

Malagasy Dialect Divisions: Genetic versus Emblematic Criteria

Alexander Adelaar

UNIVERSITY OF MELBOURNE

This paper gives an overview of the literature on Malagasy dialect variety and the various Malagasy dialect classifications that have been proposed. It rejects the often held view that the way Malagasy dialects reflect the Proto-Austronesian phoneme sequences *li and *ti is a basic criterion for their genetic division. While the linguistic innovations shown in, respectively, central dialects (Merina, Betsileo, Sihanaka, Tanala) and southwestern dialects (Vezo, Mahafaly, Tandroy) clearly show that these groups form separate historical divisions, the linguistic developments in other (northern, eastern, and western) dialects are more difficult to interpret. The differences between Malagasy dialects are generally rather contained and do not seem to be the result of separate migration waves or the arrival of linguistically different migrant groups. The paper ends with a list of subgrouping criteria that will be useful for future research into the history of Malagasy dialects.

1. INTRODUCTION.¹ This paper investigates some of the early linguistic changes that have contributed to the dialect diversity of Malagasy, as well as the various classifications that have been proposed for Malagasy dialects.

Malagasy is an Austronesian language directly related to some of the languages spoken in Central Kalimantan Province and South Kalimantan Province in Indonesian Borneo. Together with these languages, it forms the South East Barito (henceforth SEB) subgroup, which belongs to the Malayo-Polynesian branch of the Austronesian language family.

A historical classification of dialects or languages should be based in the first place on the oldest linguistic changes that have happened in the dialect or language group in question. These changes are not necessarily the most salient ones from a typological perspective: later changes can be just as emblematic and may have occurred in large numbers, to the point that historical changes may have become opaque or may have been wiped out

1. Research for this paper was supported by a Discovery Grant (DP120100390) from the Australian Research Council. An earlier version was presented at the 12th International Conference of Austronesian Linguistics in Bali, Indonesia, July 2–6, 2012. In writing this paper, I benefited a great deal from the generous help and vast expertise of Emeritus Professor Noel Gueunier (Paris). I am also grateful to Dr. Anthony Jukes (Centre for Research on Language Diversity, La Trobe University, Melbourne), Professor Waruno Mahdi (Fritz Haber Institute of the Max Planck Society, Berlin) for their careful reading of earlier drafts of this article, and to Mrs. Chandra Jayasuriya (University of Melbourne's cartographer) for preparing maps. The usual disclaimers apply.

altogether by these later changes. However, in an attempt to classify historically the members of a linguistic group, it is necessary to start at the very beginning. This is done in this paper by investigating the earliest forms of dialect divergence within the group. Later on, subsequent changes that occur in already formed dialects and that are due to borrowing, areal features, and the like will have to be investigated using methodology that is more specifically dedicated to dialect study.

In section 2, I present conventions, abbreviations, and sources. I give a summary and critical evaluation of previous Malagasy dialect studies in section 3, and present possible further criteria for the classification of Malagasy dialects in section 4. In section 5, I make concluding remarks and show the relevance of Malagasy dialect classification for the settlement history of Madagascar.

2. SOURCES, CONVENTIONS, AND ABBREVIATIONS. Publications dealing with Malagasy dialect variety generally recognize the North, Southwest, West, East, and Center of Madagascar as major areas of dialect variation, based on the commonalities that the dialects in these areas are perceived to share (see map 1).

For ease of reference, I will use this areal division as a preliminary classification, although the historical-genetic status and delineation of this division remains to be tested, especially concerning the internal coherence of the eastern dialects, and the position of western dialects in relation to southwestern ones.

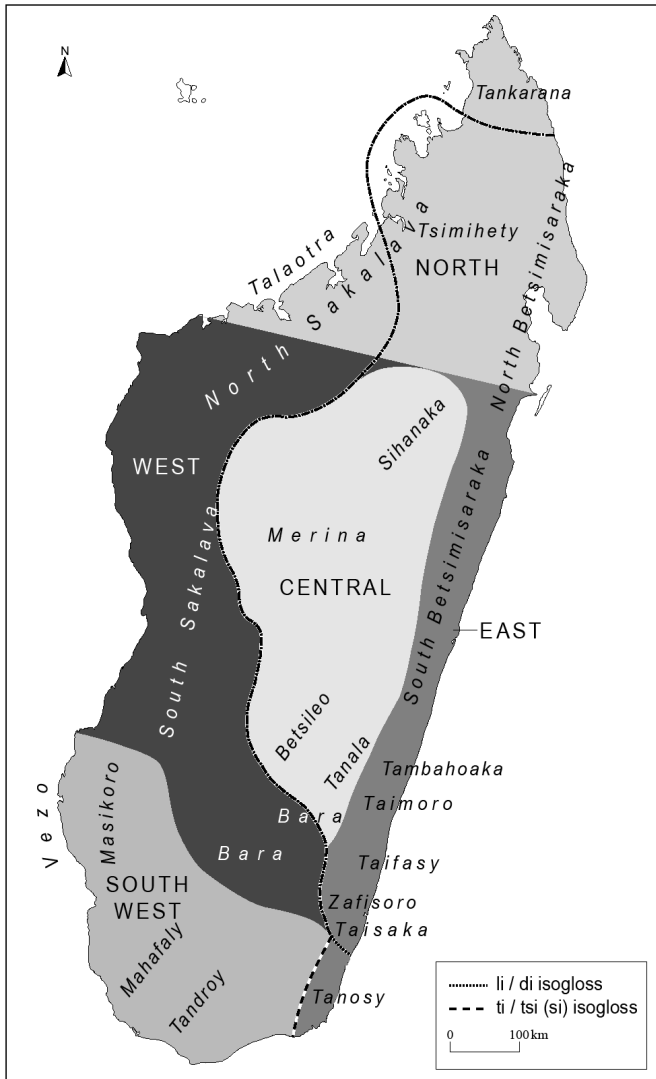
Map 2 shows the Malagasy dialects that are referred to in these pages. The western area includes dialects in the central Sakalava region as well as western varieties of Bara. The southwestern area includes (southern) Vezo, Masikoro, Mahafaly, and Tandroy. Northern dialects consist of Tankarana, Tsimihety, northern Sakalava (spoken to the north of Majunga town), and some northern varieties of Betsimisaraka. Merina, Sihanaka, Tambahoaka, Betsileo, Tanala, and eastern Bara belong to the central dialect area. Betsimisaraka dialects along the central and southern parts of the east coast, as well as Taimoro, Tanosy, Taisaka, and Zafisoro, all belong to the eastern area. Finally, Madagascar's oldest literature, written in Sorabe (an adaptation of the Arabic script), also originated in East Madagascar and its language is an early form of coastal South East Malagasy (henceforth Old South East Malagasy).²

The language data used in this paper are taken from published and unpublished written sources. For some of the dialects (Merina, Sakalava, Bara, Tandroy), there are good dictionaries available, but for many other dialects the sources are incomplete. In the case of Betsimisaraka, there are various subdialects, and the dictionaries, which are several centuries old, are phonologically imprecise and tend to represent unsystematic and rudimentary information. For other dialects, such as Tsimihety and Velonandro's (1983) Northern Malagasy, the dictionaries in question focus more on providing data contrasting these dialects with other forms of Malagasy than on presenting the lexicon that is frequently used in daily language. As a result, they contain many unusual entries, but often lack the basic vocabulary needed to test the lexical, semantic, and phonological history of

2. The language in Sorabe texts is an old form of coastal South East Malagasy. It is often incorrectly assumed to be an old form of Taimoro (including by the present author in previous publications). The association with Taimoro Malagasy is based not on linguistic criteria but on the fact that the use of the Sorabe script is nowadays limited to the Taimoro region (Noel Gueunier, pers. comm.).

a given dialect. (Moreover, Velonandro [1983] includes words from a wide range of dialects that are not labeled according to their provenance, and is therefore not appropriate for dialect comparison). In other cases again (for example, Taimoro), there are no dictionaries at all, and I collected data from wordlists or written texts instead. For Old South East Malagasy, I used data from a sixteenth century text, which was transcribed, translated, and annotated by Ferrand (1904).

MAP 1. MAJOR MALAGASY DIALECT DIVISIONS



**MAP 2. MALAGASY DIALECTS AND TOPONYMS
REFERRED TO IN THIS PAPER**



In practice, sources are not always adequate, and I sometimes had to consult a number to be able to double-check and obtain a fuller picture of a particular phenomenon. This occasionally lead to finding contradictory data due to the fact that some dialect labels designate more than one dialect variety. This is because many assumed Malagasy dialect

names are not linguistically defined. They are in fact geopolitical or ethnic terms, which is a cause of much confusion. The terms “Sakalava,” “Bara,” “Betsimisaraka,” and even “Vezo” each designate very different speech forms depending on the location where they are spoken. In many cases, the speech forms referred to are not just regionally different but also belong to different major dialect divisions. For instance, Sakalava spoken in Morondava is a western Malagasy dialect, whereas Sakalava spoken north of Majunga belongs to the northern division and is more similar to Tsimihety and Tankarana (see below), while Sakalava spoken south of Morondava is a southwestern dialect that is more appropriately called Masikororo. In what follows, I use the term “Sakalava” specifically for the dialect spoken in Morondava and thereabout.

