KONINKLIJK MUSEUM
VOOR MIDDEN-AFRIKA
TERVUREN, BELGIË

MUSÉE ROYAL DE
L'AFRIQUE CENTRALE
TERVUREN, BELGIQUE

ANNALEN
MENSWETENSCHAPPEN

VOL. 162

ANNALES
SCIENCES HUMAINES

CONTINUITY AND DIVERGENCE in the Bantu languages: perspectives from a lexicostatistic study

Yvonne BASTIN, André COUPEZ, Michael MANN

1999

Contents

Avant-Propos — Foreword

L'enquête de lexicostatistique qui est ici présentée a été entreprise en 1953 par A. Meeussen, A. Coupez et J. Vansina. Ses principales étapes sont les suivantes, compte tenu du fait que la récolte des données ne s'est pas interrompue et se poursuit encore:

- 1956 premiers calculs sur des couples de langues: A. Meeussen; A. Coupez;
- 1963 première élaboration linguistique extensive: A. Coupez; transmission à É. Evrard pour calcul sur cartes perforées;
- 1966 première publication statistique (à partir de l'élaboration de 1963): É. Évrard;
- 1975 première publication linguistique extensive (à partir de l'élaboration de 1963): A. Coupez, É. Évrard et J. Vansina;
- 1983 secondes élaboration et publication linguistiques extensives: Y. Bastin, A. Coupez et B. de Halleux (calcul sur ordinateur effectué par B. de Halleux);
- 1990 troisième élaboration linguistique extensive: Y. Bastin; transmission à M. Mann pour calcul sur ordinateur.

La méthode lexicostatistique, présentée à partir de 1951 par M. Swadesh, visait au départ une classification des langues doublée d'une chronologie absolue. Très rapidement (voir Hymes 1960), l'objectif de chronologie absolue a été abandonné face aux critiques et l'on s'est limité à envisager la chronologie relative. Même réduite à cet aspect, la méthode offre la seule perspective de classification globale du bantou qui soit accessible dans un avenir prévisible.

Notre attitude est pragmatique. Pendant près d'une vingtaine d'années, l'application de la lexicostatistique au bantou a été soumise à des essais, des contrôles et des adaptations.

Les premiers calculs portant sur des couples de langues (1956) ont donné des résultats vraisemblables. La première classification linguistique extensive (1975; 58 langues), dont les calculs ont été effectués par E. Évrard au L.A.S.LA. (Laboratoire d'analyse statistique des langues anciennes, Université de Liège) au moyen de cartes perforées, a permis de multiples contrôles sur la base des classifications partielles établies antérieurement à partir d'études fiables. Ainsi la zone interlacustre, désignée par le sigle J, que Meeussen avait posée en 1953 en regroupant l'est de la zone D et l'ouest de la zone E de M. Guthrie, apparaît telle quelle dans l'arbre lexicostatistique.

La seconde publication linguistique extensive (1983; 214 relevés), à laquelle a collaboré Y. Bastin et dont le calcul statistique a été effectué sur ordinateur par B. de Halleux au C.I.D.A.T. (Centre d'informatique appliquée au développement et à l'agriculture tropicale, Musée royal de l'Afrique Centrale, Tervuren), a apporté de nouvelles confirmations. Ainsi, comme l'avaient établi respectivement au colloque de Viviers sur l'expansion bantoue (1977) A. Meeussen et J. Voorhoeve, le mbagani, situé géographiquement dans la zone L et classé L 22 par Guthrie, quitte cette zone pour se grouper avec des langues situées plus à l'ouest; plusieurs langues des groupes A 40 et A 60 se détachent de la zone A pour former un groupe distinct. Par ailleurs, une statistique grammaticale effectuée par Y. Bastin (1979) confirme dans l'ensemble les statistiques lexicales.

Ce n'est qu'à cette époque qu'il a été décidé d'étendre l'enquête à l'ensemble du bantou.

Le même pragmatisme a guidé l'application de la méthode, dont le principal obstacle tient à l'imprécision sémantique. Si des mots de la liste se prêtent à la polysémie, l'enquête risque d'être faussée par des choix qui diffèrent selon les relevés. Sur la base de l'expérience acquise dans les premières années d'application, nous avons procédé à trois types de modifications de la liste, tout en visant à sauvegarder l'esprit de la méthode. Huit mots ont été éliminés, soit parce qu'ils ne correspondent pas à des concepts courants en bantou, soit parce qu'ils se prêtent mal à la comparaison linguistique; ce sont: "celui-ci, celui-là, glace, jaune, ne...pas, tous, vert, neige". Quelques mots ont été remplacés par d'autres, de sens proche, mais plus courants en bantou; par exemple "bras" a été substitué à "main", "jambe" à "pied"; "pluie" à "pleuvoir". Enfin, quelques précisions sémantiques ont été introduites, par exemple "rond" à prendre comme "rond comme une boule"; "semence" à prendre comme "graine mise en terre"; "terre (sol)" à prendre comme "par terre".

Pour éviter les erreurs, les relevés sont effectués de préférence en interrogeant des spécialistes d langues intéressées ou des locuteurs sûrs. Dans ces cas, qui constituent la majorité, la liste de 92 mots e généralement complète, avec quelques exceptions éventuelles (souvent "rond", par exemple). À défa on utilise des dictionnaires de bonne qualité ou des listes lexicales publiées par des linguistes po d'autres enquêtes statistiques; un peu plus nombreuses alors, les lacunes ne dépassent néanmoins pas d mots en général. Ce n'est que dans des cas exceptionnels et de manière provisoire qu'il est fait usage d'i formations moins fiables ou plus lacunaires.

Le nombre des listes disponibles est supérieur à celui des listes utilisées et mentionnées ci-après. Lo qu'il y a double emploi entre deux listes, nous avons choisi celle qui possède la meilleure transcriptio D'autre part, nous n'avons retenu qu'un parler dans les groupes qui, lors des enquêtes précédentes, s apparus comme composés de dialectes. Par convention, nous considérons comme ayant une relation di lectale les parlers dont le rapport lexicostatistique est égal ou supérieur à 86%.

L'établissement des étymologies sur lesquelles se fondent les calculs ne pose pas de problèm majeurs. Au début de l'enquête, les langues bantoues du nord-ouest étaient peu documentées; au avaient-elles été omises dans l'élaboration de 1963 (publication de 1975). Depuis lors, les connaissanc ont progressé dans cette région; nous y avons bénéficié en particulier des travaux de B. Janssens, qui no a prêté son concours. Une difficulté apparaît d'autre part en liaison avec la variabilité lexicale, mécanis linguistique présenté à partir de 1975 par A. Coupez, selon lequel une certaine proportion du lexique ba tou, empreinte d'iconisme, présente un jeu de variantes libres distinguées par de petits écarts formels. pourrait donc envisager, pour la dizaine de mots de la liste lexicostatistique qui sont impliqués, d'étab des étymologies qui n'offrent pas tous les parallélismes formels requis, mais on s'en est abstenu par pr dence. Peut-être à l'avenir un double calcul serait-il souhaitable dans ces cas.

