

STUDIES IN  
LANGUAGE  
COMPANION  
SERIES 90

# From Linguistic Areas to Areal Linguistics

*Edited by*  
Pieter Muysken

JOHN BENJAMINS PUBLISHING COMPANY

## From Linguistic Areas to Areal Linguistics

## *Studies in Language Companion Series (SLCS)*

This series has been established as a companion series to the periodical *Studies in Language*.

### **Editors**

Werner Abraham  
University of Vienna

Michael Noonan  
University of Wisconsin-Milwaukee USA

### **Editorial Board**

Joan Bybee  
University of New Mexico

Ulrike Claudi  
University of Cologne

Bernard Comrie  
Max Planck Institute, Leipzig  
University of California, Santa Barbara

William Croft  
University of New Mexico

Östen Dahl  
University of Stockholm

Gerrit J. Dimmendaal  
University of Cologne

Ekkehard König  
Free University of Berlin

Christian Lehmann  
University of Erfurt

Robert E. Longacre  
University of Texas, Arlington

Brian MacWhinney  
Carnegie-Mellon University

Marianne Mithun  
University of California, Santa Barbara

Edith Moravcsik  
University of Wisconsin, Milwaukee

Masayoshi Shibatani  
Rice University and Kobe University

Russell S. Tomlin  
University of Oregon

### **Volume 90**

From Linguistic Areas to Areal Linguistics  
Edited by Pieter Muysken

# From Linguistic Areas to Areal Linguistics

*Edited by*

Pieter Muysken

Radboud University Nijmegen

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

#### Library of Congress Cataloging-in-Publication Data

From linguistic areas to areal linguistics / edited by Pieter Muysken.

p. cm. (Studies in Language Companion Series, ISSN 0165-7763 ; v. 90)

Includes bibliographical references and index.

1. Areal linguistics. I. Muysken, Pieter.

P130.F76 2008

409--dc22

2007034442

ISBN 978 90 272 3100 0 (Hb; alk. paper)

© 2008 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands  
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

# Table of contents

Editor and contributing authors	VII
1. Introduction: Conceptual and methodological issues in areal linguistics <i>Pieter Muysken</i>	1
2. The problem of the Caucasian Sprachbund <i>Viacheslav A. Chirikba</i>	25
3. East Nusantara as a linguistic area <i>Marian Klamer, Ger Reesink and Miriam van Staden</i>	95
4. The Guaporé-Mamoré region as a linguistic area <i>Mily Crevels and Hein van der Voort</i>	151
5. An integrated areal-typological approach: Local convergence of morphosyntactic features in the Balkan Sprachbund <i>Olga Mišeska Tomić</i>	181
6. Zhuang: A Tai language with some Sinitic characteristics: Post-verbal ‘can’ in Zhuang, Cantonese, Vietnamese and Lao <i>Rint Sybesma</i>	221
Language index	275
Author index	281
Subject index	285
Place index	291



## Editor and contributing authors

Pieter Muysken is professor of linguistics at Radboud University Nijmegen, after previously working at the Universities of Amsterdam and Leiden. He was the recipient of a Spinoza Research Prize in 1998, and the current volume grew out of the research program Lexicon and Syntax set up with the funds from this prize.

Slava (Viacheslav) Chirikba is assistant professor at the department of Middle Eastern languages at Leiden University and a specialist on Caucasian languages.

Mily Crevels is working as a postdoctoral researcher at Radboud University Nijmegen, on a project describing and comparing endangered Bolivian languages. Her specialism is language typology and Bolivian languages.

Marian Klamer is assistant professor at the department of Indonesian languages at Leiden University and the author of *A Grammar of Kambera*.

Ger Reesink is a postdoctoral researcher at the Radboud University Nijmegen, on a project describing and comparing endangered

Papuan languages. His expertise lies in that area.

Miriam van Staden is currently a postdoc at the University of Amsterdam and author of *Tidore*, a linguistic description of a language of the North Moluccas.

Rint Sybesma is assistant professor at the Chinese department of Leiden University, and carries out a research project on southern Chinese languages and their neighbors.

Olga Mišeska Tomić is emeritus professor of English at the University of Novi Sad (Serbian Republic) and was a research fellow at Leiden University while working on this project. She recently edited a volume on the grammar of the Balkan languages with Benjamins (*Balkan Syntax and Semantics*) and has completed a monograph on the same topic.

Hein van der Voort is currently a postdoctoral researcher at Radboud University Nijmegen, on a project describing and comparing the endangered languages of the Brazilian State of Rondonia. He is the author of *A Grammar of Kwaza*, and a specialist on the languages of Rondonia.





# Introduction

## Conceptual and methodological issues in areal linguistics

Pieter Muysken

Linguistics, Radboud University Nijmegen

### 1. Preliminary

This book aims at further exploring the approach to language description and comparison in terms of the regional context in which languages are spoken. It does this by presenting a number of case studies in areal linguistics, and it covers linguistic areas from four continents. Some of these concern a well-established linguistic area, such as the Balkan, others document regions which have never been studied in an areal perspective before, such as East Nusantara and the Guaporé-Mamoré region (involving parts of lowland Bolivia and the adjacent Brazilian state of Rondonia), yet others involve areas which have been the subject of debate and recent scrutiny, such as the Caucasus and South-East Asia. All studies present cultural and historical background, as well as detailed information. However, they differ in their scope and linguistic theoretical framework, although they all stress the significance of language contact factors in process of change.

In an extensive survey, Slava Chirikba convincingly shows that there is good reason to accept the reality of a Sprachbund in the **Caucasus**, basing himself on a host of little known materials, as well as on his own field research. Chirikba lists a whole range of features, phonological, morphological, syntactic, semantic, and lexical which are shared by the Kartvelian and North Caucasian language groups, and as such are characteristic of the Caucasus as a whole. Chirikba also discusses in considerable detail processes of historical intermingling etc. that produced this result.