I write Malagasy language data in the original orthography used in the sources, except for the following changes: *o* is replaced by *u* and final *y* by *i*, paragogic vowels (which are whispered vowels) are marked with a breve (*ă, ě, ĭ*, and so on), and velar nasals are rendered as *ŋ*. Stress is indicated wherever the original sources do so (basically, it is indicated in the Merina, Sakalava, Vezo, and Tandroy dictionaries). Exceptions to these spelling conventions are data from Houtman (1603) and Flacourt (1658), as well as toponyms and dialect names, which I left in their original spelling. *N* refers to a nasal when it occurs with a hyphen attached to it (*-N, -N-*, and so on).-

I present names of dialects without the *aN-* prefix with which they are often found: thus I write Taimoro, Tankarana, Tandroy, Tanosy instead of Antaimoro, Antankarana, Antandroy, Antanosy, and so on, as is also found in the literature.

I add between parentheses the letter ‘R’ or ‘F’ followed by a numeral in some cases where language data are not taken from dictionaries or alphabetically ordered lists: this code stands for a source and the relevant page number, for example, ‘(R48)’ stands for Ruud (1955:48), ‘(F114)’ for Ferrand (1904:114).

Proto-Austronesian etyma are given in their Proto–Malayo–Polynesian (PMP) equivalent forms, as these are closer to their Austronesian reflexes and thus easier to interpret.

Unless indicated otherwise, the language sources used in this paper are those listed in the appendix. I use ‘id.’ to mean ‘same meaning as previous foreign word’.

3. PREVIOUS STUDIES DEALING WITH MALAGASY DIALECT VARIETY AND CLASSIFICATION. There are several studies dealing with Malagasy dialect variation. Although these generally agree on distinguishing the North, Southwest, West, East, and Center of Madagascar as major areas of variation, they differ in their interpretations of how these areas relate to one another from the perspective of a basic genetic classification. In what follows, I give a summary and critical evaluation of these studies, treating them in chronological order.

3.1 THE EARLIEST SOURCES ON DIALECT VARIETY. From the nineteenth century onward, various initiatives were taken to map the dialect variety of Malagasy, beginning with a list commissioned by Queen Ranavalona I, one of Madagascar’s monarchs, who ruled from 1828 to 1861. These attempts at making a dialect atlas were usually based on lexical data only. Some of the resulting data collections were never published; others were published as supplements to linguistic reference works. While these

early collections are of great value for the study of the history and spread of Malagasy dialects, they never became the basis of a systematic analysis. Noel Gueunier and Solo Raharinjanahary (2011) present an overview. Gueunier and Raharinjanahary are also preparing a very detailed and systematic dialectological study of the southern part of Madagascar (Gueunier and Raharinjanahary to appear).

3.2 JACQUES DEZ'S DIALECTOLOGICAL STUDY. It is fair to say that modern Malagasy dialectology started with the work of Jacques Dez (1963a,b,c). This author proposed a dialect classification on the basis of phonological, lexical, and morphological differences that he found in the various regional forms of Malagasy. He concluded that there was a major division between southwestern and western dialects on the one hand, and central and eastern dialects on the other (see figure 1). This was most typically exemplified in the phonological histories of these dialects.

Most critical for the demarcation of an eastern/central and western/southwestern division is the way dialects reflect the sequences **li* and **ti*, namely as *li* and *ti*, respectively, in southwestern and western dialects, and as *di* and *tsi* (or *si*), respectively, in central and eastern dialects. Compare the cognate sets in table 1, in which Merina, Betsimisaraka, Sakalava, Tandroy, and Tankarana each represent a major dialect area.

The east/central versus west/southwest division also appears to some extent in the way dialects reflect PSEB and PMLG final **t*. This generally became a retroflex affricate *-tr* in eastern and central dialects and a palatal affricate *-ts* (in both cases followed by a paragogic vowel) in western and southwestern dialects. However, here the picture is

FIGURE 1. MALAGASY DIALECTS PROPOSED BY DEZ (1963a,b,c,d)

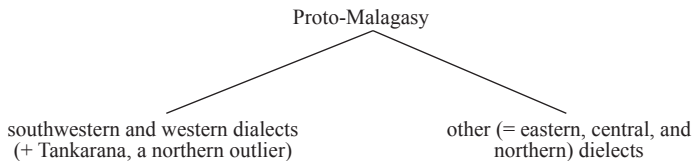


TABLE 1. **li* AND **ti* REFLEXES ACROSS MALAGASY DIALECTS

PMP	Central (Merina)	East (South Betsimisaraka)	West (Sakalava)	Southwest (Tandroy)	North (Tankarana)
*li:					
*lima 'five'	dimi	dime (R47)	lime	lime	dimi
*kulit 'skin'	hùditra	—	hùlitse, hùlitsi	hùlitse	huditri
*bøli 'to buy'	vidi 'value'	vua-vide (R52) 'bought'	vili 'value'	vili 'id.'	?
*ti:					
*teyup 'to blow'	tsiuka 'wind'	—	tiuke 'wind'	tiuke 'id.'	tsiku 'id.' (< *tsiuku)
—	ràtsi 'bad'	?	ràti	ràti	ratsi, rati
*putiq 'white'	fùtsi	fuse (R53)	fùti	fùti	futsi
*ha-timuR 'dry monsoon' [†]	atsimu	atsimu (R47)	atimu, antimu	atimu	?

[†] Meaning 'South' in all dialects.

complicated by Betsimisaraka, which represents several distinct subdialects, some of which exhibit *ts* in old European sources. Furthermore, various Malagasy dialects exhibit both final *tr* and *ts* (although one of these consonants usually predominates). See table 2 for examples.

3.2.1 The *li* : *di* and *ti* : *tsi* distinctions. Dez's phonological criteria may seem emblematic for a major division between southwestern and western dialects on the one hand, and central and eastern dialects on the other (with northern dialects falling on both sides of the dividing line), but they are not critical for the following reasons.

A methodological problem is that whereas *di* (< *li) and *tsi* (< *ti) represent sound changes, *li* and *ti* do not. Therefore, only *di* and *tsi* represent innovations that can be used as subgrouping evidence. (In the present case, they suggest that the central and eastern dialects form a subgroup within the overall Malagasy dialect variation). In contrast, the original sequences *li* and *ti* do not have any subgrouping value as they represent the original phonological situation of PMLG (as well as Proto-SEB and even PMP). This point has already been made by Pierre Simon in relation to a possible subgroup including western and southwestern dialects; see 3.6).

From a distributional perspective, Dez's dividing line is fuzzy in places. One thing that becomes immediately obvious is that it divides Madagascar into three areas rather than two. Tankarana in the northern tip of the island, a dialect presumably exhibiting *li* and *ti*, is cut off from other dialects that have retained these reflexes. Its outlier position is in itself hardly alarming, but further investigation shows that this dialect exhibits *di* and *tsi* instead of the expected *li* and *ti* in almost all instances. Compare the instances of *li and *ti reflexes in the three Tankarana lists compiled by Serva and Petroni (2011) given in table 3.

Of all nine instances reflecting *li, only *liu/li* 'blood' has maintained it as such, and of all eight instances reflecting *ti, only *rati* and *mi-tilinjī* (which occur in only one subdialect each) have maintained *ti*. We can only surmise that, at the time Dez wrote his dialect study, the forms containing *di* and *tsi* were not yet widespread in Tankarana, and if they are now, this must be the result of lexical replacement and/or influence from standard Malagasy pronunciation. There is positive evidence for this in at least one case: Vérin, Kottak, and Gorlin (1969) registered *atika* (not *atsika*) for the 1st person plural inclusive.

TABLE 2. FINAL *t REFLEXES ACROSS MALAGASY DIALECTS

English	Central (Merina)	East (Betsimisaraka)	West (Sakalava)	Southwest (Tandroy)	North (Tankarana)
four	éfatrā	efara (R47) [†]	éfatsē	éfatsē	efatrā
to steal	mangālatrā	mangalar (R55)	mangālotsē	mangālotsē	—
strong wind	rivutrā	—	rivutsē, rivutsī	rivutsē	rivutrū
wish	sāutrā	mi-saur (R54)	sāutsē 'thanks'	sāutsē 'a wish'	—
fear	tāhutrā	—	tāhutsē, tāhutsī	tāhutsē	ma-tahutrū
sky	lānitrā	—	lānjtsē, lānjtsī	lānjtsē	lānjitrī
skin; bark	hūditrā	—	hūlitsē, hūlitsī	hūlitsē	huditri
state, condition, character	tūetrā	i-tuer (R50)	mi-tūetsē, mi-tūetsī 'stay, live'	tūetsē	man-tutri 'sit' (< man-tūetri)

[†] Although Ruud (1955) writes final *r* (sometimes *rr*), from his own description (on p. 35) it appears that this sound must be a retroflex affricate, much like the corresponding consonant in Merina.