Par un malheureux concours de circonstances, les données que D. Schoenbrun nous avait commu quées à partir de son enquête sur le nord-est du bantou n'ont pas été incorporées dans nos calculs. Il est néanmoins tenu compte dans le dernier chapitre.

Une enquête parallèle à la nôtre a été menée en 1996 par P. Piron sur les langues bantoïdes en re tion avec le bantou. Elle permet de préciser la limite nord-occidentale du groupe bantou et probableme d'incorporer au bantou une partie du groupe bantoïde.

Les premiers calculs de la présente enquête ont été communiqués en 1993 à J. Vansina, qui en a don sa propre interprétation historique.

Il reste à envisager l'avenir, c'est-à-dire essentiellement à inclure les langues encore absentes. Si circonstances s'y prêtent, nous utiliserons les nombreuses données qui ont été publiées récemment ai que de nouvelles listes que des collaborateurs bénévoles nous ont adressées, c'est-à-dire à ce jour: C Tetela, parlers divers (Tassa Okombe); zone J tembo, région de Kalemye (P. Petit); D 32 bira, parl divers (C. Kutsch Lojenga); zone E? Rumi (B. Van Houdt); E 70 malakote-ilwana (A. J. Rossbac E 73 digo (M. de Groot); M 30 ndali (R. Botne); M 41 bwile (P. Petit).

D'autre part et dans la mesure du possible, des relevés de meilleure qualité seront substitués à ce qui n'ont pas été effectués dans les conditions idéales.

Enfin sera introduite une amélioration tardive, qui est préparée depuis plusieurs années par une ad tion dans les relevés: la substitution d'"aller" à "marcher", concept souvent élusif en bantou.

Nous faisons appel, pour cette phase de l'enquête que nous espérons terminale, à toutes les contri tions et observations que pourraient nous apporter les bantouistes.

ANDRÉ COU

1. Introduction

The research reported in this study is the fruit on the one side of collaboration among a large number of scholars and their assistants, animated principally by Jan Vansina, André Coupez and the late A.E. Meeussen, in accumulating standard vocabularies of Bantu languages over more than 30 years. This work is described in more detail in the Foreword by André Coupez, and major collaborators have been named in the Acknowledgements. All contributions are acknowledged in the listing of sources in section 2.2, and their contribution is warmly acknowledged. Analyses based on earlier subsets of this data, including Coupez (1956), Meeussen (1956), Evrard (1966), Coupez–Evrard–Vansina (1975) and Bastin–Coupez–de Halleux (1983), are discussed in the Foreword.

The collection of the 542 vocabularies presented in the present study was complete by 1990; Michael Mann was invited to join André Coupez and Yvonne Bastin in the analysis and presentation of the data, which at his request was extended to consider the data both in its historical and geographical dimensions. He regrets that the price of this ambition has been a delay that has dismayed his Belgian colleagues, to whom on that score he owes every expression of apology. André Coupez and Yvonne Bastin undertook the initial correlation and cognation judgement of the data (as described in Chapter 2), and Yvonne Bastin has further contributed the documentation (section 2.2) and the localisation that forms the basis for Chapter 3. Other aspects of the analysis have been undertaken by Michael Mann, who bears full responsibility for the text. Time brooking no further delay, this study has the character of observation and report, largely in graphic form; it presumes to offer no elaborated historical interpretation, except to insist to historians on the many shades of grey.

It was the hope of scholars associated with this project that a more exhaustive collection of Bantu vocabularies would resolve the ambiguities of earlier analyses. It has rather in my judgement confirmed those ambiguities and forced us to question the basis of the (phylogenetic) clarity that has been looked for in comparative linguistic studies. Rather, the ambiguities are part of the evolving linguistic situation which historians and linguists together may eventually illuminate. The contribution of Jan Vansina as a colleague who represents both disciplines has been especially valued, and his anticipatory synthesis of this data (Vansina 1995) presents an interpretation from which I would diverge only in detail.

There have been many hints that relation among the Bantu languages (and indeed within language families in general) has as much to do with geography as history. When Henrici (1973) set out to review (and redo) Guthrie's use of statistics in *Comparative Bantu* (1967-71), in one analysis he applied Multi-Dimensional Scaling (MDSCAL) to a table of linguistic (dis)-similarities among Guthrie's 28 'test languages'; with input restricted to linguistic data, the analysis arranged the languages faithfully in (2-dimensional) space in a way that diverged from geographical reality chiefly in the stretching of relative distances as one approached the North-West of the Bantu area (Henrici 1973: 90). I have repeated this experiment on several occasions with other linguistic data with similar results, suggesting that if geography can be reconstructed from comparative linguistic data, geography must have played a large part in shaping those relationships.

Secondly, varied attempts to classify the distribution of reflexes of Guthrie's lexical reconstructions have pointed more to geography than history. Guthrie's own study of 'topology' (a considerable innovation in comparative linguistics) attributed distributions to over-generously defined Eastern and Western regions, and in many contexts did not differentiate between exclusive and predominant distribution within one or other regions. If the categories are sharpened, the clarity disappears. Considering the distribution of Guthrie's non-General roots in his 'test' languages, roots are attested on average in less than six languages, while attributing each root to a node on a tree representing a 'latest common ancestor' implies a potential distribution over more than 20 languages on average for all trees considered – an implausible rate of loss. On the other hand calculating a 'centre of gravity' for the distribution of each root and plotting these on a map reveals a fairly even geographical spread.

There is in fact fairly clear statistical evidence that many of Guthrie's roots, including perhaps some of his General roots, are not inherited from Proto-Bantu, but subsequent innovations. If we assume a com-

mon origin in Proto-Bantu, we can calculate a rate of loss for each language, and predict that the common inheritance for two languages will be the product of the two rates of loss. Any higher figure for the actual tally of shared forms will be good news and suggest a period of common history for the languages concerned. Any lower figure challenges the assumption of originality. For some 12% of comparisons between languages this test fails (the figure is 4.5% for General roots). We cannot ascribe Guthrie's whole corpus to the common lexical stock, and must assume that many roots spring from local innovations with an areal rather than a historical focus.[1]

Séguy (1971) has given some insights into lexical differentiation from a study of French dialectology. Drawing lines across the map and sampling language at approximately equal intervals along the line, he counted similarities between pairs of samples and plotted linguistic difference against geographical distance. He found near neighbours to be sharply differentiated, but instead of increasing uniformly with distance, after about 50 km linguistic difference reached a plateau and remained largely independent of distance. In explanation he suggests two conflicting principles in operation. On the one hand, a sense of identity prompts each language to differentiate itself from its neighbours (even if it does so by making a different selection from a wider lexical pool); beyond the range of immediate inter-communal competition, variety is constrained by the needs of intercomprehension.