Similarly, Marian Klammer, Ger Reesink, Mirjam van Staden present the historical and the linguistic evidence for **East Nusantara** (the archipelagoes of eastern Indonesia and the adjacent regions of Papua) as a linguistic area. In this area Austronesian languages and Papuan languages have influenced each other. Five features are focused upon: possessor-possessum order in adnominal possession, the overt marking of the distinction alienable vs. inalienable possession, and clause-final negation have diffused into the Austronesian languages from the Papuan languages, while two features have gone the other way: SVO as primary constituent order, and an inclusive/exclusive opposition in the pronominal paradigm.

With respect to the Amazonian state of Rondonia in Brazil and the adjacent lowland areas of Bolivia, Mily Crevels and Hein van der Voort explore the possibility that what they term the **Guaporé-Mamoré** region be considered a linguistic area, citing both ethno-historical and archeological evidence and data from recent research by themselves and members of their research group. Among the linguistic features explored we find various types of classifiers, as well as a high incidence of prefixes, the use of evidentials and directionals, occurrence of verbal number but not so much of nominal number, and the inclusive/exclusive distinction.

Using a more theoretical generative approach, Olga Mišeska Tomić explores areas of local convergence in the **Balkan Sprachbund**, predominantly on the basis of morpho-syntactic features. These include various nominal and pronominal case markings, the pronominal clitic system (clitic doubling, possessive clitics, impersonal clitics), the verbal perfect, and evidentials. In addition to the data in the vast research literature, Tomić adduces evidence from her own fieldwork. Not just the standard varieties of the Balkan languages are taken into account, but a host of individual local varieties.

Finally, Rint Sybesma explores the syntactic and semantic properties of a single feature, postverbal ‘can’ in four languages from **South East Asia**: Zhuang, Cantonese, Vietnamese and Lao. Zhuang is argued to show some Sinitic characteristics, even though it is a Tai language. This is an areal feature shared by languages spoken in Indo-China and Southern China; the modal element occurs in a for these languages a-typical, post-verbal position, which is further evidence of its contact-induced occurrence.

Thus these five studies illustrate different approaches to the concept of linguistic area. The insight has been rapidly gaining ground that languages owe many of their characteristics to the languages with which they have been in contact over time, languages either related to them or not. Yet the nature of influences and the way these areal influences took hold remains a matter of debate. To cite Sally Thomason, “. . . what we understand about linguistic areas is depressingly meager, compared to what we don’t understand about them” (Thomason 2001: 99). Furthermore, areas are often hard to define. Hence the title of this book, which reflects a shift in emphasis from linguistic areas as concrete, well circumscribed and relatively unique entities, to a way of doing linguistics in an areal perspective. Researchers such as Stolz (2002) have called for an absolute cessation of research on Sprachbund phenomena; while his methodological observations on how difficult it is to define an area certainly make sense, I do not take this as a call to stop doing linguistic research from an areal perspective.

New findings in this book include the observation that there may be many more language areas than previously been recognized, leading to a more general ‘areal’ perspective on comparative and descriptive linguistics. Inevitably, this will lead to a new series of methodological issues, e.g., involving the sampling of both languages and features, and methods for establishing the significance of findings. Also, there is

the need for increased awareness of new technique for data storage and comparison, relying on typological databases and map representation of features and items.

It is also clear that comparative research is needed, now that new ‘linguistic areas’ are being discovered in different parts of the world, involving different language types and different contact situations. This comparative approach will lead to a new typology of linguistic areas. In addition to the “core area / peripheral area” model implicit in many descriptions, there will be attention for chains of overlapping areas, diaspora communities as linguistic areas, clusters of immigrant languages in a migration setting as linguistic areas, etc. What to do with the white holes in an area, i.e., the non-converging languages? What is the role of size: should we distinguish micro / meso / macro linguistic areas? I return to some of these questions below.

Chirikba’s paper contains an excellent historical overview of the history of the notion of linguistic area or Sprachbund. The amount of attention paid to areal approaches in comparative linguistics is considerable and still growing, particularly at the descriptive level. This book is a further example of this trend, and presents, as stated, five detailed case studies of linguistic areas. We can only gain understanding of linguistic areas through two approaches: further detailed examination of specific linguistic areas to discover their properties, on the one hand, and reflection upon scenarios for their genesis and upon ways of identifying linguistic areas and studying them, one the other.

Complementing the case studies, this introduction will primarily take the second approach, that of theoretical reflection. It contains four main parts. In section 2 I dissect the definition of linguistic area provided by Thomason, as a way of tackling some conceptual and methodological issues. Section 3 contains a discussion of some of the scenarios that might be invoked in explaining the emergence of linguistic areas. This is illustrated where possible with a large linguistic area, the Atlantic, to illustrate some of the issues we are faced with in section 4, where I also draw some conclusions from the earlier discussion and relates it to the perspective taken in the five case studies presented in this book.

## 2. A definition and its dimensions

Thomason (2001: 99) defines a linguistic area or, alternatively Sprachbund (Trubetzkoy 1928), as “... a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor.” This definition contains six key concepts, which may serve to illuminate the problems at hand.

- a. Geographical region
- b. Three or more languages
- c. Shared structural features
- d. Contact

- e. Not accident
- f. Not inheritance

In the following subsections I will discuss each of these concepts in turn.