More telling evidence comes from Tsimihety, another northern dialect. According to Dez, it reflects *li and *ti as *di* and *tsi*, respectively, and therefore has a position to the right of his dividing line. However, under close inspection it appears that inherited vocabulary with *li* and *ti* is still extant in Tsimihety: table 4 shows that some words reflect *li and *ti as *li* and *ti*, whereas other words reflect them as *di* and *tsi*, and there are also doublets contrasting *li* with *di* and *ti* with *tsi*, with or without different meanings.

I found no Tsimihety doublet forms contrasting *ti* and *tsi* in Faridanona (1977). However, this source is not complete. Among other gaps, it fails to include instances of *li > *di* such as *dimi* ‘five (compare Merina *dimi*, PMP *lima ‘id.’) and *vidi* ‘price’ (Merina *vidi* ‘id.’, PMP *bəli ‘to buy’), and instances showing doublets with *ti* and *tsi* for *ti such as *tiŋànana*, *atsiŋànana* ‘east’ (Merina *antsinànana*, Sakalava *atiŋàna* ‘id.’) and *atimu*, *atsimu* ‘south’ (Merina *atsimu* ‘id.’, PMP *qatimuR ‘wet monsoon’).³

While Tsimihety has forms with *li* and *ti* as well as with *di* and *tsi*, it seems that forms with *li* and *ti* have more idiosyncratic meanings, whereas those with *di* and *tsi* are more similar to their Merina counterparts in terms of meaning and valency. The latter are therefore suspect of borrowing. An interesting case is Tsimihety *tinàini* ‘Monday’, a loanword from Arabic. Its first consonant was originally a voiceless fricative (compare Arabic [θnayn]), but this was adopted into Malagasy dialects as a stop.⁴ It has remained a stop

TABLE 3. *li AND *ti REFLEXES IN TANKARANA SUBDIALECTS

PMLG	Tankarana (Ambiloŋa)	Tankarana (Antalaha)	Tankarana (Voŋemar)
*-ali ‘to fight’	mi-adi	mi-adi	mi-adi
*atuli ‘egg’	atudi	atudi	atudi
*-hali ‘to dig’	man-gadi	man-gadi	man-gadi
*hulit ‘skin’	uditri	huditri	huditri
*iime ‘five’	dimi	dimi	dimi
*-lili ‘to cut’	man-didi	man-didi	—
*liu ‘saliva’	liu ‘blood’	li ‘id.’	liu ‘id.’
*tali ‘rope’	tadi	tadi	tadi
*wali ‘spouse’	vadi	vadi	vadi
*puti ‘white’	—	futsi	—
*rati ‘bad’	rati	ratsi	ratsi
*ti ‘no(t)’	tsi	tsi	tsi
*tikkaN ‘we (INCL)’	atsika	atsika	atsika
*-tiliŋ ‘to fly’	—	mi-tiliŋi	—
*tinai ‘guts’	—	tsinai	—
*tiu[p] ‘wind’	tsikū	—	tsikū
(Malay buah betis) ‘calf of leg’	—	vu/vitsi ‘leg’	—

3. These words were kindly given to me by Professor Hideo Fukazawa from the Department of Anthropology at Tokyo University of Foreign Studies. He has worked and lived in northwest Madagascar for many years and is a fluent speaker of Tsimihety.

4. It is unclear why the original Arabic dental fricative first became *t in Malagasy dialects. It may have been borrowed from an intermediate language in which the fricative had become a stop. A large number of Arabic loanwords came into Malagasy via Kiswahili, and one may be tempted to see the latter as the immediate source language for *tinàini* and *tinèni*. However, tracing these terms to Kiswahili is problematic because Kiswahili does not use Arabic names for days of the week (Adelaar 2009).

TABLE 4. TSMIHETY *li AND *ti REFLEXES (FARIDANONA 1977)

Historical		Tsimihety	Merina
*li > li:			
PSEB *tilij 'to fly'	>	tilij 'id.'	tsidina 'id.'
PMP *liaR 'wild'	>	li 'wild, shy'	dia 'id.'
PMP *liuR 'saliva'	>	liu 'blood'	—
*li > di:			
PMP *uliq 'to get, obtain'	>	a/üdi 'remedy, charm'	üdi 'id.'
Malay sulij 'flute'	>	an/tsüdiñi 'id.'	südina 'id.'
PMP *um-uliq 'to return home'	>	müdi Zañahäri '(to go back to God =) to die'	müdi 'to go back home'
*li is reflected in doublet forms contrasting li and di:			
PMP *talih 'rope'	>	tädi 'rope, rope-like thing', täli 'plait, braid, tress'	tädi 'rope'
PMLG *wali 'spouse; member of spouse's family'	>	vädi 'spouse', vali-lähi 'sib-ling-in-law'	vädi 'id.'
Malay pali, pämal 'prohibition'	>	fädi 'id.' bibi/fäli, bibi/fäli 'big animal, black in front but with a yellowish hind part'	fädi 'id.'
*ti > ti:			
PSEB *tilij 'to fly'	>	tilij 'flight of birds'	tsidina 'to fly'
Malay tingi 'high'	>	tiñgi 'mons veneris(?) [†]	tsiñgi 'mountain top'
—		tipakä 'a kick with the leg'	tsipakä 'id.'
Arabic (yaum al-)iñnain 'Monday'	>	tinäini 'id.'	tsinäini 'id.'
*ti > tsi:			
PMP *punti 'banana'	>	füntsi, üntsi 'id.'	—
PMP *heyup 'to blow'	>	tsiuku 'wind'	tsiukä 'breeze, wind'
Malay buah bätis 'calf of leg'	>	vavitsi 'id.'	vuavitsi 'id.'
Javanese patek 'pin, peg'	>	fätsiki 'thorn, spike'	fätsikä 'id.'
PMLG *tika-(i)kam	>	atsika, atsikana 'IPL.INCL'	isika 'id.'

[†] In *fädi-n-drahütsikitsiki* 'prohibition observed in part (like drinking the stock but not eating the meat)'.
[‡] Faridanona has *ny maha-vehivavy* referring to the female genitals.

up to the present in Tsimihety, but changed into an affricate in Merina *alatsinaini* 'Monday' (which also maintained *ala-*, a reflex of the Arabic article *al-*). An example like this seems to indicate that, at one point in time, *ti without affricatization was the default pronunciation and was felt as more emblematic than *tsi* in Tsimihety; the latter pronunciation may have been associated with outside influence from central and eastern dialects. Furthermore, the *ti* pronunciation was maintained until very recently, and the affricatization in *tsi* is a relatively new development. In summary, there are strong indications that historically Tsimihety (and probably also the more northern dialect Tankarana) followed the southwestern dialects in maintaining *li and *ti, and that instances of *di* and *tsi* reflexes in these dialects are the result of contact with Merina or other mainstream dialects that have undergone the *li > di and *ti > tsi changes.

Another fuzzy area exists on Madagascar's South East coast. Here, the changes in question do not occur in tandem. According to Dez's main dialect isoglosses, Tanosy on Madagascar's extreme southeast coast is the only dialect showing a change from *ti to tsi while maintaining *li as li (see map 1). This situation is also reflected in other sources

representing southeastern Malagasy dialects, namely in Flacourt's (1658) dictionary (which is also a Tanosy source as it was compiled in the region around Fort-Dauphin), and in the Old South East Malagasy texts.⁵ Examples are given in table 5.

However, in contrast to these sources, modern spoken South East Malagasy dialects have generally undergone both changes, and *li is realized as *di*.⁶ Simon (1988:267) suspects that Old South East Malagasy *li* is not a retention but a spelling convention due to northern Malagasy influence. However, two factors militate against this interpretation.⁷ One is Dez's observation that *li remained *li* in Tanosy, suggesting that, rather than a spelling convention, we may be dealing with a feature in the spoken language. The other factor is that all other Malagasy dialects that have maintained *li* have also maintained *ti*, and the northern dialects are no exception to that (allowing for their occasional doublets with *di* and *tsi*, which must be due to borrowing). So, if Old South East Malagasy writers had modeled their spelling of *di* after the (historical) pronunciation of a northern Malagasy dialect, they would in all probability also have done so for *ti and written it as *ti* (without affricatization), which is not the case.