The influence of geography on language history has long been recognised, but generally an attempt has been made to abstract it out. Following the work of Maurice Swadesh, the Belgian originators of the present project confined their wordlists to 'basic' vocabulary, which is supposedly more resistant to change than cultural vocabulary. However cross-cultural comparative studies such as Kruskal–Dyen–Black (1971) have demonstrated that what vocabulary is most resistant to change differs considerably from one culture to another even within basic vocabulary. Guthrie relied on regular sound correspondences to exclude later innovations, though he conceded a few items – fewer than we now suppose – 'probably originated within the Bantu period'. In contrast I take the view that language has developed within overlapping communities, and that we should take note of all ambiguities of relationship for which ultimately some historical account must be given.

Vansina (1995) in previewing the present study expounded an understanding of the wave theory of linguistic diffusion, suggesting a mechanism whereby internal dialect differentiation within a static community can result in a 'core and periphery' situation, with languages more differentiated the further their distance from the core. A tree representation of the same situation might suggest earlier splits of the peripheral languages from the core, which might be interpreted historically as implying successive location of the ancestral communities near the boundaries of the offshoots concerned. The principle of 'least moves' favours the simpler interpretation of contemporaneous differentiation of languages at different removes from some focal point.

I would refine the position taken by Vansina by suggesting that the model of differentiation within a static core is not restricted to dialect relations. Dialect is usually defined with reference to inter-comprehension, but the ability to comprehend is relative to experience, exposure and motivation, and further neighbouring peoples can and do learn each other's languages, with greater facility and probability the greater the linguistic similarity. Thus an area of high linguistic differentiation may be no less likely to show something of the phenomenon of core and periphery. Indeed the Bantu languages as a whole may appear as a continuum with core and periphery, with many repetitions in microcosm at greater or less distance. (It will be suggested later that some trees are to be read with the vertical dimension representing closeness of communication rather than separation in time.)

If cores are seen as centres of influence and communication, individual language communities may be in relation to more than one. Centres may wax and wane, and allegiances alter. Moreover the substance of differentiation is innovation, and innovation is continually forging new communities. To the confusion of the tree model, shared innovation is independent of ancestry. The most we can say of two languages that display a number of innovations in common is that they have shared a period of linguistic community.

I had hoped before publication of this study to respond to the challenge of Colin Flight (1988) and familiarise myself with developments in taxonomy to be found within the pages of the journal *Systemati*

[1] Informally, Guthrie used to describe the 'source items' underlying his lexical reconstructions as 'located somewhere in prehistory' a very insightful expression; formally, however, he generally assumed this to be 'before the Bantu threshold' when supposedly Proto-Bantu started differentiating into daughter languages.

Zoology. Flight takes previous writers on Bantu classification severely to task for failing to differentiate between phenetic (observational) and cladistic (evolutionary) classification. I believe it would be useful to survey some of the methods developed, particularly perhaps methods that operate on the incidence matrix (effectively the data as set out in section 2.3) rather than the correlation matrix (the count of shared vocabulary between languages taken two at a time). However, my regret at not having been able to pursue this is tempered by a sense of a radical difference between biological evolution and linguistic history: biological features are inherited, linguistic features are acquired.

The theme of the continuity of geography and history run through this study. Chapter 2 presents the 'raw' data in a graphic form that it is hoped will enable readers to form their own impression of the diversity of the data and the prolific local innovation that generates it.

Given the multiple indications of local influences in linguistic relationship, Chapter 3 looks at interrelationship on the ground, noting areas of continuity, where languages share a measure of community with their neighbours irrespective of hierarchical ('historical') relationship, but more particularly at lines of discontinuity, possibly at the border between competing cores, possibly representing physical barriers, and in some cases representing a real historical discontinuity (as with the Nguni-speakers of Zimbabwe and Malawi).

Chapter 4 approaches the construction of trees conscious of the strong effect of geography on local relationships; it constructs a series of trees progressively emphasising internal 'connectivity', in the hope that this will discount more immediate relationships and reveal connections at greater distance in time or space. At the same time since trees inherently simplify reality by imposing discrete groups, it provides measures for evaluating the well-formedness of each putative group.

2. Presentation and correlation of the lexical data

2.1. Introduction

This chapter presents in as visual form as can be contrived the data on which the analyses in the following chapters are based, consisting of 542 standard vocabularies. It details the lexical questionnaire, lists the contributors and sources of the vocabularies, maps their distribution, and finally presents the data gloss by gloss on symbolic maps.

The linguistic questionnaire, published in English and French versions as a supplement to the *Bantu Newsletter* 2 (February 1975), listed 92 glosses, reduced from the standard 100-word list of Maurice Swadesh by the omission of glosses that were climatically inappropriate or that belonged more to grammar than lexis. Precisions were added to avoid ambiguity in the meanings of some glosses, e.g. 'skin' was to refer to human skin (rather than treated animal hides or vegetable rinds). Generic terms were to be preferred to specific, the grammatical category of the gloss was not binding (for example, an adjective might be rendered by a verb or noun if that was more natural in the language being collected), and multiple responses were allowed where several terms were in equal currency. Respondents were asked to leave blanks in cases of uncertain comprehension or interpretation. The vocabulary is set out in Table 2.1.1 with added precisions.

all/tous
arm/bras
ashes/cendres
bark *n*/écorce
belly/ventre
big/grand (≠ long)
bird/oiseau
bite/mordre
black/noir
blood/sang
bone/os
breast/sein
burn *intr*/brûler (être en feu)
cloud/nuage
cold (weather)/froid (temps)
come/venir
die/mourir
dog/chien
drink *v*/boire
dry/sec
ear/oreille
eat/manger
egg/oeuf
eye/oeil
fat *n*/graisse (animale)
feather/plume
fire/feu
fish *n*/poisson
fly *v*/voler (oiseau)
full/plein
give/donner
good/bon
ground (on the ~)/terre (par terre)
hair (of head)/cheveu
head/tête
hear/entendre
heart/coeur
horn/corne
kill/tuer
knee/genou
know/savoir
leaf/feuille (d'arbre)
leg/jambe
lie (down)/couché
liver/foie
long/long (≠ grand)
louse/pou
man/homme (être humain masculin)
many/beaucoup (nombreux)
meat/viande
moon/lune
mountain/montagne
mouth/bouche
nail/ongle
name/nom
neck/cou
new/nouveau
night/nuit
nose/nez
one/un (nombre)
path/chemin
person/personne (être humain)
rain *n*/pluie
red/rouge
root/racine
round (as ball)/rond (comme une boule)
sand/sable
say/dire
see/voir
seed *n*/semence (graine mise en terre)
sit/assis
skin (human)/peau (humaine)
sleep *n*/sommeil
small/petit
smoke *n*/fumée
stand/debout
star/étoile
stone/pierre
sun/soleil
swim/nager
tail/queue
tongue/langue (corps)
tooth/dent
tree/arbre
two/deux
walk/marcher
warm (weather)/chaud (temps)
water/eau
what?/quoi?
white/blanc
who?/qui?
woman/femme (être humain féminin)