## 2.1 Geographical region

The notion of geographical region would seem uncontroversial, but actually is somewhat complex, for various reasons. First of all, because in geography, and consequently in fields like dialect geography there has been an evolution from space to spatiality (Britain 2002: 604). In the changing conceptions of spatiality (Britain 2002: 606 sq. citing Massey 1984, 1985), before 1960: space was viewed as region (with the key words: place, difference, distinctiveness), but due to the quantitative revolution in the social sciences in the 1960s: the notion of space exploded. Euclidean space gave way to social space (defined in terms of network links), and social space in turn was redefined as perceived space (social distance). For the study of linguistic areas this view of space has the implication that geographical contiguity is not the crucial notion, but rather communicative networks, affinities, and social ties. Adjacent languages need not influence one another profoundly; in contrast, language influence may involve communities geographically distant but linked by terrestrial or maritime trade or migration. In the final part of this introduction I illustrate such a non-contiguous linguistic area.

Second, there is the issue of *scale*. In current typological work very large areas are postulated, on the basis of the distribution of typological features. An area of this type might be Eurasia, or the circum-Pacific region, as in Johanna Nichols' (1992) work. Yet, the scenario for the emergence of these areas is hard to link to current work in language contact studies, which involve much smaller areas in which concretely identifiable speech communities exist, as in Gumperz and Wilson's classic (1971) study of convergence in an Indian speech community, or Aikhenvald's (2002) study of the Vaupés River region in Amazonia. In between these very large areas as in linguistic typology and the small areas that are the subject of the language contact literature we find the middle sized areas of historical linguistics, such as the Balkan or Ethiopia. On an abstract level, these are all areas, but from the perspective of concrete research, they are quite distinct. It is clear that the smaller the scale of the postulated area, the easier it is to apply what Campbell (1985) has dubbed the historicist approach, which involves the actual documentation of what influences there were, and in which direction. For large areas, the 'circumstantialist' approach, which simply catalogues similarities between languages, is probably the maximum attainable. In Table 1 I present the differences between the approaches schematically.

## 2.2 Three or more languages

The requirement that three or more languages be involved in a linguistic area is a practical one for Thomason: with just two languages any old language pair which has

**Table 1.** Levels of scale in the analysis of linguistic areas

Level	Space	Time	Sources	Scenarios
<i>Micro</i>	Bilingual community	20–200 years	Fieldwork data	Specific contact scenarios
<i>Meso</i>	Geographical region	Generally 200–1000 years	Comparative data; historical sources	Global contact scenarios
<i>Macro</i>	Larger areas of the world	Deep time	Typological, genetic, archeological data	Vague or no contact scenarios

converged would be treated as a linguistic area. This is a sensible restriction, of course, but it does raise a complicated issue, namely that of exclusiveness: how special are linguistic areas? In the original perception of researchers, linguistic areas are somewhat exceptional: we find them here and there on the globe, but their number is in the tens rather than hundreds. However, this perception also differs from sub-discipline to sub-discipline and from continent to continent.

For North America, there is a long tradition of areal comparisons, dating back to Boas (1911) and Kroeber (1939), and with links to anthropological studies about cultural diffusion. More recent scholars working on linguistic areas here include Mary Haas and Joel Sherzer. Originally, little distinction was made between genetic and diffusionist perspectives on areal resemblances, but later this changed, and more recently the issue of cultural diffusion and culture areas triggered a considerable interest in areal approaches. This is evidenced in the large number of areas distinguished in North America, following the earlier tradition of culture areas. The same thing cannot be said for Central and South America, although in the 1970s Mesoamerica was identified as a linguistic area. For the Old World, the attention paid to areal phenomena differs widely from region to region. In Table 2 an overview is given of a number of linguistic areas in the world that have been identified or postulated.

It is clear that on the one hand, a large variety of different regions has been identified, but on the other hand, the areas covered do not involve the whole globe. Further studies are continuously appearing.

### 2.3 Shared structural features

Given the ease with which words are borrowed, there is widespread agreement that linguistic areas based on shared vocabulary would not be interesting. This conclusion is entirely reasonable but it leads to a paradox. Languages in linguistic areas are assumed to have distinct lexicons, but shared grammatical features. Yet there are implicational borrowing hierarchies such as the in (1).

- (1) Words < Sounds < Grammar patterns < Word order < Meaning distinctions

**Table 2.** Selective overview of reported linguistic areas

Region	Source
<i>The Americas</i> (based on Campbell 1997a).	
Northern Northwest Coast	Leer 1991
Northwest Coast	e.g., Sherzer 1973
Plateau	e.g., Sherzer 1973
Northern California	Haas 1976
Clear Lake	Callaghan 1964
South Coast Range	Hinton 1991
Southern California/Western Arizona	Hinton 1991
Great Basin	Sherzer 1973, 1976
Pueblo	Bereznak 1995
Plains	Sherzer 1973
Northeast	Sherzer 1976
Southeast	Haas 1969
Mesoamerica	Kaufmann 1973
Colombia/Central America	Constenla 1991
Vaupés	Aikhenvald 2002
Guaporé/Mamoré	Crevels/van der Voort (this volume)
Southern Cone	Klein 1992
Venezuelan/Antillean	Constenla 1991
Ecuadorean-Colombian	Constenla 1991
Amazon	Derbyshire 1987
Lowland South America	Doris Payne 1990a, David Payne 1990
<i>Eurasia and Pacific</i>	
Balkan	Trubetzkoy 1928
Baltic	Koptjevskaja-Tamm and Wälchli 2001
Caucasus	Chirikba (this volume)
Mediterranean	Ammann (in prep.)
Europe	Haspelmath 1998, 2001
Ethiopian highlands	Leslau 1945; Tosco 2000
Benin/Surinam	Muysken and Smith (in prep.)
South Asia	Masica 1976; Emeneau 1980
South China/Indochina	Enfield (2003), Sybesma (this volume)
East Indonesia/Papua	Klamer, Reesink, van Staden (this volume)
Sepik River basin	Foley 1986

Grammatical features are much harder to borrow than lexical elements, and among lexical elements there are various types of hierarchies as well. This can be expressed in terms of the *embeddedness* paradox, which holds that the more embedded a linguistic feature, the more significant is its spread. At the same time, the more embedded a linguistic feature, the less likely it is borrowed. This paradox

is formulated by Campbell (1997b: 51) as follows: ‘The more internally motivated the universal or typological pattern, the less likely it is that areal diffusion will be permitted to introduce departures from the expected principled arrangements of elements in a language.’