In a lexically very frequent form such as the first person plural inclusive pronoun (Merina *isika*, Sakalava *tsika* < *i + *tik(k)aN),⁸ affricatization of *ti is also observed in many western and southwestern dialects. In fact, in Serva and Petroni's (2011) comparative wordlists, it is shown in all dialectal forms except in Tandroy *itikanje*. (However, compare Vêrin, Kottak, and Gorlin [1969] who list Tankarana and Vezo *atika*, Mahafaly *tika*, Tandroy *tika(ŋe)*). Although the retention of PMP *ti as *ti* is particularly distinctive in southwestern Malagasy dialects, even these dialects occasionally show *tsi*. Or, in a

TABLE 5. *li AND *ti REFLEXES IN TANOSY AND OLD SOUTH EAST MALAGASY

Merina	Old South East Malagasy	Tanosy	Flacourt
*li > di :			
hadinu 'forgotten'	haliŋu 'id.' (F115)	—	talinzou 'shadow'
hùditrā 'skin'	hulitri 'id.' (F87)	hulisi	houlitse 'skin'
tādi 'rope'	talimpali 'bow string' (F115)	tali 'rope'	tali 'id.'
tadini 'ear'	taliŋi 'id.' (F88)	tadiŋi	talinhe 'id.'
*ti > tsi :			
atsimu 'South'	—	—	adcimou 'id.'
futsi 'white'	futsi 'id.' (F86)	fusi	foutchi 'id.'
PMP *punti 'banana'	afuntsi 'id.' (F112)	—	fontsi 'id.'
*ti > Merina tsi 'no(t)'	tsi 'id.' (F114)	tsi 'id.'	tsi 'id.'
vuavitsi 'calf of leg'	vuavitsi 'id.' (F92)	—	—

5. As already noted in Beaujard (1998b:7).

6. Here again, Dez's information needs qualification: although classificatorily ("historically") correct, it does not take into account the many forms with *di* and *tsi* in current Tanosy (as reflected in Serva and Petroni's wordlists, for example). The latter must be due to interference from standard Malagasy and/or other dialects.

7. Simon's suspicion is most likely based on claims made by members of the Zafiraminia, Onjatsy, and Taimoro ruling classes that they hail from the island's northeast. I am not contesting these claims, but I have as yet found no indications of the maintenance of *li (and a concomitant change of *ti to *tsi* or *si*) in northern (or northeastern) forms of Malagasy.

8. Adelaar and Kikusawa (in preparation).

more generalizing way, the change from *ti to *tsi* affected all dialects but has progressed farthest in the central and eastern ones, and least in the southwestern ones.

Finally, as shown above, published sources also show variation in exhibiting *li* : *di* and *ti* : *tsi*, *si*. For the first person plural inclusive in Tankarana, Vérin, Kottak, and Gorlin (1969) have *atika*, but Serva and Petroni (1983) as well as Velonandro (1983) have *atsika*, which may be due to geographical variation within the Tankarana region or, as discussed above, to variation over time. Influence from the hegemonic Merina dialect no doubt plays a role in the spread of *di* and *tsi*.⁹

In conclusion, the arguments presented above clearly demonstrate that a major classification of Malagasy dialects based on the distinctions between *li* versus *di* and *ti* versus *tsi* (*si*) such as proposed by Dez must be rejected on methodological and distributional grounds.

Note also that Simon (1988:222, 239) made the point that the *li > *di* and *ti > *tsi* changes proposed by Dez are not innovative because they had already taken place long before in some East Barito languages, including Maanyan. He agrees with Dahl (1951:111) that these changes were “recurrent” and that they happened several times and in many places independently. However, rather than considering these changes to be recurrent, one could also think of a sustained period of phonological variation between two forms, sooner or later resulting in one of the two forms becoming permanent in dialect forms. The *li > *di* and *ti > *tsi* changes must have existed as tendencies for a long time, starting before the split of Proto-East Barito into PSEB and Proto-Central East Barito, and lasting until after the present-day Malagasy dialects came into being.¹⁰ While this may seem to be a long time, the pattern shown here is not essentially different from that of most sound changes in action. Similar processes are not unusual, although in actual practice we are not able to see them taking their full course. Within the context of the dialect variety in Madagascar, an island far away from the homeland in Borneo, we can still consider their outcomes as innovations.

3.2.2 The distinction between final *tr*(V) and final *ts*(V). Dez’s other phonological evidence, the contrast between final *tr* and final *ts*, does not suffer from the same methodological drawback as the *li and *ti cases, as both *tr* and *ts* are innovative in final position. However, the problem here is that, while Ferrand (1905) has *tr* and Ruud has *-r* (which in all likelihood stands for a retroflex affricate *tr*; see Ruud 1955:35), the older sources for Betsimisaraka show the (presumably palatal affricate) *ts* reflex. Compare the data in table 6.

TABLE 6. OCCURRENCE OF FINAL *tr*(V) AND *ts*(V) IN BETSIMISARAKA SOURCES

English	Merina	(North) Betsimisaraka (Houtman 1603)	Tanosy (Flacourt 1658)	(South) Betsimisaraka (Ferrand 1905)
four	èfatrà	effas, effats	effats	efatrà
skin	hùdirà	hoedits	oulits, houlitse	holitrà
fear	tàhutrà	mattahots ‘afraid’	tahotse	tahotrà
finished	tàpitrà	tapits ‘end’	tapets, tapere ‘end’	tapitrà ‘finished’
to teach	mànanatrà	mang’anjarts	manganats	mànanatrà

9. It is possible that this dialect is also the original center of dispersal of these changes, but this is a matter for further investigation.

10. An explanation that in its essence was also considered by Dahl (1983:21).

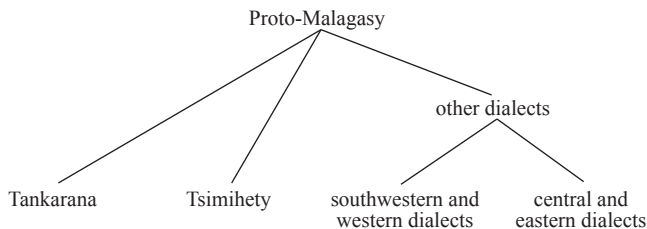
Ferrand (1905:14) assumed that this final *ts* stood for the same retroflex affricate as final *tr* in other central and eastern dialects, and that previous authors had used it for want of a proper symbol for the retroflex affricate. Nevertheless, the fact that Flacourt did distinguish between *ts* and *tr* in nonfinal positions militates against this assumption.¹¹ Moreover, other dialects (including Old South East Malagasy and the northern dialects) exhibit both final *-tr* and *-ts*, with a preponderance of the former. An obvious additional problem is that Betsimisaraka is not a uniform dialect. It is spoken in a very long stretch of land along the east coast and represents a dialect variety that is poorly understood. The matter remains unclear.

3.3 VÉRIN, KOTTAK, AND GORLIN'S (1969) LEXICOSTATISTICAL APPROACH. A second systematic dialectological survey by Pierre Vérin, Conrad Kottak, and Peter Gorlin (1969), made use of lexicostatistics, a method much in vogue in the 1960s. On the basis of Swadesh's 100-wordlist, it distinguished three primary subgroups: (1) Tsimihety, (2) Tankarana (both in North Madagascar) and (3) all other forms of Malagasy. The latter can again be classified into (3a) southwestern and western varieties and (3b) central and eastern varieties (figure 2).

While Vérin, Kottak, and Gorlin (1969) are meticulous in their methodology and present many interesting insights, their conclusions are rather weakened by a number of facts: the database for their research is very small (100 words); they published at a time when Malagasy dialects were much less well documented than they are today; and they based their study on lexicostatistics, a method flawed by its inability to distinguish between inherited and borrowed vocabulary as well as between retentions and innovations.

3.4 MAHDI (1988). Waruno Mahdi (1988:106–22) combines Vérin, Kottak, and Gorlin's Malagasy dialect data with Hudson's (1967) wordlists of Barito languages in order to reach a classification of all languages involved based on lexicostatistics. His calculations show a tripartite division of Malagasy dialects into a northern branch, Tankarana and Tsimihety; a western branch, including Sakalava, Vezo, Mahafaly, and Tandroy; and an eastern branch, including Merina, Betsileo, Betsimisaraka, Sihanaka, Tamba-hoaka, Zafisoro, Taimoro, and Taisaka (Mahdi 1988:117). The western branch is further divided into Vezo, a group consisting of various forms of Sakalava, and a (southwestern)

FIGURE 2. MALAGASY DIALECTS PROPOSED BY VÉRIN, KOTTAK, AND GORLIN (1969)

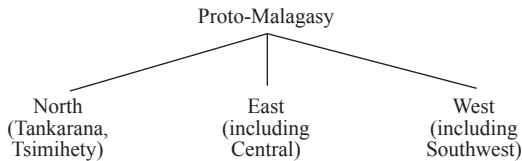


11. For instance, 'to build a house' is rendered as *minboutse tranghe* in Flacourt (1658) and as *mambuatrà traju* in Ferrand (1905).

subgroup made up of Tandroy and Mahafaly; the eastern branch is divided into a central group (Merina, Sihanaka, Betsileo), an eastern group (Betsimisaraka, Tambahoaka), and a southeastern group (Taimoro, Taisaka, Zafisoro). See figure 3 for a diagrammatic representation of Mahdi's classification.

In his application of lexicostatistics, Mahdi (1988) is aware of the distorting effects of language contact, and tries to account for the loanwords that he is able to spot. However, his classification of Malagasy dialects is largely based on the limited data provided by V erin, Kottak, and Gorlin (1969), and it is (at least partly) subject to the same criticism.