Table 2.1.1: word-list

The sources of the vocabularies are set out in section 2.2. Most have been collected from named informants by numerous collaborators, a few have been excerpted from published sources, not always contemporary. Not infrequently, several vocabularies have been collected of the 'same' language; though some show a close concurrence, most commonly the similarity is between 80% and 90%, and in a few cases that may be suspected of mislabelling, is as low as 60%. We have treated each vocabulary as a sep-

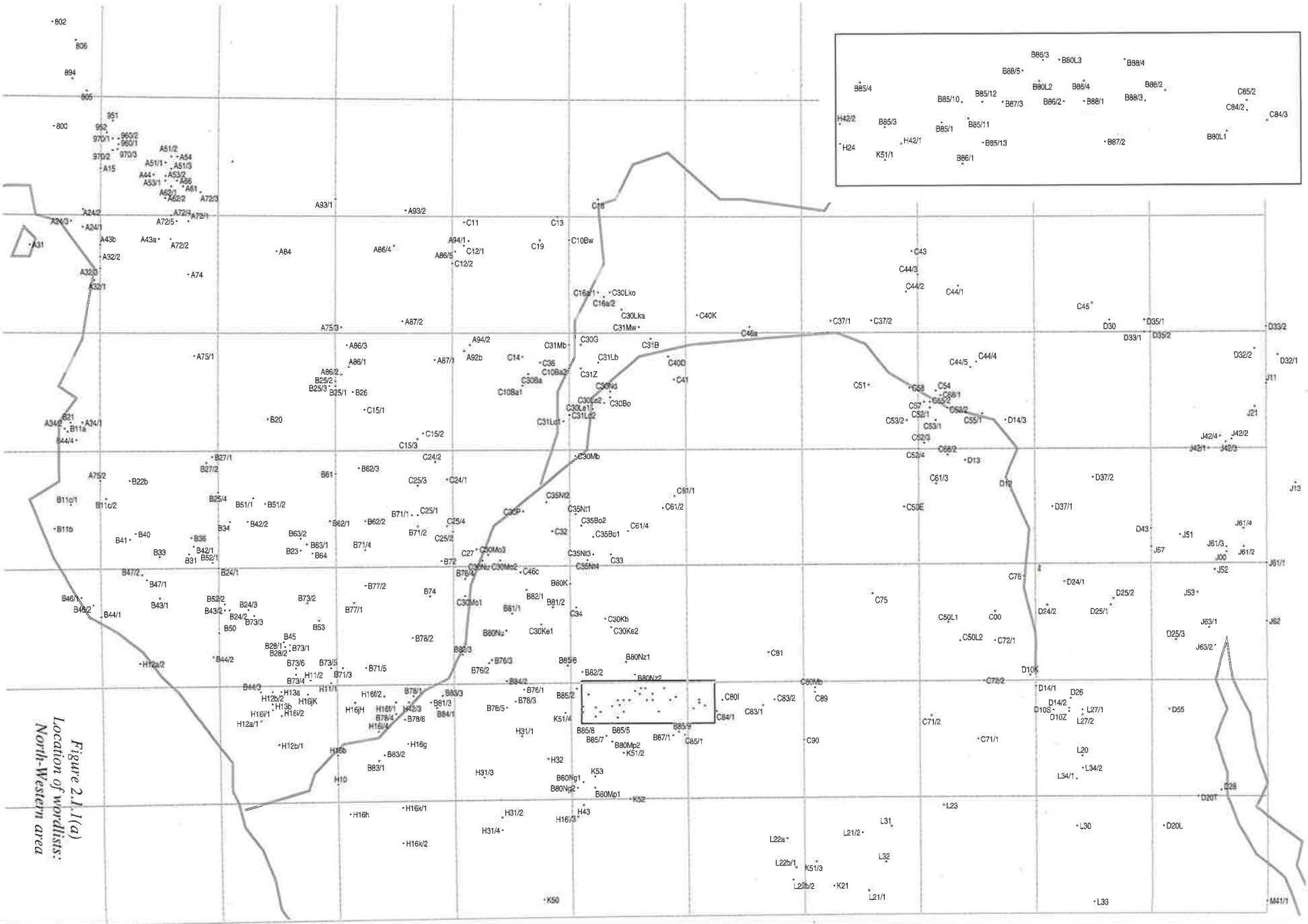

Figure 2.1.1(a)
Location of wordlists: North-Western area

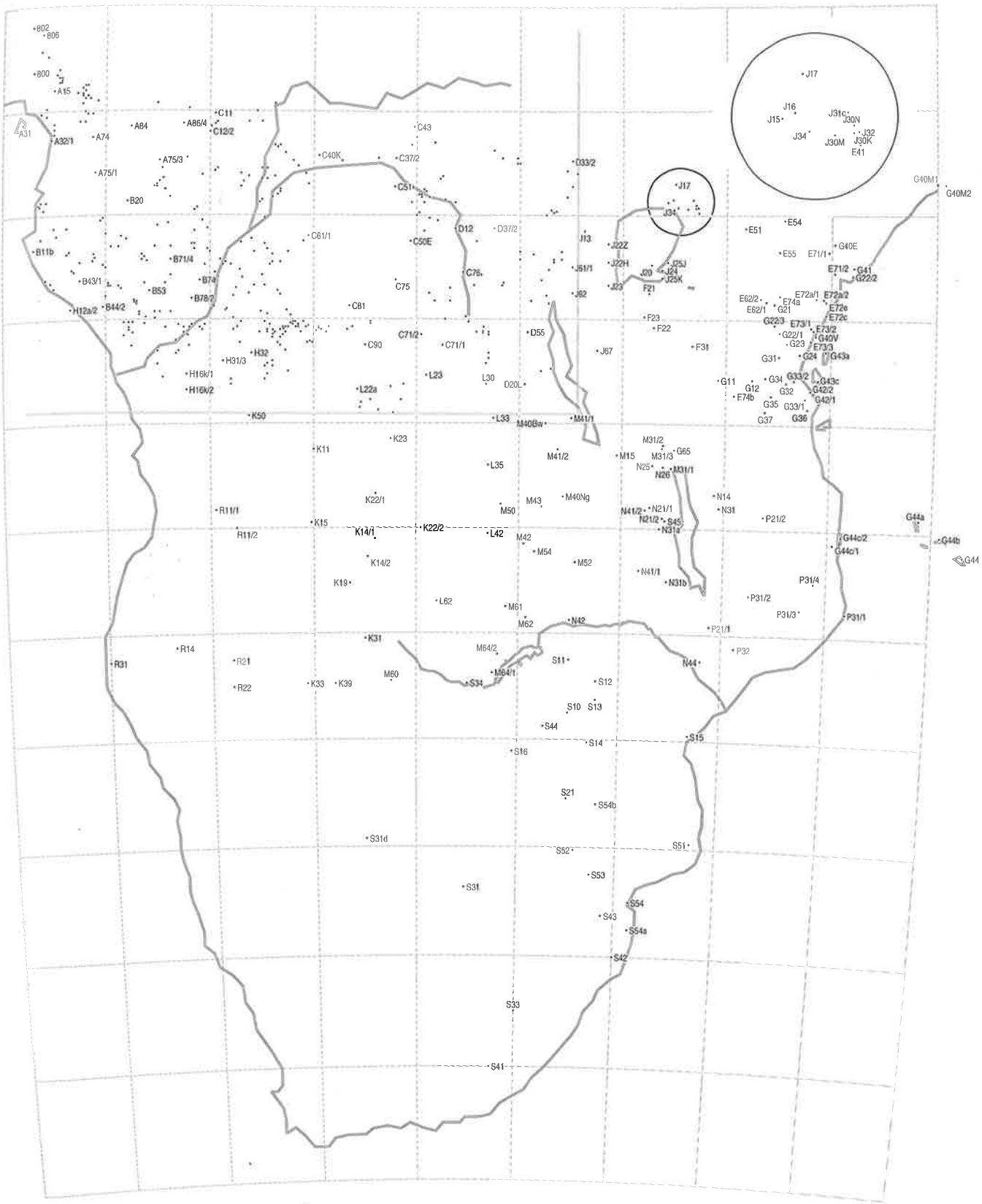

Figure 2.1.1(b)
Location of wordlists: Eastern and Southern areas

arate entity rather than collective testimony to a more abstract language, and we have associated each with a locality (whether the place where the vocabulary was recorded or a claimed home area), so that the vocabularies are to be thought of as language recorded at sample-points rather than representative of language 'areas'. (In a few cases we have slightly displaced these locations for visibility in mapping.)

It will be seen from the key maps (Figures 2.1.1(a)/(b)) that the density of the sampling is much higher in the North-West than in the rest of the area, to the extent that this area is shown in a separate enlargement. While this may reflect in part the nature of the network that has yielded the greater part of the data, it is also a just reflection of the greater linguistic variation in this area. Elsewhere sampling in some areas is relatively thin, and some language groups are missing entirely: F1 Tongwe, G5 Pogoro, M2 Nyika-Safwa, P1 Matumbi, S6 Chopi.[1]

Use has been made throughout this work (as in most contemporary Bantu linguistic studies) of language-codes deriving from Malcolm Guthrie's *Classification of the Bantu Languages* (1948). It should be made clear that this is a referential classification without any claim to historical interpretation: in Guthrie's own words it is 'a geographical classification with as much linguistic justification as possible'. (The classification distinguished 16 zones, later reduced to 15, identified by an initial capital, and within them a total of 80 language groups, identified by the first of two digits following, the second identifying the language; by general consensus an additional zone J has been extracted from the eastern part of zone D and the western part of zone E.) It will be noted in section 4.3 that Guthrie's groups frequently but by no means always emerge with coherence from the present analyses.

