOURT André-Georges, 1965, «Problems of Austronesian comparative philology», *Lingua* 14, p. 315-329.
- HENRI Agnès, 2005, *L'élaboration d'un modèle théorique au fil des travaux de phonétique et de phonologie historiques dans le domaine austronésien*, Mémoire de Maîtrise, Université de Paris-Sorbonne.
- HYMAN Larry M. et Jan VORHOEVE (éds.), 1980, *L'expansion bantoue*, I, Paris, SELAF.
- LEMARÉCHAL Alain, 1986, «Syntaxe, morphologie et genèse de la forme dite "hypothétique" du palau», *Cahiers linguistiques Asie Orientale*, 15/1, p. 129-170.
- LEMARÉCHAL Alain, 1989, *Les parties du discours*, Paris, PUF.
- LEMARÉCHAL Alain, 1991, *Problèmes de sémantique et de syntaxe en Palau*, Paris, Editions du CNRS.
- LEMARÉCHAL Alain, 1997, «Superposition des marques, zéro et morphologisation», in *Mémoires de la SLP* 5 (Nouvelle série), Paris, Klincksieck, pp. 25-61.
- LEMARÉCHAL Alain, 1998a, *Etudes de morphologie en f(x,...)*, Paris, Peeters.
- LEMARÉCHAL Alain, 1998b, «Cliticisation vs autonomisation d'affixes: genèse des marques de voix et grammaire comparée des langues austronésiennes», in Cl. Muller, *Actes du Colloque sur les clitiques* (Bordeaux, 1998).
- LEMARÉCHAL Alain, 2001, «Problèmes d'analyse des langues de Formose et grammaire comparée des langues austronésiennes», *BSLP* 96/1, p. 419-480.
- LEMARÉCHAL Alain, 2003, «Hypothèses sur les marques de 2ème pers. de l'austronésien: PAN *Su = "2pl"», *BSLP* 98/1, p. 485-482.
- LEMARÉCHAL Alain, 2004a, «(Pré)histoires d'articles et grammaire comparée des langues austronésiennes», *BSLP* 100/1, p. 395-456.
- LEMARÉCHAL Alain, 2004b, «Les langues de Formose: "évoluées" ou "primitives"?», communication à la journée d'études sur «Les premiers austronésiens: langues, gènes, systèmes de parenté» (org. L. Sagart), le 5 mai 2004.
- LI Paul Jen-kuei, 1972, «On comparative tsou», *Bulletin of the Institute of History and Philology* 44, Taipei, Academia Sinica, p. 311-337.
- MEILLET Antoine, 1903, *Introduction à l'étude comparative des langues indo-européennes*, Paris (rééd. 1966, University of Alabama Press).
- MEILLET Antoine, 1927, *Linguistique historique et linguistique générale*, Paris, Champion.
- NOTHOFFER Bernd, 1975, *The reconstruction of proto-malayo-javanic*, S'Gravenhage, Nijhoff.

- ODÉ Cecilia et STOKHOF Wim, 1997, *Proceedings of the Seventh International Conference on Austronesian Linguistics*, Amsterdam, RODOPI.
- OGAWA Naoyoshi et ERIN ASAI, 1935, *Taiwan Takasagozoku Densetsu-shu* («Myths and Traditions of the Taiwan Formosan Native Tribes»), Taipei.
- PÄTZOLD Klaus, 1968, *Die Palau-Sprache und ihre Stellung zu anderen indonesischen Sprachen*, Berlin, Reimer.
- REID Lawrence A., 1971, *Philippine Minor Languages. Word Lists and Phonology*, Oceanic Linguistics Special Publ. n° 8, Honolulu.
- ROSS Malcolm D. (éd.), 1992, *Papers in Austronesian Linguistics n° 2* (= *Pacific Linguistics A-82*).
- ROSS Malcolm D., 1988, *Proto Oceanic and Austronesian languages of western Melanesia* (= *Pacific Linguistics C-98*), Canberra.
- SAGART Laurent, 1999, *The Roots of Old Chinese*, Amsterdam, Benjamins.
- SAPIR Edward, 1916, «Time perspective in aboriginal American culture» (= 1949, *Selected Writings of Edward Sapir* (éd. Mandelbaum), p. 389-462, Berkeley.
- SAUVAGEOT Aurélien, 1951, Compte rendu de M. Swadesh 1950, *BSLP XLVII/2*, p. 25-28.
- SCHACHTER Paul et F. OTANES, 1972, *Tagalog Reference Grammar*, Berkeley, University of California Press.
- STAROSTA Stanley, 1988, «A Grammatical Typology of Formosan Languages», *The Bulletin of the Institute of History and Philology* 59/2, Taipei, Academia Sinica.
- STAROSTA Stanley, 1995a, «A grammatical sub-grouping of Formosan languages», *Austronesian Studies*, p. 683-726.
- STAROSTA Stanley, A. K. PAWLEY et L. REID, 1982, «The evolution of focus in Austronesian», in S. A. Wurm et L. Carrington (éds.), *Papers from the 3rd International Conference on Austronesian Linguistics*, vol. 2 (= *Pacific Linguistics C-75*), Canberra.
- STEINHAUER H. (éd.), 1991, *Papers in Austronesian Linguistics n° 1* (= *Pacific Linguistics A-81*).
- STEINHAUER H. (éd.), 1996, *Papers in Austronesian Linguistics n° 3* (= *Pacific Linguistics A-84*).
- SWADESH Morris, 1950, «Salish internal Relationship», *IJAL* 216, p. 157-167.
- SWADESH Morris, 1952a, «Lexicostatistic dating of the prehistoric ethnic contacts», *PAPS*, 96, p. 452-463.
- SWADESH Morris, 1952b, Réponse au CR de Sauvageot 1951, *BSLP XLVIII*, p. 1-5.
- SWADESH Morris, 1955, «Towards a greater accuracy in lexicostatistic dating», *IJAL* 21, p. 121-137.
- SWADESH Morris, 1959, «The Mesh principle in Comparative linguistics», *Anthropological Linguistics*, 1, p. 7-14.
- TEEUW A., 1959, «The history of the Malay language», *Bijdrage tot de taal-, land- en volkenkunde van Nederlandsch-Indië* 115, p. 138-156.
- TRYON Darell T. (éd.), 1995, *Comparative Austronesian Dictionary I-V*, Berlin, Mouton de Gruyter.
- TSUCHIDA Shigeru, 1971, *List of words of Formosan Languages*.
- TSUCHIDA Shigeru, 1976, *Reconstruction of Proto-Tsouic phonology*, Tokyo.
- VLIS C. J. van der, 1842, «Formosaansche Woorden-Lijst volgens een Utrechtsch Handschrift», *Verhandelingen van het Bataviaasch Genootschap*, 18, p. 437-488.

- WU Jing-lan, 1995, *Complex sentences in Amis* (Master thesis), National Taiwan Normal University.
- ZORC R. David, 1982, «Where, o where, have the laryngals are gone? Austronesian laryngeals reexamined», in Amran Halim, Lois Carrington et Stephen A. Wurm (éds.), *Papers from the Third Conference on Austronesian Linguistics*, II, p. 111-144.
- ZORC R. David, 1990, «The Austronesian monosyllabic root, radical and phones-theme», in Baldi (éd.), *Linguistic change and reconstruction methodology*, Berlin, Mouton De Gruyter, p. 175-104.

Alain LEMARÉCHAL
Université de Paris-Sorbonne
et LACITO (CNRS)
80 rue des Archives, 75003 Paris