FIGURE 3. MALAGASY DIALECTS ACCORDING TO MAHDI (1088)

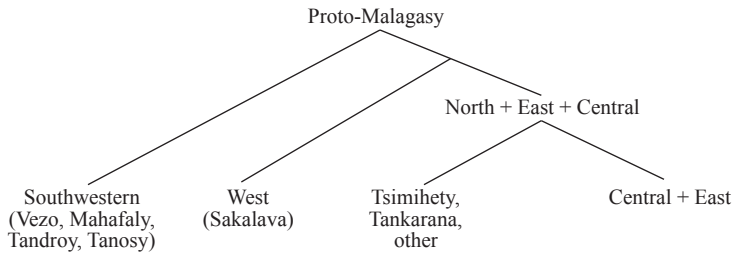


3.5 GUEUNIER'S (1988) DIALECT STUDY OF MALAGASY ON MAYOTTE (COMOROS). Noel Gueunier (1988) also uses Malagasy dialect geography but only in order to determine the history and affiliations of the two Malagasy varieties spoken on the island of Mayotte in the Comoros. While making explicit some of the problems involved in Malagasy dialectology, he does not venture into an overall classification of Malagasy varieties.

3.6 SIMON'S (2006) CLASSIFICATION BASED ON QUANTITATIVE AND QUALITATIVE EVIDENCE. Pierre Simon (2006:183–202) proposes a basic division between southwestern Malagasy and other varieties, which are divided again between western Malagasy (Sakalava, West Bara) and all other varieties; finally, the latter are split between northern varieties (Tsimihety, Tankarana) and central-eastern varieties (including Merina, Betsimisaraka, and Taimoro); see figure 4. Simon bases his classification on V erin, Kottak, and Gorlin's lexicostatistical calculations augmented with additional phonological, grammatical, and lexical criteria. His Southwestern subgroup includes Vezo, Mahafaly, Tandroy, and Bara, and is based on the following exclusively shared features:

1. phonetic values reflecting a more limited phoneme system; Vezo represents the simplest version of this system, missing the retroflex series present in other dialects;
2. maintenance of a glottal stop (more particularly, in the dependence marker *ʔe*, which is absent in other dialects);
3. a demonstrative pronoun *i* (**ihi*);
4. the article *te* (instead of *ʔe* or *ni* elsewhere);
5. the occurrence of *ʔzu* '2SG oblique pronoun' instead of *anau* (*ʔanau*) in other dialects; and
6. many roots ending in *-k e* instead of reflecting expected final *-N (> Merina *-na*, Sakalava *-∅*, and so on.).

FIGURE 4. MALAGASY DIALECTS ACCORDING TO SIMON (1988)



Simon remains unsure about parts of this classification. The position of western varieties (Sakalava) and northern varieties (Tsimihety and Tankarana) are difficult to ascertain. Western varieties may also turn out to subgroup with southwestern ones, although the latter are definitely closer to one another than to Sakalava, and the agreement in maintaining *li and *ti in both western and southwestern dialects provides no guidance for subgrouping because it is a retention from PMP (Simon 2006:189). Northern varieties could also be an independent primary branch as opposed to two other branches, one including the western and southwestern varieties, and another including all eastern and central varieties. Simon tends to favor an original North-South opposition with Tsimihety and Tandroy at the extremes, a primary genetic split that was soon after overlaid by an East-West division blurring the original pattern. He also gives an overview of differences between Malagasy dialects at all linguistic levels. Some of these differences may be helpful for a subgrouping (2006:191–232). However, he does not attempt a systematic interpretation of this variety along classificatory lines, which would be a very difficult task anyway considering the diffuse pattern and lack of a clear signal in the data.

Simon's classification is considerably more solid than the others, as it combines lexicostatistics with substantial qualitative evidence. The Southwestern dialects are very different from Merina, and they hold together as a subgroup. However, although Simon's six typological criteria may prove to be critical for the subgrouping of southwestern dialects, they are obviously not sufficient as evidence for an internal classification of the Malagasy dialect group as a whole. Moreover, it remains to be investigated whether the southwestern dialects form a primary branch of Malagasy in their own right (as claimed by Simon), or whether they also include Sakalava. They admittedly seem to be more similar to one another than they are to Sakalava, but this does not rule out the possible inclusion of Sakalava in the same branch but at a higher cladistic level. Moreover, some of the typical features of southwestern dialects may in fact be retentions—forms that these dialects have kept as a result of their isolated history. For instance, the occurrence of a 2nd person singular oblique pronoun *azu* in Tandroy (corresponding to the 2nd person singular oblique form *aŋau* in Betsileo, *anau* in other dialects is clearly a retention, and therefore not solid subgrouping evidence. It is related to Maanyan *hañu* and Samihim *hayu*, which are general second person singular pronouns, from Proto-SEB *hayu (Adelaar and Kikusawa in preparation).¹² The same goes for Sakalava, Bara, and Vezo *iha* and Tandroy *ihe* '2nd person singular (nominative)'.¹³ while in this case a Sakalava form is included in the cognate set, these cognates are related to East Barito forms like

Paku *ikɔʔ*, Dusun Deyah *ikɔ* (both < *ika), Bayan *ika* ‘2nd person (unmarked)’, and must be retentions. On the other hand, Merina *hianau* (‘2nd person singular nominative’) and its corresponding forms in all central, eastern, and northern dialects are innovative, and can be considered as subgrouping evidence inasmuch as these forms have not spread through borrowing.¹⁴

3.7 LEXICOSTATISTICS AND THE LEVENSHTAIN METHOD APPLIED BY SERVA ET AL. (2011). Very recently, Maurizio Serva, Filippo Petroni, Dima Volchenkov, and Søren Wichmann (2011) used a mathematical method to investigate the early linguistic history of Madagascar. The aims of their investigation included the origins of Malagasy, the classification of its dialects, the time of arrival of the first Malagasy settlers in Madagascar, and the place where they first landed. Their method makes use of Swadesh 200-word lists for dialect varieties collected from 23 locations throughout Madagascar. It goes further than conventional lexicostatistics in that it not only calculates lexical retention rates among the dialects (the number of basic vocabulary items they hold in common), but also measures the extent of phonological change between the members of each word pair, applying a special form of the Levenshtein (or “edit”) distance. The two parameters in measuring this distance are sound similarity and word length.

The outcomes of the calculations made by Serva et al. generally vindicate the conclusions reached by other scholars of Malagasy linguistic history and dialectology. The language is most closely related to Maanyan and the SEB languages. The settlement of Madagascar is estimated to have happened some 1,350 years ago (around 650 AD). The dialects are basically divided into a southwestern plus western group and a central plus eastern group (see figure 5). The southwestern plus western group is further subdivided into Tandroy versus all other dialects (that is, Vezo, Masikoro, Tanosy, Mahafaly, Bara as spoken in Betroka, and the Salakava dialects spoken in Maintirano and Morondava). The central plus eastern group has a northern subgroup consisting of Tsimihety, the Antakarana dialects of Antalaha, Vohémar, and Ampilobe, the Sakalava dialect of Ambanja, and the Betsimisaraka dialect of Fenoarivo-Est; the other, more southern, dialects are Merina, Betsileo, Antambahoaka, Sihanaka, Taimoro, and the Betsimisaraka dialect of Mahanoro. The subgrouping of three dialects, namely Antaisaka, Zafisoro, and the Sakalava dialect of Majunga, are variable depending on the phylogenetic algorithm applied to calculate their positions. Based on the “neighbor joining algorithm,” they belong to the (non-Tandroy) subgroup of southwestern dialects, but if the “unweighted pair group method with arithmetic mean” is applied, they belong to the southern subgroup of central-eastern dialects.¹⁵ The authors argue that, according to Sapir’s (1916) principle,¹⁶ the initial land-

12. Compare also Dusun Witu *kayo*. Note that Hudson (1967) has *hañu* for both Maanyan and Samihim, but Samihim consistently has *hayu* ‘2nd person singular’ and *haye* ‘3rd person singular’ in a nineteenth century liturgical text (Soerat n.d.).