In order to convert the vocabularies into a countable form, judgement had to be made of which forms could be regarded as comparable. The work of cognation judgement was undertaken by André Coupez and Yvonne Bastin. For each gloss they recognised a number of roots to which they assigned numbers (root-codes), and built up in sections a table with a row for each vocabulary and a column for each gloss, entering in each cell the appropriate root-code or codes. It is this coded data that has been manipulated to produce the symbolic maps in section 2.3 in which each root-code is converted into a unique symbol and each vocabulary appears at a fixed position in each map.

In making cognation judgements, Coupez and Bastin were informed by philological principles (that is to say, they took into account the systematic sound-changes known to affect each language), but faced with minor variation they preferred to be inclusive. Coupez referred in justification to his study of 'variability' in the Bantu languages (Coupez 1975), which uncovered a high degree of freedom to generate near-homophonous synonyms (MM has referred to them as 'homeophones'). While this phenomenon is perhaps associated most especially with words that carry a high emotive load, equal justification can be found in Guthrie's 'osculant' synonyms, which appear in many cases to reflect imperfect operation of a sound-change within a complex linguistic continuum – a phenomenon familiar enough in European dialectology.

In looking at the maps, it may be useful to realise that the most widespread roots were generally given low root-codes by Coupez and Bastin, and appear on the maps with the solid black shapes ● ■ ▲ ▼ etc., while block numbers such as ❻ were used only where the number of distinct roots was very high and I was running out of alternative symbols.

Perusal of the symbolic maps will, I suggest, convince that there is no easy categorisation. There are a few glosses with largely uniform realisation, e.g. 17 die, 22 eat, 62 person and 82 tongue, and a number with reasonably clear regionalisation, e.g. 27 fire, 40 knee, 85 two; some glosses show one realisation predominating but with a variety of local innovations: 1 all, 16 come, 35 hand, 43 leg, 53 mouth. Some glosses have extremely diverse realisations, e.g. 4 bark, 32 good, 87 warm. In 3 ashes the Outer Bantu realisation is curiously echoed in the North-East. Some glosses proved difficult to obtain, e.g. 29 fly, 64 red, 66 round, 80 swim, while others yielded frequent multiple responses, e.g. 32 good, 46 long (adjectives sometimes realised as verbs), 42 man, 92 woman.

From this data the computer calculated a measure of similarity for each pair of languages, 146611 comparisons in all. There is little obvious human interest in the similarity between say C32 Bobangi and S54 Ronga, but these measures play their part in the analyses presented in the next two chapters, and it is frequently relevant to consider which languages are most closely related to a language of current interest.

[1] Derek Nurse generously offered to complement our data, but faced with accumulating delays I was unable to take up his offer.

For convenience of exposition we speak of two vocabularies/languages being 'linked' at a particular level of similarity.