Related to this is what might be called the *regularity* paradox, namely that language contact phenomena are rule-governed. Yet the results of contact in areal linguistics may appear haphazard. Thus regarding certain rare word order types (Greenberg 1966: 7, 18, 19, 20) Campbell (1997b: 51) remarks: ‘Most (perhaps all) languages representing these types owe significant aspects of their word order patterns to borrowing from other languages.’

Methodologically, the embeddedness and regularity paradoxes lead us to the following dilemma. To establish a certain linguistic area we may take either route (2a) or route (2b):

- (2) a. *Top-down*  
Establish a universal check list of key typological features and calculate the distance between the languages in an area in terms of their values for these features;
- b. *Bottom-up*  
Establish one-to-one correspondences of very specific features between individual languages and then generalize the resulting list to all languages in the area.

The advantage of method (2a) is less bias, and the possibility of establishing significance across linguistic areas: which languages show a clustering of properties? The disadvantage of this might be that there is interference from typological patterning. What could be areal may alternatively be the result of typological pressures (cf. also Campbell 1998).

The advantage of (2b) is that it relies on specific borrowed features, in many cases an ad hoc unorganized list, with probably less interference from typological pattern pressure. The disadvantage would be that there is no guarantee that the feature set chosen is representative for language patterns in general. Another feature set might yield an entirely different, half overlapping linguistic area.

## 2.4 Contact

The next issue is the precise notion of contact we need to invoke in accounting for linguistic areas. This will be the topic of the next section, but a few preliminary remarks are in order here.

A first question is, how important is the degree of bilingualism? Obviously, very slight bilingual contacts would not suffice to bring about any relevant changes. However, by no means all cases of linguistic areas documented show evidence of intimate and



widespread community-level bilingual language use in everyday settings. Some of the situations, such as the Circum-Baltic linguistic area, showed extensive commercial contacts etc., but no large scale population movements. Possibly, prestige plays an important role here as well.

A second question would be: how important is time depth? Again, no hard and fast rules can be given. We know from the genesis of creoles that in the right circumstances linguistic change can be extremely rapid, e.g., 50 years in the case of the Surinam creoles. On the other hand, some of the putative linguistic areas may have taken thousands of years to form.

Of course we would like to have non-circular historical research results confirming that there was contact, possibly with written sources for the languages involved in the pre-contact period. Sometimes there is archaeological evidence for the contact, only reliable when the remains can be directly linked to specific language groups. Finally, the geographical distribution of language families may be a secondary source of evidence for language contact.

## 2.5 Not accident

It is immediately obvious that we would not want the similarities between the languages involved to be accidental. Nobody will claim areal influence if the similarity between two languages is limited to the vowel [a] and the category plural. It is more interesting if two adjacent unrelated languages share the category trialis and the rounded front vowel [ü]. Convincing claims of areal influence involve a number of highly specific and unusual shared features. However, this immediately leads to the *markedness* paradox: The more marked a shared feature, the more likely it is due to contact, but at the same time the more marked a shared feature, the less likely it is borrowed. Marked features do not transfer as easily as unmarked features.

A special role is reserved here for discourse properties. Are these completely separate from grammatical properties or is their dynamism the result of conversational and intonational salience?

## 2.6 Not inheritance from a common source

This seemingly speaks for itself, since the very idea of a linguistic area involves to some extent unrelated language varieties. Paradoxically, however, as pointed out by Dahl (2001), many cases of areal spread involve closely related languages, most strikingly of course dialects, where we find that non-coinciding isoglosses are the rule rather than the exception. This often makes it difficult to establish whether a feature has been inherited or not. Suppose we have three related languages A, B, and C sharing a feature X. Then it is easily possible that the feature spread from A to B and C (after entering A from an unrelated language Q), rather than being inherited from the ancestor to A, B,

and C. This can not be decided on the basis of the distribution of the feature in A, B, and C by itself.

### 3. Contact scenarios

Having surveyed the definition of a linguistic area in some detail, I now turn to the possible contact scenarios that could potentially be responsible for the genesis of such an area. How can we explain the many typological similarities between the languages in a large area? For this, it is important to establish a typology of more or less understood contact phenomena. In the context of this paper, I can only mention these briefly, referring to the vast amount of literature in this area, most recently Winford (2003). Invoking known scenarios to understand historical contact developments is in line with the uniformitarian methodological principle that changes in the past must be understandable in terms of changes that have been observed more recently or that are currently taking place. The following contact scenarios can be distinguished. There is some overlap between them, but there is sufficient ground to keep them apart. In each case, a proto-typical example will be given.

*Borrowing.* As has been the case in English with French words after the Norman invasion, languages may borrow massive amounts of vocabulary from one another, without concomitant grammatical changes. In the wake of lexical borrowing, there is often the borrowing of a set of derivational affixes and lexically based word order restrictions, but often no other more fundamental grammatical changes. Distinguishing features of borrowing are:

- (3) a. It is frequent;
- b. It is generally asymmetrical: from dominant superstrate to a socially subordinate language;
- c. It involves relatively concrete features, generally 'fabric' (audible word forms, rather than patterns);
- d. It generally follows borrowability hierarchies.

Borrowing could lead to a linguistic area if there had been a shared superstrate language. Thus a group of South American languages could all separately have borrowed features from Spanish. Alternatively, there could have been chain borrowing (as we find in the spread of Quechua numerals among various languages of southern Latin America). Finally, there could have been one prestigious model language from which features spread in various directions. In all cases, however, a borrowing scenario would be best to explain the areal spread of 'surface' features such words, sounds, and possibly word order.