13. Compare also Betsileo *iha*, an informal 2nd person singular pronoun (Noel Gueunier, pers. comm.).

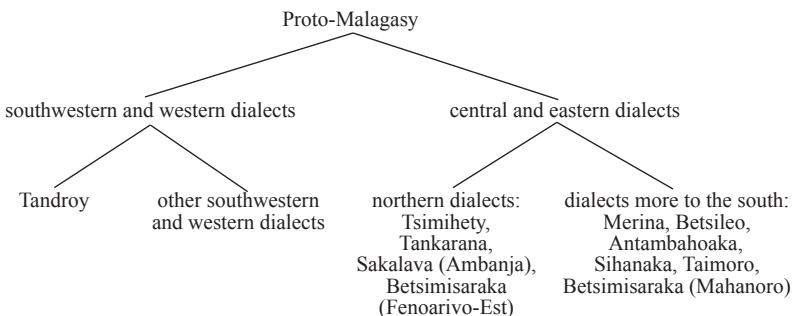
14. Merina *hianau* is originally a compound (< *iha + *nau(n)), and must have been a plural or polite counterpart of *ika ‘2nd person singular’, which survived in western and southwestern dialects and in East Barito languages (see above). The initial *h* in *hianau* is the result of metathesis; cf. Old South East Malagasy *ihanau* ‘2sg nominative’, which maintained the original compound form without metathesis (Adelaar and Kikusawa in preparation).

ing point of the first Malagasy immigrants from Asia should be in the region where these three dialects are spoken, that is, on Madagascar's southeast coast roughly between Mahanoro (midway up the island's east coast) and Ambovombe (towards the island's southern tip). Antaisaka, Zafisoro, and the Sakalava spoken in Majunga are of uncertain classification: they could belong to either the southwestern + western branch or the central + eastern branch.

Although Serva et al. use a lexicostatistical method that is more advanced than the one used by Vérin, Kottak, and Gorlin and by Mahdi, and their lexical sample is twice as big, this method ultimately suffers from the same fundamental deficiency in that it fails to distinguish (i) between inherited and borrowed vocabulary, and (ii) between retentions and innovations. Furthermore, their identification of the southeastern coast as the point of debarkation of the first SEB speakers in Madagascar is not corroborated by qualitative evidence, whether from linguistics, archaeology, or any other discipline. The claim made by various historians and anthropologists that some immigrants from insular Southeast Asia did turn up on Madagascar's southeastern coast (see Beaujard 2012) is not supported, as it concerns later migrations, taking place half a millennium after the first SEB speakers had arrived. If anything, these migrations must have a skewing effect on Serva et al.'s outcomes and on Sapir's principle, rather than showing the place where SEB speakers first arrived and from where they dispersed all over Madagascar.

3.8 AN INTERIM EVALUATION OF PREVIOUS STUDIES. The evidence adduced in the above publications remains insufficient for a full appreciation of the prehistory of Malagasy varieties and their classification. While there seems to be a fundamental division between western and southwestern dialects on the one hand, and northern, central, and eastern dialects on the other, the evidence that is most emblematic for such a division is not critical, and there is still some uncertainty about the position of the western dialects (which include

**FIGURE 5. MALAGASY DIALECT CLASSIFICATION
ACCORDING TO SERVA ET AL. (2011)**



15. The “neighbor joining algorithm” allows for differences in evolutionary rates between tree branches, whereas the “unweighted pair group method with arithmetic mean” assumes that these rates do not vary (Serva et al. 2011:2, column 2).

16. Sapir (1916) argued that, all things being equal, the homeland of a linguistic group must be sought in the area where the greatest genetic linguistic variety is found in relation to the size of this area.

southern Sakalava and western Bara). For Simon, this uncertainty is enough to include the western dialects with the northern, central, and eastern dialects in one primary branch. The position of the western dialects remains problematic: whereas the southwestern dialects qualify for a separate branch status vis-à-vis other dialects on account of several innovations they exclusively share (see 3.6), the western dialects share only some of these innovations.

In the case of the *li* : *di* and *ti* : *tsi* (*si*) oppositions, evidence from Tankarana and Tsimihety as well as the spread of the first person pronouns related to Merina *isika* show that *di* and *tsi* (*si*) forms are not always inherited, but can also have spread through borrowing. This spread may be described more adequately according to a wave model.

The quantitative evidence provided by V \acute{e} rin, Kottak, and Gorlin (1969), Mahdi (1988), and Serva et al. (2011) is inherently unreliable for reasons discussed earlier in section 3. It also loses much of its relevance with the availability of the qualitative evidence in Dez (1963) and, even more so, Simon (2006). Lexicostatistical evidence for one or two separate branches of northern dialects (shared by V \acute{e} rin, Kottak, and Gorlin and by Mahdi, but not by Serva et al.) can easily be discounted by the fact that Tankarana, Tsimihety, and northern Sakalava are spoken in a part of Madagascar that is relatively close to the Comoros and the African coast, and therefore the main gateway to the world outside. As the northern (and especially the northwestern) region is more exposed to foreign influences than other parts of the island, it should come as no surprise if this has affected the lexical retention rate of its local dialects.

Although it is not yet possible to reach a basic classification, let alone a detailed one, we have at least been able to identify some of the bottlenecks in our research, and we are also able to formulate some guidelines for further research.

4. HOW TO PROCEED FROM HERE. The following general guidelines for classifying Malagasy dialects along genetic lines may appear as self-evident to some, but they nevertheless need to be made explicit.

First, a historical classification of Malagasy dialects should be made from the vantage point of PMP and Proto-SEB. The approaches by Dez and by V \acute{e} rin, Kottak, and Gorlin fail to do this, exploiting difference per se instead of difference as a consequence of change. They thus miss the fact that there is directionality in the changes in question, and that only exclusively shared innovations are pertinent to a classification, whereas retentions from an earlier protolanguage are not.

Second, it is important to keep in mind that dialects are constantly in contact with one another (even the isolated ones), and that some dialects have had a disproportionate amount of influence on others. This is especially the case with Merina, but it is also true of some other dialects (like southern Sakalava). The point is clearly made in Kikusawa (2012). This means that certain subgrouping criteria have become less obvious through the effects of contact. A clear example of contact elements becoming default forms while marginalizing inherited elements are the many words containing *di* and *tsi* in Tsimihety, which are in the process of replacing and/or marginalizing *li* and *ti*.

Third, all dialects should be treated as equally important in the comparison. In other words, one should avoid falling into the usual trap of considering the standard language (based on the Merina dialect) as a model of normality and a point of departure. In the con-

text of Malagasy dialect classification, this is not a statement of the obvious, because Merina is not only the dominant dialect but it is also one of the most innovative of all Malagasy dialects.

Some qualitative criteria at all linguistic levels are needed for more comprehensive research into the classification of Malagasy dialects. These criteria will have to be constantly tested, added to and renewed. I suggest the following ones for inclusion.

4.1 OTHER SOUND CHANGES. These include the distribution of $*y > z$ and $*w > v$ and the merger of PMLG $*i$ and $*e$ (<PMP $*ə$) as i in final (open and closed) syllables.

The spirantization of semivowels has affected all dialects. However, it has done so most completely in Merina, and less so in some other dialects. South Betsimisaraka (Ruud 1955) and (in a few cases) Tandroy and Tanala have maintained $*y$, and Old South East Malagasy has maintained both $*y$ and $*w$ (see table 7).

The merger of final syllable $*i$ and $*e$ occurred in Merina and is widespread, but it has not taken place in Sakalava and some other dialects. This merger is diagnostic of a central group, as it seems to be limited to its members (see table 8).

4.2 THE DISTRIBUTION OF PERSONAL PREFIXES DERIVED FROM $*i$ AND FROM $*ra-$. The form $*i$ is a reflex of the PMP person marker $*si$. It is fairly general among East Barito languages. In the history of Malagasy, it functioned as a topicality marker (Adelaar and Kikusawa in preparation). On the other hand, $*ra-$ adds a notion of respect, and must be due to Javanese influence (Adelaar 2010:173–74). It has a limited distribution among Malagasy dialects, being prominent

TABLE 7. THE MAINTENANCE OF $*y$ IN MALAGASY DIALECTS

English	South Betsimisaraka	Merina
wood, tree	kāyo	hāzu
finger nail	vayan/kòo (lit., ‘tooth of finger’)	vāzanā ‘molar’
your parents-in-law	rafoyaŋ-ô (-ô ‘your’)	rafūzan-o
English	Tandroy	Merina
3SG.NOM	ie	izi
dark	ie(ŋe)	izinā
English	Old South East Malagasy	Merina
molar	wayaŋ	vazanā
nerve, tendon	uyarri	ūzatrā
alive	weluŋ	vēlunā
neck	wuyunŋ	vūzunā

TABLE 8. THE DISTRIBUTION OF FINAL SYLLABLE $*i$ AND $*e$ IN SAKALAVA

English	Sakalava	Merina	PMLG
to dig	hāli	hādi	*hāli
flight, flying	tili	tsidinā	*tūliŋ
skin	hūlitsē	hūitrā	*hūliŋ
death; corpse	fāte	fāti	*fatey
navel	fūetsē	fūitrā	*fūet
dove	limūhe	dumūhinā	*limūken

and even productive in Merina and some other central Malagasy dialects but occurring rather sporadically in some other dialects. Nevertheless, as Simon (2006:215) points out, both prefixes occur in all dialects. Sometimes *ra-* is stacked on to a lexicalized **i-*, as in *razandri* ‘my (younger) sister/brother’ (< **ra-* + **i-* + **andri* ‘younger sibling’). This stacking suggests that **ra-* is more recent than **i-*. Its history and spread may be partly explained as the result of dialect contact, especially contact of southeastern and central dialects with one another and with dialects elsewhere (see Beaujard 2003). However, this remains to be investigated.