In calculating the similarity indices, what was to be done about missing data and multiple responses? For multiple data, I scored a match if any root-code for language *x* agreed with any root-code for language *y*. This may have slightly boosted similarities for languages rich in doublets, but in many cases it will be noted that the same pair of codes repeat in vocabulary after vocabulary, so the effect is not likely to be large. For missing data, I counted the glosses for which data was jointly available for *x* and *y*, and expressed the matches as a percentage of this figure. To the extent that data is most often missing for 'difficult' glosses where roots are diverse and matches rarer, this may again have boosted similarities for ill-documented languages. I considered the alternative of eliminating some combination of glosses and vocabularies for which the data was most frequently defective, but total elimination was too severe, and any lesser measure would still have required a similar treatment for the remaining gaps. The number of glosses for which data is available in each language is included in the documentation in section 2.2.

2.2 Identification and sources of the vocabularies

The location of the vocabularies collected is shown in Figures 2.1.1(a)-(b). The table following was originally assembled in French by Yvonne Bastin; parts have been left in French, notably details of location referring to countries where French is among the national languages.

The first line of each entry gives a unique name (incorporating a language code indicating zone, group, language and occasionally dialect), and to the right, an abbreviated code used in the labelling of maps and diagrams. There may be more than one vocabulary representing a single language, in which case there will be a serial number after the language name followed perhaps by some other differentiating label. Note that languages with codes ending in 0 are uncertainly assigned within the group; similarly languages with codes ending in 00 are uncertainly assigned within the zone; since these codes are not unique, further mnemonic letters may appear in the labelling code for these languages.

The following line(s) give first an indication of the area in which the vocabulary was collected, followed by precise geographical latitude and longitude, which in principle represent the place where the vocabulary was recorded, or which the informant claimed was his home district; in a few cases marked by an asterisk (*) this information was not available, and the vocabulary has been placed arbitrarily within area where the language is spoken.

The final line(s) of each entry give credit to the contributors of each vocabulary. Speakers who supplied the data are identified with the symbol 🗣; the symbol ✍ identifies the many scholars who noted the vocabulary (from informants) or excerpted it from written sources (🕮: see *Bibliography of Sources* at the end of this section). An asterisk (*) following the contributors means that the vocabulary was included in the sample of 214 languages (Bastin—Coupez—de Halleux 1983). The final figure in [square brackets] gives the number of items recorded in the vocabulary (out of 92 maximum).

800 Ejagham 800
9.2°E, 5.5°N
✍ J. Watters, 1984 [92]

802 Tiv 802
Gboko Benue State, Nigeria; 9.15°E, 7.25°N
🗣 Thomas Achia (✍ B. Janssens, 1989) [92]

805 Amasi 805
Cameroun; 9.75°E, 6.1°N
🕮 Jungraithmayr & Von Funck 1975 [83]

806 Ambele 806
Cameroun; 9.55°E, 6.95°N
🕮 Jungraithmayr & Von Funck 1975 [83]

894 Asumbo 894
Cameroun; 9.5°E, 6.3°N
🕮 Jungraithmayr & Von Funck 1975 [78]

951 Ngyemboong d. bangang 951
Cameroun; 10.2°E, 5.6°N
🗣 J. Ketchankoue (✍ F.M. Rodegem, 1973) [91]

952 Yemba d. atsang 952
Cameroun; 10.1°E, 5.4°N
🗣 P.B. Nkasse (✍ F.M. Rodegem, 1973) [92]

960 Ghomala 1 960/1
Cameroun; 10.3°E, 5.2°N
🗣 J. Mandjo (✍ F.M. Rodegem, 1973) [89]

960 Ghomala 2 d. central 960/2
Cameroun; 10.3°E, 5.3°N
🗣 S. Sofo (✍ F.M. Rodegem, 1973) [92]

970 Fe'fe' 1 970/1
Cameroun; 10.2°E, 5.3°N
🗣 P. Njinte (✍ F.M. Rodegem, 1973) [92]

970 Fe'fe' 2 d. central fa' 970/2
Cameroun; 10.2°E, 5.1°N
🗣 O. Tientcheu (✍ F.M. Rodegem, 1973) [91]

970 Fe'fe' 3 d. central 970/3
Cameroun; *10.25°E, 5.15°N
🗣 M. Keou Ngassa (✍ F.M. Rodegem, 1973) [92]

A15 Mbo A15
Cameroun; 10°E, 4.8°N
🗣 A. Kingue Nseke (✍ F.M. Rodegem, 1973) [92]

A24 Duala 1 A24/1
Cameroun; 9.7°E, 3.8°N
🗣 J.E. Mouelle Mounkouri (✍ F.M. Rodegem, 1973) [92]

A24 Duala 2 d. pongo A24/2
Cameroun; 9.7°E, 4.1°N
🗣 L. Katto (✍ F.M. Rodegem, 1973) [92]

A24 Duala 3 d. malimba A24/3
Cameroun; 9.5°E, 3.9°N
🗣 C. Etongo (✍ F.M. Rodegem, 1973) [92]

A31 Bubi A31
Basakato; 8.8°E, 3.5°N
🗣 B. Boriko-Lopeo (✍ G. Guarisma) [88]

A32 Batanga 1 d. puku A32/1
Kribi, dép. de l'Océan, Cameroun; 9.9°E, 2.9°N
✍ M. Van Hille, 1989 [72]

A32 Batanga 2 d. batanga A32/2
Cameroun; 10°E, 3.3°N
🗣 A. Sadey & D. Bandiga (✍ F.M Rodegem, 1973) [92]

A32 Batanga 3 d. noho A32/3
Cameroun; 10°E, 3.1°N
🕮 Adams 1907 (✍ A.E. Meeussen) [83]

A34 Benga 1 Corisco A34/1
Corisco, Guinée Equatoriale; 9.7°E, 0.5°N
🗣 G. Menz (✍ J. Vansina) [92]

A34 Benga 2 Gabon A34/2
Estuaire (Libreville, quartier de Cointet), Gabon; 9.4°E, 0.4°N
🗣 Maman Jeanne (✍ H.A. Hazoumé) [89]

A43a Basaa A43a
Dép. Nyong et Kellé, Sanaga Maritime, Cameroun; 11°E, 3.6°N
🕮 Lemb & de Gastines 1973 (✍ B. Janssens) [92]

A43b Bakoko A43b
Songeland, Cameroun; 10°E, 3.5°N
🗣 R. Ntimba (✍ F. Rodegem, 1973) & 🗣 M. Kino (✍ E. Dodo Moundjiegou Bounguendza, 1989) [92]

A44 Nen A44
Arr. Ndikinimeki, dép. Mbam Cameroun; 10.9°E, 4.7°N
🗣 H. Boyeleba Balehen (✍ B. Janssens) [92]