*Grammatical convergence under prolonged stable bilingualism.* Cases have been reported of communities where bilingual speakers have let their languages converge

over the generations. The most famous case is the also somewhat controversial paper by Gumperz and Wilson (1971). It is easy to imagine this, but we know little of the actual mechanisms involved. Distinguishing features of this kind of convergence are:

- (4) a. It may be relatively frequent, though not as frequent as borrowing by far;
- b. It is potentially symmetrical;
- c. It may lead to surface convergence, e.g., in semantic categories and in word order.

The precise linguistic effects of this type of convergence are as yet unknown.

Some researchers have claimed a link between convergence and code-switching. An example is Hudson (1996: 44), who writes: 'Features presumably spread across language boundaries as the result of bilingualism, and the preference of syntactic features among areal features may be due to a tendency among bilingual individuals to mix languages in mid-sentence. The more similar the sentence-structures are in the two languages, the easier this is; so language-mixing may encourage the suppression of syntactic differences. The areal diffusion of syntactic features is otherwise hard to understand, since syntax generally seems relatively impervious to historical change.' However, it is not known whether typical areal features correspond to frequent switching sites.

*Shift and substrate formation.* Speakers who have massively shifted to a socially dominant language may have retained patterns from their own native languages in the newly acquired second language. If these remain in the resulting new variety once the original native languages have disappeared, we speak of substrate phenomena. A typical situation would be substrate languages like Basque and Celtic that have influenced the course of development of Vulgar Latin in southwestern Europe, resulting in the Romance languages. Potentially distinguishing features of shift-induced language change are:

- (5) a. It is frequent;
- b. It is asymmetrical: from a subordinate language to a socially dominant language;
- c. It involves relatively abstract language features.

Areal phenomena that appeal to shift-induced change would most likely propose a shared substrate. The result would be shared 'deep' features, such as meaning distinctions, and patterns of grammatical organization.

*Relexification and intertwining.* Bilingual speakers have been reported to massively replace the vocabulary of their native languages with that of another language, while retaining most of the morpho-syntax, phonology and morphology of their original language. This process has been termed relexification. An example is Media Lengua, as reported in Muysken (1981; 1997), where speakers of Ecuadorian Quechua relexified their Quechua with Spanish words. The more general term intertwining has been used particularly by Bakker (1997), and leaves open the possibility that the traditional language provides the lexicon rather than the grammar. Relevant features are:

- (6) a. It is highly infrequent in its stable forms;
- b. It is often asymmetrical, in that the grammar of the original community language is retained;

- c. It results in a division of labour between the lexicon of language A and the grammar of language B.

To assume that a linguistic area has emerged from intertwining would be rather far-fetched, given the rarity of the process and its specific outcome.

*Pidgin and creole genesis.* A number of pidgins have emerged in different parts of the world, more or less stable non-native languages used in occasional communication between different groups. Creole languages have emerged in situations of intense contact with highly unequal social relations and limited presence of the dominant language, such as plantation slavery. These generally derive most of their lexicon from the dominant language, often a European colonial language, and their structures are the result of the interaction between substrate and superstrate languages, also showing the result of universal developments and simplification of inflectional morphology. Sometimes a creole has resulted from an earlier pidgin.

*Linguistic levelling.* In periods of population movements and increased communication between hitherto separate populations, intense dialect contact may lead to linguistic levelling: morphological simplification and compromise phonological traits. To the extent that leveling leads to greater similarities, conjoint displacement of languages might contribute to the overall end-result of a linguistic area, which in this case would be characterized by the dominance of unmarked features.

#### 4. The Atlantic as a linguistic area

Having briefly surveyed a number of language contact processes, as an example of the approaches one could take to tackle the problem of very large linguistic scale, I want to explore the many complex relations of mutual influence that have characterized the languages bordering on the Atlantic, which for this reason may be termed a linguistic area. As such, it takes its inspiration from the massive work by the Bremen social historian Dirk Hoerder on migrations in the second millennium (2002). However, where Hoerder takes a global perspective, I focus on the Atlantic, one of the key arenas of such migrations. Furthermore, this is an exploratory overview outlining some of the basic trends rather than the detailed treatment that the topic at hand undoubtedly deserves, but which would take an entire book. Historical linguists may scoff at looking at the Atlantic as a linguistic area since it surely has only a limited time depth, and as such could be considered more the product of global processes of globalization than of a linguistic area in the traditional sense. However, the processes involved in the linguistic development of the Atlantic region cannot be seen as really distinct in principle from what has happened before in other parts of the world.

Given the complexities of the interactions and influences involved, the Atlantic should be seen as a contact superposition zone (Koptjevskaja-Tamm and Wälchli 2001), both multiplex in time, in the sense that it has been the backdrop to different interactions over a large time span, and multiplex in space, in that it consists of a

number of sub-networks. The purpose of this section is then to briefly look at a large and diverse set of in themselves familiar phenomena from a single perspective, and moreover, a perspective that has met with some success in other parts of the world and in other branches of linguistics (Aikhenvald and Dixon 2001, Matras et al. 2006). The wider and novel perspective may help us uncover larger regularities not apparent at first sight, and help come to grips with the scale problem mentioned.

A further reason to take the Atlantic region as a point of departure is that it is a recent linguistic area; intensive links across the Ocean were only established after the European invasions in the wake of Columbus' voyages into the Caribbean. Many traditional treatments of linguistic areas take the era of European colonial expansion as the end point, the breaking off point of their analysis. For this reason, I cannot use data bases such as those in the *World Atlas of Linguistic Structures* (Haspelmath et al. 2005), which largely rely on the pre-expansion situation, as our point of reference either.