4.3 THE FORMATION OF NUMERAL COMPOUNDS. In SEB languages in Borneo, and in most other Malayo-Polynesian languages, numeral compounds are made up of higher order units preceding lower order ones, as in English. For instance, the number 999 is expressed in both English and Maanyan as nine hundred (+) ninety (+) nine:

- (1) MAANYAN
 suei jatuh suei-pulu suei
 nine hundred nine-tens nine
 ‘999’

However, Malagasy dialects show various different patterns. In Merina, the lower order digits precede the higher ones, and are connected by *àmbi* ‘with’:¹⁷

- (2) MERINA
 sivi àmbi sivi-pùlu si sivi-n-jàtu
 nine with nine.ligature-tens and nine-ligature-hundred
 ‘999’ (Dahl 1951:199).

Sakalava follows a different pattern again: multiplications of ten precede digits, which are followed by *àmbi* ‘in addition’; compare the numeral ‘eleven’:

- (3) SAKALAVA
 fùlu ràik’ àmbi
 ten one-with
 ‘11’

However, hundreds do follow lower order numbers and are also preceded by *àmbi*, as in the numeral ‘352’:

- (4) SAKALAVA
 lima-m-pùlu rùe àmbi, àmbi telu-n-jàtu
 five-ligature-tens two with with three-ligature-hundred
 ‘352’ (Dahl 1968:14)

Finally, South Betsimisaraka shows the following pattern (for ‘356’):

- (5) SOUTH BETSIMISARAKA
 telu-n-jàtu se dima-m-pùlu ènin-àmbe
 three-ligature-hundred and five-ligature-tens six-with
 ‘356’ (Ruud 1955:48)

17. The form *-pùlu* consists of the root *fùlu* ‘ten’ preceded by a ligature *-N-*.

The Merina, Sakalava, and Betsimisaraka structures are innovative, and each of them is potential subgrouping evidence for Malagasy dialects that exhibit them (provided these dialects have not acquired them through contact).

4.4 THE EVOLUTION OF <in> IN MALAGASY DIALECTS. The PMP perfective marker **ni-/*<in>* survived as an undergoer voice affix in Malagasy dialects. In Merina, <*in*> survived as a fossilized infix in lexical roots. In Sakalava and Tandroy, however, *ni-* and <*in*> are productive and also mark past tense: *ni-* is prefixed to undergoer verbs that have no other undergoer voice markers, and <*in*> is infixed to undergoer verbs corresponding to verbs with the undergoer voice suffixes *-a* or *-e* in present tense (Dahl 1951:217; Rajaonarimanana and Fee 1996). Inasmuch as the Sakalava and Tandroy data are innovative, they are evidence for a subgroup including western and south-western dialects.

4.5 AN INVENTORY OF NONACTOR VOICE DISTINCTIONS, AND THE WAY THEY ARE EXPRESSED. Malagasy dialects still have undergoer, locative, and conveyance voice, although some of these distinctions tend to get blurred in many verbs, and the selection of the different signifiers is often subject to lexicalization. Meanwhile, a new category, the circumstantial voice, has developed, which allows a large variety of noncore arguments to become the subject. According to Rajaonarimanana and Fee's (1996) summary description, Tandroy has no locative voice, but it has an instrumental and an applicative voice instead (expressed by the prefixes *i-* and *a-*, respectively). This requires further exploration, as does the distribution of nonactor voices in other dialects, and whether their distribution constitutes a new development (and hence is innovative).

4.6 THE USE OF LEXICAL TAXONOMY. Lexical evidence is important, but as it became clear in Dez's attempt at classification, it is difficult to select diagnostic words among the endless supply of lexical items. Closed systems such as personal pronouns, kinship terms, body-part terms, directional terms, and terms pertaining to agriculture and cattle breeding are good places to start. Furthermore, not only the occurring terms are relevant but also the oppositions they imply and the systemic reductions that may have taken place in the domains they belong to. For instance, Tsimihety, Tankarana, and Tandroy have two terms for sibling-in-law. Tankarana *rajàu*, Tsimihety *rajàu*, and Tandroy *rajàutsě*, refer to 'siblings-in-law' in general, while Tankarana, Tsimihety, and Tandroy *vali/làhi* specifically refer to the brother-in-law of a man. Merina makes no such distinction but instead has a single term *ranàutrà* 'sibling-in-law'.¹⁸

5. CONCLUDING REMARKS. Dez's phonological arguments for a dialect division should be rejected because they are neither innovations nor are they exclusive to the dialects in the divisions in question. This undermines an important classification attempt, and one that has been authoritative for half a century. Nevertheless, refuting these arguments will provide a better platform for further research. It does not necessarily follow that Dez's classification was wrong per se, but it needs to be reassessed with critical

18. However, this term is now obsolete and has been replaced by *zau-bàvi* 'sister-in-law' and *zau-dàhi* 'brother-in-law' (Noel Gueunier, pers. comm.).

evidence. Simon's division into a southwestern branch and a branch including all other dialects is a possible outcome, but the position of western Malagasy dialects such as Sakalava remains uncertain: is Sakalava an early offshoot of the non-southwestern branch, or does it subgroup with the southwestern dialects after all, albeit at a higher cladistic level?

The northern dialects do not seem to differ from the central and eastern ones in a historically significant way, and I tentatively subgroup them together in a Northern-Central-Eastern branch. The fact that lexicostatistically they show lower retention rates than other dialects must reflect the fact that they have been particularly exposed to foreign influences, being close to the Comoros. They may be a primary group within that branch, but their exact relationship with central and eastern dialects needs to be sorted out further. Finally, the nature and delineation of eastern dialects remains unclear, and the label "Bet-simisaraka" needs to be defined more clearly.

The history and classification of Malagasy dialects has some important implications for the settlement history of Madagascar, about which there are several theories. Some (like Dahl 1951) believed that Madagascar was already populated by speakers of an early form of Comorian¹⁹ before the arrival of SEB speakers. These speakers would have shifted to early Malagasy, but not without leaving a strong Comorian imprint on modern Malagasy dialects. Others (Blench 2007, Adelaar 2007, 2009) believe that the Indonesian ancestors of the present-day Malagasy first migrated to the coast of East Africa, where they mixed with local Bantu speakers before they finally settled on Madagascar. First proposed by the historian Hubert Deschamps (1960), this idea has gained much currency since the publication of test results of Hurlés et al. (2005) and subsequent teams of human geneticists, showing that the Malagasy population as a whole shares a mixed Asian and Bantu gene pool along both the Y-chromosome and mitochondrial DNA lines. Philippe Beaujard (2011) believes that the first migrants to Madagascar were SEB speakers coming directly from Indonesia, and that they came into contact with Bantu speakers in the Comoros and north Madagascar at the end of the first millennium AD. According to Pierre Simon (2006), the first SEB speakers in East Africa migrated to the Comoros before they ended up in Madagascar. Meanwhile, there is growing multidisciplinary support for the idea that Madagascar already had a human population before the arrival of SEB speakers and Bantu speakers, although the identity of this early population remains unknown; see Blench (2007) for a summary of the evidence.

The linguistic and human genetic evidence does not favor Beaujard's scenario of SEB and Bantu migrants moving to Madagascar in separate groups. It is difficult to see how it could have provided the conditions leading to the current genetic and linguistic homogeneity. The latter suggests that a mixed Asian/Bantu population had been in Madagascar from the initial settlement, and it makes more sense to locate the hybridization process of their Asian and Bantu ancestors in East Africa, in a period prior to the settlement of Madagascar.

Eastern and central parts of Madagascar remained in contact with Southeast Asia, leaving an imprint of influences (particularly from Malay and Javanese) that are generally missing in western, southwestern, and northern dialects (cf. Beaujard 2012). The imprint may be strong, but as far as the eastern and central dialects are concerned, it is

19. This term refers to the languages spoken on the Comoros Islands, which are closely related to Kiswahili and with it belong to the Sabaki or coastal East Bantu subgroup of Bantu languages (Nurse and Hinnebusch 1983).

mainly lexical and had little impact (if any impact at all) on the formation of dialect groups. When it comes to phonology and morphosyntax, these dialects are neither more Malayo-Polynesian, nor less Bantu, than the northern, western, and southwestern ones. Significantly, all dialects have adopted the same key elements of Bantu morphosyntax (including nominal prefixes and tense).

Although there seems to be a basic division running through Madagascar between southwestern (and western?) dialects on the one hand, and central, eastern, and northern dialects on the other, the differences in question are relatively minor, suggesting that the island has been overwhelmingly monolingual and SEB for a long time. The seeds of the relatively homogeneous speech community of the Malagasy today may have already been planted before they settled in Madagascar.

APPENDIX.