A51 Lefa 1 d. tingong A51/1
Bangong, Bafia, Cameroun; 11.1°E, 4.9°N
🗣 Kadi (✍ M. Sachnine, G. Guarisma & C. Venot, 1973) [88]

A51 Lefa 2 A51/2
Ngam, Bafia, dép. Mbam, Cameroun; 11.2°E, 5°N
🗣 A. Mbue (✍ C. Venot & M. Sachnine, 1973) [75]

A51 Lefa 3 d. maja A51/3
Bafia, Cameroun; 11.2°E, 4.8°N
🗣 G. Ntsila (✍ G. Guarisma, 1973) [86]

A53 Rikpa 1 A53/1
Bafia, dép. Mbam, Cameroun; 11.1°E, 4.6°N
✍ G. Guarisma 1975 [91]

A53 Rikpa 2 A53/2
Bafia, arr. Bafia, Mbam, Cameroun; *11.1°E, 4.6°N
🗣 R. Ribana (✍ C. Venot & M. Sachnine, 1973) [90]

A54 Tibea A54
Nyafianga, Bafia, dép. Mbam, Cameroun; 11.3°E, 5°N
🗣 A. Ngo (✍ G. Guarisma, 1975) [86]

A61 Tuki A61
Cameroun; 11.4°E, 4.5°N
🗣 D. Zogo & R. Okala (✍ F.M. Rodegem, 1973) [89]

A62 Yambasa central 1 d. mmaala A62/1
Bényi, dép. Mbam, sous-préfecture de Bokito et d'Ombessa, Cameroun; 11.2°E, 4.5°N
🗣 Benenegnie Moussa (✍ Y. Nzang-Bie, 1990) [85]

A62 Yambasa central 2 d. nu yangben A62/2
Yangben, Préfecture Bafia, dép. Mbam, Cameroun; 11.1°E, 4.3°N
🗣 A. Oloko (✍ C. Paulian) [92]

A66 Nu Gunu A66
Dép. Mbam; arr. d'Ombessa et de Bokito, Cameroun; 11.3°E, 4.6°N
🗣 Br. Kamana-Bodiong (✍ J.P. Rekanga, 1989) [81]

A72 Ewondo 1 A72/1
Yaounde, Cameroun; 11.5°E, 3.9°N
🗣/✍ J.M. Essono [92]

A72 Ewondo 2 A72/2
Cameroun; 11.2°E, 3.6°N
🗣 L. Etogo Mbezele (✍ F.M. Rodegem, 1973) [92]

A72 Ewondo 3 d. eki A72/3
Ngoro, arr. Ntui, dép. Mbam, Cameroun; 11.7°E, 4.4°N
🗣 B. Awom (✍ G. Guarisma) [89]

A72 Ewondo 4 d. eton 1 A72/4
Cameroun; 11.2°E, 4°N
✍ F.M. Rodegem, 1973 [92]

A72 Ewondo 5 d. eton 2 A72/5
Cameroun; 11.3°E, 3.9°N
🗣 R. Ndongo (✍ F.M. Rodegem, 1973) [92]

A74 Bulu A74
Cameroun; 11.5°E, 3°N
🗣 G. Mimbo (✍ F.M. Rodegem, 1973) [91]

A75 Fang 1 d. ntumu A75/1
Adzan-Esamdom, région Oyem, prov. Woleu-Ntem, Gabon; 11.6°E, 1.62°N
🗣/✍ P. Ondo-Mebiame, 1989 [91]

A75 Fang 2 d. make A75/2
Aloumendong, prov. Estuaire, Gabon; 10°E, 0.5°S
🗣/✍ Y. Nzang Bie, 1989 [89]

A75 Fang 3 A75/3
Souanke, région Sangha, Congo; 14.1°E, 2.1°N
🗣 Souanke (✍ J. Ndamba & Ch. Lia, 1989?) [90]

A84 Koozime A84
Eko', Haut-Nyong, Cameroun; 13°E, 3.4°N
🗣 S. Mpiale (✍ K. Beavon) [92]

A86 Mpo 1 d. bekwil Gabon A86/1
Gabon; *14.15°E, 1.35°N
🗣 J.V. Mesa-Mvele (✍ M. Dufeil, 1986) [82]

A86 Mpo 2 d. bekwil Gabon: Mayibot A86/2
Mayibot 2, Hakoku, Gabon; 14.1°E, 1.3°N
🗣 G. Igindzita & Mayibot (✍ C. Marchal-Nasse, 1988) [91]

A86 Mpo 3 d. bekwil Congo A86/3
Messok (Souanke), région Sangha, Congo; 14.2°E, 1.8°N
🗣 J. Aselam (✍ G. Elounga) [92]

A86 Mpo 4 d. mpiemo Cameroun A86/4
Yokadauma, Boumba et Ngoko, région Sangha, Cameroun; 15°E, 3.5°N
🗣 L. Wanabetsia (✍ K. Beavon) [92]

A86 Mpo 5 d. mpiemo Centrafrique A86/5
Nola, Centrafrique; 16.05°E, 3.4°N
✍ E. Voeltz, 1989 [92]

A87 Bomwali 1 A87/1
Kandeko, Ouesso, région Sangha, Congo; 15.7°E, 1.55°N
🗣 C. Essakiba (✍ Ch. Lia & J. Ndamba) [89]

A87 Bomwali 2 d. Iino A87/2
Moloundou (Cameroun) + région Ouesso, Congo; 15.15°E, 2.2°N
🗣 Awitiku Samuel (✍ K. Beavon) [87]

A92b Pomo A92b
Ouesso, région Sangha, Congo; 16.2°E, 1.7°N
🗣 G. Mokossi Djomai (✍ J. Ndamba & Ch. Lia) [90]

A93 Kako 1 Cameroun A93/1
Cameroun; 14°E, 4.3°N
℘ G. Moung (✍ F.M. Rodegem, 1973) [91]

A93 Kako 2 Centrafrique A93/2
Gamboula, Centrafrique; 15.2°E, 4.1°N
✍ E. Voeltz, 1989 [92]

A94 Kweso 1 Centrafrique A94/1
Ngula, Sangha, Rép. Centrafricaine; *16.2°E, 3.5°N
℘ E. Bébet (✍ K. Beavon) [92]

A94 Kweso 2 Congo A94/2
Sangha, Congo; 16.3°E, 1.8°N
℘ D. Goumalengue (✍ K. Beavon) [87]

B11a Mpongwe B11a
Libreville, Estuaire, Gabon; *9.45°E, 0.3°N
℘ E. Bebedi (✍ C. Marchal-Nasse) [91]

B11b Rungu-Nkomi d. nkomi B11b
Région de Fernan Vaz, Gabon; 9.22°E, 1.3°S
℘ E. Ogandaga (✍ Cl. Grégoire, 1989) [91]