In discussing the Atlantic as a linguistic area, I have chosen for convenience sake the different continents rather than the relevant time periods as the organizational principle for my exposition, and will start with Europe, before turning to Africa, the Caribbean, and South and North America. The reason is that European nautical activities and, subsequently, the slave trade and the colonial expansion, were the motor behind many of the developments which led to the Atlantic linguistic area. In contrast with most treatments, my focus in this section will be on contact processes rather than linguistic outcomes, given the preliminary nature of this exploration.

In Table 3 some of the developments in the European context are summarized which have lead to increased contact:

Table 3. Developments in Europe that have led to increased contact

Process and region	Period	Resulting areal phenomenon
Migrations in the post-Roman era throughout Europe	From pre-history	Gradual convergence
Trading in the Mediterranean	Since Roman empire	Lingua francas, gradual convergence
Basque whale hunting in the North Atlantic		Pidgin, spread of vocabulary
North Atlantic trade in arctic regions	Since 1000	Pidgin, spread of vocabulary
Baltic	Since 1000	Gradual convergence through trade
Most of Europe	Since the Roman Empire until the present	Superstrate influence (borrowing and calquing) from (successively) from Latin, French, and English
Western Europe	Since 1000	Convergence through Norman conquest of Britain

The results of early patterns of convergence between the languages of **Europe** have been well documented in the enormous Eurotype typology project. This project has unveiled a number of structural parallels or similarities between the languages of Europe, only some of which can be attributed to common genetic roots (Haspelmath 1998; 2001). These similarities have emerged in millennia long processes of convergence and parallel development.

Haspelmath (2001) presents the following list of core grammatical features of what he terms Standard Average European:

- a distinction between definite and indefinite articles (*the woman/a woman*)
- relative clauses formed with relative pronouns (*the woman who came*)
- a 'have' perfect construction (*I have just eaten.*)
- nominative experiencers (*I think, I feel*)
- participial passives (*he was beaten*)
- anti-causative prominence (*change* [tr.] > *change* [intr.], not *change* [intr.] > *change-cause* [tr.])
- dative external possessives (*Die Mutter wäscht dem Kind die Haare.* (lit.) 'The mother washes the(dative) child the hair.')
- negative pronouns and lack of verbal negation (*I saw nobody.*)
- particles in comparative constructions (*taller than Jack*)
- relative-based equative constructions (*so Z wie X* (German), *tan Z com X* (Catalan) 'so Z as X')
- subject person affixes as strict agreement markers (*nous mangeons / wir essen*)
- distinction between the intensifier and the reflexive (*self* versus *himself*)

Haspelmath presents five possible explanations for the emergence of these common traits, and discards all but one of these:

1. Shared Proto-Indo-European features. This explanation is flawed because many of the features listed were not part of Proto-Indo-European, as far as we know.
2. A common substrate. This explanation is flawed for a similar reason; some of the features emerged later, as far as we know, at a time when a possible substrate would have been long gone.
3. The role of Latin as a lingua franca since the time of the Roman conquests. This explanation suffers from the fact that many of the common features are not part of Latin.
4. A common pan-European culture since the Renaissance and the Enlightenment. Many of the changes are too deeply embedded in the grammatical structures of the languages involved to have arisen in a relatively short period.
5. The great political transformations and migrations that characterized the transition from the Roman Empire to the early medieval period. This explanation has most possibilities, in Haspelmath's view, but it is rather unspecific.

It could well be that the shared patterns are due to a chain of events, and that we should conceive of a number of interlocking networks. After the Middle Ages, a number of developments are worthy of note.

In southern Europe, Mediterranean contact vernaculars have played an important role. A number of languages have been used in the Mediterranean in the period of trade at least since the Alexandrian period), when Koine Greek was spoken in the eastern and central Mediterranean. Greek was later replaced in part by Latin, and in the medieval period Italian dialects and the Romance-based contact vernacular *Lingua Franca* played an important role. Other languages of wider communication in the Mediterranean included Catalan, Italian, and Arabic. Thus the Mediterranean can be viewed as an important contact area (Ramat and Stolz 2002; Ammann, in prep.).

In northern Europe, the Hanseatic League, a federation of trading guilds mostly active in the Baltic and North Seas, played an important role in linking merchants in a large number of cities. The League was active from the 13th until the 17th centuries, and was centered around Lübeck and later on Dantzig. Since Low German speaking merchants settled in a number of towns, there was convergence and some levelling, with Low German as a point of reference (Koptjevskaja-Tamm and Wälchli 2001).

Dating from the period immediately preceding the European colonial expansion are the Atlantic whaling pidgins and nautical languages. Bakker (1987) describes an Icelandic pidgin with a Basque lexical base and other traces of Basque presence. The Basque have explored the waters of the northern Atlantic while whale-hunting from the 15th century onward. This has led to the creation of a Basque-Icelandic nautical pidgin (Bakker 1987) and to the presence of numerous Basque loans in coastal North-Atlantic Amerindian languages (Bakker 1989). A later pidgin is *Russenorsk*, which developed along the North Cape of Scandinavia from contacts between Norwegian and Russian (Jahr and Broch 1996).

Throughout the Middle Ages, Latin played an important role in unifying large parts of western Europe, but in the Renaissance and Reformation Latin was eventually replaced by the vernaculars, but particularly in learned vocabulary and phrases there was extensive calquing as well as direct borrowing. In the 18th century, French took over the role of Latin, again with extensive calquing and direct borrowing, and since World War II English has been acquiring the same role. In Central Europe, the Habsburg Empire played an important role in spreading German over a large region, as far as the northern Balkan. This influence affected several Slavic languages, as well as Hungarian. The effects of these developments is currently still being investigated.