Language (or language group)	Abbreviation	Source
Proto-Malagasy	PMLG	as indicated in text
Proto-Malayo-Polynesian	PMP	Blust and Trussel (in progress)
Proto-South East Barito	PSEB	Adelaar (work in progress)
Arabic		Wehr (1994)
Bara		Elli (1988)
Betsimisaraka		as indicated in text
Javanese		Pigeaud (1938)
Maanyan		Rubay (1997)
Malay		Wilkinson (1959)
Merina		Abinal and Malzac (1888)
Old South East Malagasy		Ferrand (1904)
Sakalava		Gueunier (n.d.)
South East Barito	SEB	Hudson (1967)
Tanala		Beaujard (1998a)
Tandroy		Rajaonarimanana and Fee (1996)
Tankarana, Tanosy		Serva and Petroni (2011)
Tsimihety		Faridanona (1977)
Vezo		Poirot (2011)

REFERENCES

- Abinal, Antoine, and Victorin Malzac. 1888. *Dictionnaire malgache-français*. Paris: Éditions maritimes et d'Outre-mer (reprinted in 1970).
- Adelaar, Alexander. 2007. Language contact in the Austronesian Far West. Paper presented at the Third Conference on Austronesian Languages and Linguistics, SOAS, London, September 21–22.
- . 2009. Loanwords in Malagasy. In *Loanwords in the world's languages: A comparative handbook of lexical borrowing*, ed. by Martin Haspelmath and Uri Tadmor, 717–46. Berlin: Mouton de Gruyter.
- . 2010. The amalgamation of Malagasy. In *A Journey through Austronesian and Papuan linguistic and cultural space: Papers in honour of Andrew K. Pawley*, ed.

- by John Bowden, Nikolaus P. Himmelmann, and Malcolm Ross, 161–78. Canberra: Pacific Linguistics.
- Adelaar, Alexander, and Ritsuko Kikusawa. In preparation. Proto-Malagasy pronominal systems: A reconstruction.
- Beaujard, Philippe. 1998a. *Dictionnaire malgache–français: Dialecte Tañala, sud-est de Madagascar avec recherches étymologiques*. Paris: L'Harmattan.
- . 1998b. *Le parler secret arabico-malgache du sud-est de Madagascar: Recherches étymologiques*. Paris: L'Harmattan.
- . 2003. Les arrivées austronésiennes à Madagascar: Vagues ou continuum? *Études Océan Indien* (INALCO, Paris) 35/36:59–147.
- . 2011. The first migrants to Madagascar and their introduction of plants: Linguistic and ethnological evidence. *Azania* 46(2):169–89.
- . 2012. *Les mondes de l'Océan Indien*, tome 1: *De la formation de l'état au premier système-monde afro-eurasien*. Paris: Armand Colin.
- Blench, Roger. 2007. New palaeozoogeographical evidence for the settlement of Madagascar. *Azania* 42:69–82.
- Blust, Robert A., and Stephen Trussel. In progress. *Austronesian comparative dictionary*. <http://www.trussel2.com/ACD/>.
- Dahl, Otto Christian. 1951. *Malgache et maanyan: Une comparaison linguistique*. Avhandlinger utgitt av Instituttet 3, Oslo: Egede Instituttet.
- . 1968. *Contes malgaches en dialecte sakalava*. Instituttet for sammenlignende kulturforskning. Oslo: Universitetsforlaget.
- . 1983. *Sorabe révélant l'évolution du dialecte antemoro*. Antananarivo: Trano Printy Loterana.
- Deschamps, Hubert. 1960. *Histoire de Madagascar*. Paris: Berger-Levrault.
- Dez, Jacques. 1963a. Aperçu pour une dialectologie de langue malgache. *Bulletin de Madagascar* 204.
- . 1963b. Aperçu pour une dialectologie de langue malgache. *Bulletin de Madagascar* 205.
- . 1963c. Aperçu pour une dialectologie de langue malgache. *Bulletin de Madagascar* 206.
- . 1963d. Aperçu pour une dialectologie de langue malgache. *Bulletin de Madagascar* 210.
- Elli, Luigi. 1988. *Dizionario Bara–Italiano*. Fianarantsona: Ambozontany.
- Faridanona. 1977. *Rantimbôlaña diksionera Tsimihety*. Antananarivo: Akademia Malagasy.
- Ferrand, Gabriel. 1904. *Un texte arabico-malgache du XVII^e siècle. Transcrit, traduit et annoté d'après les mss 7 et 8 de la Bibliothèque Nationale*. Paris: Imprimerie Nationale.
- Flacourt, Étienne. 1658. *Dictionnaire de la langue de Madagascar*. Paris: Georges Josse.
- Gueunier, Noel J. 1988. Dialectologie et lexicostatistique: Cas du dialecte malgache de Mayotte (Comores). *Études Océan Indien* (INALCO, Paris) 9:143–69.
- , ed. n.d. *Dictionnaire des dialectes malgaches du sud ouest (fondé sur le Dictionnaire sakalava–merina–français de Victor Denis Mahavere [1925–1926], revu et complété par Noel J. Gueunier)*. Unpublished MS.
- Gueunier, Noel J., and Solo Raharinjanahary. n.d. *Atlas linguistique de Madagascar: Situations et perspectives*.
- . To appear. A dialect-geography of the southern part of Madagascar.
- Houtman van Gouda, Frederick de. 1603. *Spraeck ende woord-boeck inde Maleysche ende Madagaskarsche talen [...]*. Amsterdam: J. E. Cloppenburg.
- Hudson, Alfred B. 1967. *The Barito isolects of Borneo*. Southeast Asia Program (Department of Asian Studies), Data Paper no. 68. Ithaca, NY: Cornell University Press.

- Hurles, Matthew E., Bryan C. Sykes, Mark A. Jobling, and Peter Forster. 2005. The dual origin of the Malagasy in island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. *American Journal of Human Genetics* 76:894–901.
- Kikusawa, Ritsuko. 2012. Standardization as language loss: Potentially endangered Malagasy languages and their linguistic features. *People and Culture in Oceania* 28:23–44.
- Mahdi, Waruno. 1988. *Morphophonologische Besonderheiten und historische Phonologie des Malagasy*. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg, Band 20. Berlin: Dietrich Reimer.
- Nurse, Derek, and Thomas J. Hinnebusch, eds. 1993. *Swahili and Sabaki. A linguistic history*. University of California Publications in Linguistics Volume 121. Berkeley and Los Angeles: University of California Press.
- Pigeaud, Th. 1938. *Javaans–Nederlands woordenboek*. Groningen-Batavia: J. B. Wolters.
- Poirot, Gérard. 2011. *Dictionnaire vezo–français suivi d'un index français–vezo*. Privately published. (ShopMyBook; www.unibook.com).
- Rajaonarimanana, Narivelo, and Sarah Fee. 1996. *Dictionnaire malgache dialectal–français: Dialecte tandroy*. Dictionnaires des Langues O', Langues de l'Océan Indien Occidental. Paris: Langues et Mondes, L'Asiatheque.
- Ruud, Joergen. 1955. Étude grammaticale du dialecte betsimisaraka du sud. *Bulletin de l'Académie Malgache* 33:33–55.
- Sapir, Edward. 1916. *Time perspective in Aboriginal American culture: A study in method*. Ottawa: Government Printing Bureau.
- Serva, Maurizio, Filippo Petroni, Dima Volchenkov, and Søren Wichmann. 2011. Malagasy dialects and the peopling of Madagascar. *Journal of the Royal Society Interface* 1:1–14 (doi:10.1098/rsif.2011.0228).
- Serva, Maurizio, and Filippo Petroni. 2011. Dialects of Malagasy (a comparative word-list of Malagasy dialects). (<http://univaq.it/~serva/languages/zlist.pdf>).
- Simon, Pierre. 1988. *Ny fiteny fahizany. Reconstitution et périodisation du malgache ancien jusqu'au XIV^e siècle*. Travaux et documents 5 du CEROI. Paris: Institut des Langues et Civilisations Orientales.
- . 2006. *La langue des ancêtres (Ny fitenin'dRazana): Une périodisation du malgache de l'origine au XV^e siècle*. Paris: L'Harmattan.
- Soerat pangadjaran agama olon kristen huang bahasa Dajak Samihim*. n.d.. Banjarmasin (Indonesia): Bandjermasinsche drukkerij.
- Velonandro (ed.). 1983. *Lexique des dialects du nord de Madagascar, par des missionnaires et séminaristes catholiques*. Tuléar (Madagascar): Centre de Documentation et de Recherche sur l'Art et les Traditions Orales à Madagascar, Centre Universitaire Régional/Valbonne (France): Centre de Documentation et de Recherche sur l'Asie du Sud-Est et le Monde Insulindien, CNRS-EHESS.
- Vérin, Pierre, Conrad P. Kottak, and Peter Gorlin. 1969. The glottochronology of Malagasy speech communities. *Oceanic Linguistics* 7:26–83.
- Wehr, Hans. 1994. *A dictionary of modern written Arabic (Arabic–English)*. 4th ed., ed. by J. M. Cowan. Ithaca, NY: Spoken Language Services, Inc.
- Wilkinson, R. J. 1959. *A Malay–English dictionary*. London: Macmillan.