B11c Galwa 1 B11c/1
Gabon; 9.5°E, 0.9°S
℘ M. Ossoucah Owanga (✍ M. Dufeil, 1986) [91]

B11c Galwa 2 Lambarene B11c/2
Lambarene, Gabon; 10.1°E, 0.8°S
℘ G. Mbezo (✍ A. Coupez, 1974) [92]

B20 Sama B20
Makokou, province Ivindo, Gabon; 12.85°E, 0.55°N
℘ Mouba (✍ J. Vansina) [92]

B21 Seki B21
Quobin, Libreville, Gabon; 9.5°E, 0.5°N
℘ G. Kinkata (✍ J. Vansina) [92]

B22b Ngom B22b
(= akele), Moyen-Ogooué, Gabon; 10.5°E, 0.5°S
℘ J.-Ch. Madouma (✍ M. Dufeil, 1986) [92]

B23 Mbangwe B23
Masuku, Haut-Ogoué, Gabon; 13.4°E, 1.7°S
B. Mbangalivoua Maka (✍ E. Eyindanga & C. Marchal-Nasse, 1988) [76]

B24 Wumvu 1 Gabon B24/1
(d'est en ouest entre Mbigou et Divenie au Congo), Gabon; 12°E, 2°S
℘ D.M. Mikeni (✍ M. Dufeil, 1986) [92]

B24 Wumvu 2 Congo B24/2
Congo; *12.1°E, 2.7°S
✍ L.Y. Bouka, 1989? [89]

B24 Wumvu 3 Kissiele B24/3
Kissiele, Niari, Congo; 12.5°E, 2.7°S
℘ F. Mouyikou (✍ J. Ndamba & J. Baka, 1989) [90]

B25 Kota 1 madungwe B25/1
Etakangai, Canton nord, Mekambo, prov. Ogoué, Gabon (kota madungwe); 14°E, 1.1°N
℘ Kadima Mbola Zamba (✍ P. Piron, 1989) [89]

B25 Kota 2 B25/2
Gabon; *14°E, 1.15°N
✍ Medjo Mve Pither & C. Marchal-Nasse, 1988 [89]

B25 Kota 3 B25/3
Mekambo, Gabon; 13.9°E, 1.1°N
℘ N. Mafomangoya & F.C. Moatekouba (✍ C. Marchal-Nasse, 1988) [92]

B25 Kota 4 Mbomo B25/4
Mbomo, région de la Cuvette, Congo; 12°E, 0.7°S
℘ A. E. Mbelibadi (✍ L. Polak, 1982) [91]

B26 Mahongwe B26
Amendemba, Canton Bogo, région Mekambo, nord-est Gabon; 14.3°E, 1°N
✍ J. Vansina [92]

B27 Sake 1 B27/1
(Booué) Ogooué-Ivindo, Gabon; 11.9°E, 0.1°S
℘ J.M. Benga-Mboudza (✍ C. Marchal-Nasse, 1986?) [85]

B27 Sake 2 B27/2
Gabon; 11.8°E, 0.2°S
✍ C. Marchal-Nasse, 1988 [85]

B28 Ndasa 1 B28/1
District de Komono, région Lekoumou, Congo; *13.2°E, 3.3°S
℘ R. Moutsimba (✍ A. Lipou, 1989) [90]

B28 Ndasa 2 B28/2
Congo; 13.2°E, 3.4°S
✍ L.Y. Bouka, 1989 [88]

B31 Tsogo B31
Mimongo, prov. Ngounié, Gabon; 11.5°E, 1.75°S
℘ J. de D. Moumbegna (✍ C. Marchal-Nasse) [88]

B33 Pinji B33
Prov. Ngounié, rive droite, en aval de Mouila, Gabon; 11°E, 1.8°S
✍ C. Marchal-Nasse, 1988 [91]

B34 Pove B34
Gabon; 12.2°E, 1.2°S
✍ C. Marchal-Nasse, 1986 [90]

B36 Himba B36
Minongo (Eteke), prov. Ngounie Gabon; 11.54°E, 1.48°S
℘ L.M. Embiault (✍ J.P. Rekanga 1989) [86]

B40 Bwali B40
Ngounié, sud Gabon; 10.6°E, 1.4°S
℘ Marly-Mounguiba (✍ C. Marchal-Nasse) [90]

B41 Shira B41
Dakartango, prov. Ngounié, Gabon; 10.5°E, 1.5°S
℘ Moussanga (✍ J. Vansina) [92]

B42 Sangu 1 B42/1
Mimongo, prov. Ngounié, Gabon; 11.58°E, 1.62°S
℘ J.M. Mombo-Tsoungou (✍ P. Ondo-Mebiame, 1988) [91]

B42 Sangu 2 B42/2
Autour de Mimongo, Koulamoutou, Gabon; 12.5°E, 1.2°S
℘ R. Nzambi (✍ C. Marchal-Nasse) [91]

B43 Punu 1 Gabon B43/1
Gabon; 11°E, 2.5°S
℘ A.B. Boulinguy (✍ F.M. Rodegem, 1973) [91]

B43 Punu 2 Congo B43/2
Divenie, Niari, Congo; 12.1°E, 2.7°S
✍ J. Ndamba & J. Baka, 1989 [91]

B44 Lumbu 1 Gabon B44/1
(Entre la zone pounoue et la mer, au nord de Mayumba), Gabon; 10°E, 2.8°S
✍ C. Marchal-Nasse, 1987? [92]

B44 Lumbu 2 Banda B44/2
Mbandu , Banda, province Niari, Congo; 11.9°E, 3.5°S
℘ J.R. Mouanda (✍ G. Elounga, 1988) [92]

B44 Lumbu 3 Nkola B44/3
Nkola, Madingo-Kayes, Kouilou, Congo; 12.7°E, 4.1°S
℘ J.P. Usangila (✍ J. Ndamba, 1989) [92]

B44 Lumbu 4 Gabon B44/4
Libreville, Estuaire, Gabon; *9.45°E, 0.3°N
℘ M.-Th. Moanda (✍ V. Koumba & C. Marchal-Nasse, 1989) [92]

B45 Bwisi B45
Loubetsi, région Niari, Congo; 13.1°E, 3.25°S
℘ D. Moutete (✍ J. Ndamba & J. Baka, 1989) [91]

B46 Varama 1 B46/1
Ogooué-Maritime, Gabon; *9.7°E, 2.5°S
✍ C. Marchal-Nasse, 1986? [90]

B46 Varama 2 B46/2
Entre Sette Cama et Gamba, Ogoué Maritime, Gabon; *9.8°E, 2.6°S
℘ Ch. Mumbo (✍ Ch. Mumbo & Cl. Grégoire, 1990) [91]