We do not know what the effect was of historical developments in the pre-colonial period in **Africa**, but it is clear that there may have been considerable convergence. Greenberg (1983) lists the following common pan-African linguistic features:

- implosives
- doubly articulated labial-velar stops like /kp/ and /gb/
- initial nasal consonant clusters;
- clicks

- the lower high (or ‘near close’) vowels /u/ and /ɪ/
- tone contrasts
- the common use of adjectival verbs
- the expression of comparison by means of a verb *to surpass*

While many of these features do occur in Atlantic Africa, they completely characterize neither Africa as a whole nor specifically the Atlantic coast.

Güldemann (2007a; cf also Güldemann 2007b) argues that there is a large linguistic area spreading across Sub-Saharan Africa from the Sudan to the Atlantic, but this area is not relevant to the concerns of this article since the core of the area is located in central sub-Saharan Africa. Benue-Congo / Chadic contacts in Nigeria have been documented e.g., by Ballard (1971) and Gerhardt and Wolff (1977), while Creissels et al. (2007) discuss general morphosyntactic properties of African languages.

According to Güldemann (2007b), there were five linguistic macro-areas before the era of recent large-scale colonization:

- I The Berber spread zone
- II Chad-Ethiopia
- III The already mentioned Macro-Sudan belt
- IV The Bantu spread zone
- V The Kalahari basin

All these areas predominantly have an east-west rather than a south-north extension, due to the fact that spread involved climatically and ecologically similar environments. Thus there was no ‘Atlantic’ area in Africa; rather the coast was split up across four major areas (all but macro-area II on the list in Güldemann 2007a). Another point stressed is that these areas may not be due to a single historical process. Note that Güldemann limits his discussion to the pre-colonial period. In Table 4 a brief overview of some of the processes that have led to greater convergence in Atlantic Africa is presented.

**Table 4.** Developments in Africa that have led to increased contact

Process and region	Period	Resulting areal phenomenon
Creation of Sudanic and Guinea kingdoms due to early contacts with Muslim Berber and Arab groups	8th–14th centuries	Increased trading across the Sahara and from the Sudanic belt to the coast
Spread of European vernaculars as part of the colonization efforts on the coast	16th–19th centuries	Increasing spread of European vocabulary in the coastal languages
Slave raids, leading to demographic upheaval	15th–19th centuries	Increasing use of African <i>linguae francae</i>

(Continued)



Table 4. Continued

Process and region	Period	Resulting areal phenomenon
(particularly in the southern part of Africa) and demographic restructuring		
Creation of European colonies and creation of spheres of influence	16th–20th centuries	Spread of European national languages
Urbanization	19th–20th century	Creation of urban contact vernaculars and spread of <i>linguae francae</i> . Expanded use of varieties based on European languages

Probably the numerous *linguae francae* spoken in the regions along the West African coast have contributed considerably to the linguistic convergence in that part of Africa. Wolf (2000: 324–330) lists the following, in the region south of the sphere of Arabic influence, and going from north to south along the coast: Wolof, Bambara (slightly further inland), Dyula, Akan, Hausa (again slightly further inland), Yoruba, Fulfulde, Lingala, Kituba, and Umbundu.

The role of African *linguae francae* became more and more important due to European contact and during the slave trade, particularly in West Africa. The role of diffused European lexifier pidgins in the period of the slave trade also remains to be studied. Along the African coast a number of European-lexifier pidgins were in use, starting with a Portuguese-lexicon pidgin, and ending up with an English-lexifier pidgin.

The colonial period, slavery, and the European linguistic presence in Africa have had a profound impact. A number of European colonial languages have developed local African varieties, which show substrate influence, evidence of levelling and simplification, and lexical borrowing. West African French is spoken as an elite language alongside of West African languages in the former French colonies, including Ivory Coast, and French Cameroon. West African English is a variety which cannot be strictly separated, although it is distinct, from West African Pidgin English. African Portuguese is spoken in Angola and Mozambique as a second language, and *de facto* functions as a *lingua franca* and official language in these countries. Finally, the position of Afrikaans, related to Dutch, is complex. It is of course the native language of a part of the South African white population, but also of part of the so-called coloured population of the western Cape province. South African English similarly is both the heritage language of settlers of British origin, and now the adopted native of European and Asian immigrants.

As to the **Caribbean** and **North** and **South America**, a number of historical developments are relevant, centered around the European invasions in the 16th and 17th centuries. I will first focus on the Caribbean, and then turn to the rest of the new World, keeping in mind that many of the developments are characteristic of the whole region. In Table 5 an overview of some of the major developments is given.

**Table 5.** Developments in Caribbean and North and South America that have led to increased contact

Process and region	Period	Resulting areal phenomenon
Early contacts with Carib and Arawak populations in the Caribbean	16th century	Spread of lexical loans from a number of Amerindian languages
Coastal settlements of the Portuguese in Brazil, followed by successive exploration of the interior	15th–18th centuries	Formation of <i>Língua Geral</i> on the basis of leveled Tupi varieties with Portuguese influence.
Spread of European vernaculars as part of the colonization efforts	16th–19th centuries	Linguistic leveling and selective spread of European dialectal and nautical features
Importation of African slaves	16th–19th centuries	Transfer of West African syntactic and semantic features to the New World. Spread of African lexical items.
Transports of slaves from one colony to another	17th–19th centuries	Mutual influence between individual creole languages, both lexically and structurally
Substitutions of colonial lexifier languages	17th–20th centuries	Mutual influence between individual creole languages, both lexically and structurally
Superstrate influence on the dominated Amerindian languages	16th–20th centuries	Spread of loanwords from European languages, structural assimilation
Labor migration in the post-slavery setting between plantation societies	19th–20th centuries	Diffusion of a number of immigrant languages
Importation of contract laborers from South Asia	19th century	Spread of Bhojpuri and Hindi
Migration to the New World from a number of countries in Europe and Asia	19th–20th centuries	Diffusion of leveled varieties of a number of languages
Urbanization	19th–20th century	Further leveling of different varieties

Historical developments in the Caribbean in the pre-colonial period are not directly relevant to the later contact situation since the Amerindian populations were mostly wiped out by diseases and overexploitation, or integrated into the surrounding communities. More directly relevant is the linguistic impact of slavery in the New World, which brought speakers of a number of West African languages to the newly founded plantation colonies. There are many instances of areal feature distributions due to this, e.g., the work by Michaelis and Haspelmath (2003) on double object constructions, and the work reported on in Muysken (in prep.) and Migge and Smith (2007) on the link between Benin and Surinam.

It should also be mentioned that there were many cases of transports of slaves from one plantation colony to another. This includes: from Brazil to Cayenne, Surinam, Curaçao (Goodman 1985); from St. Eustatius to St. Thomas (Goodman (1987); from Curaçao to St. Thomas (Hesseling 1933). These slave transports contributed to the spread of features among the different creoles. Thus we find a few properties of Papiamentu in Negerhollands, and more generally, the spread of Portuguese lexemes and some Portuguese creole structural features elsewhere in the Caribbean.

Another factor contributing to greater homogeneity among the languages of the Caribbean concerns substitutions of colonial lexifier languages due to changes in the colonial power dominating a particular island, given the turbulent history of the region. In a number of cases, one colonial language replaced another in the course of colonial domination, due to political changes but also shifting migration patterns. In the Virgin Islands Dutch settlers were influential in contributing lexicon to the newly emerging creole, Negerhollands (even though islands were under Danish rule), while later English became the dominant language. Similar developments characterize St. Lucia, Surinam, the Dutch Antilles, the Berbice river in Guyana, Belize, etc. Although not as turbulent as the earlier times, in the post-slavery period there were also transports and movements of creole speakers from one area to another, as in the movements from Haiti and Curaçao to Cuba (Lipski 1994), from the Anglo-Caribbean to Aruba, from the Caribbean to West-Africa. Haitians have moved to French Guyana and elsewhere.

With regard to **North and South America** the dominant developments, to be sure, were, as in the case of the Caribbean, the importation of African slaves and the migration of Europeans to the New World. The historical developments in the pre-colonial period are perhaps less relevant in this respect because there is no evidence of large scale convergence between the Amerindian languages in the relevant period.

There is, to be sure, the diffusion of early Amerindian borrowings, Words from a number of Amerindian languages, notably Arawak, Carib, Chibcha, Quechua, and Guaraní, have rapidly diffused after early colonization, to a number of European languages, and from them, to yet other languages. Of course, the Amerindian languages, the languages along the Atlantic coast in both North and South America have been profoundly affected. The majority of the coastal languages has died out as their speakers were pushed inland or were killed by the European invaders.

Turning to the migration of Europeans to the New World, one of the most momentous consequences of trans-Atlantic contact, and ultimately in terms of numbers of speakers the most important one, is the spread of a number of European languages to the New World. New varieties have emerged of a number of languages, and there has been the important phenomenon of selection and levelling that Foster (1960) has analyzed for the cultural domain. While Dutch was spoken for some time in settler communities in North America, it now survives only in Surinam, where it is the official language and lingua franca, and in the Netherlands Antilles, where it is the language of government, though not used by many people in ordinary life. Caribbean and North American English is extremely varied and wide-spread, and today constitutes one of

the most influential language varieties in the world. French is spoken in a number of places in the Caribbean and North America, most importantly in Quebec, of course. Brazilian Portuguese is now the most important type of Portuguese, and is developing on its own and away from older European norms. Caribbean and Latin American Spanish are the most frequently spoken, if not most prestigious (a role most often reserved for peninsular Spanish) varieties in the world.

Other languages are now less widely spoken. Yiddish survives mostly in large cities in North America and Argentina, such as Montreal, New York, and Buenos Aires. In parts of Argentina, notably in pockets of Welsh settlers in the Lower Chubut Valley in Patagonia, some Welsh is still spoken. The first Welsh settlement in southern Argentina in 1863 was on the Atlantic coastline. I know of no studies specifically of Argentinian Welsh. Finally, we should mention the language of the Mennonite communities in various parts of South America, particularly in Bolivia and Paraguay. Over 30,000 members of various Mennonite congregations speak varieties of both Low and High German; there has been some dialect levelling in the course of time.

An interesting aspect concerns the African language survivals and African language borrowings. In all European colonial languages, as well as in the Caribbean creoles, in the New World words from African languages have survived (Mufwene 1993).

A third very significant development is the diffusion of South Asian koines (Bhojpuri etc.) (Barz and Siegel 1988), Chinese, and Arabic, in the Caribbean and South Africa. After the abolishment of slavery, recruitment started of contract labourers from Asia, notably North India and China, and to some extent also Java in Surinam. This led to the development of new varieties of these languages, which often underwent levelling and simplification.

Very important on both sides of the Atlantic is urbanization and large-scale migration of rural populations, a further homogenizing factor we should take into account. All along the coasts of the Atlantic, the creation of urban conglomerations has led to more uniform dialects.

As can be seen, the historical processes listed for the Atlantic region had important sociolinguistic results. These may be summarized under a number of rubrics, which include the spread of African structural features and selected lexical items to the New World, the levelling of European dialect traits in the exported varieties of the European languages, processes of morphological reduction and creole genesis, particularly in the Caribbean, and considerable European lexical and sometimes structural influence on the American Indian languages, if these have survived at all.

If we want to look for a single set of linguistic features, key linguistic traits of the Atlantic linguistic area would be, taking the Caribbean creoles as a point of departure:

- Relatively fixed SVO word order
- Preverbal marking of tense and aspect
- Limited pro-drop

- Limited morphological endings, particularly in the verb
- Analytic marking of clausal subordination

Thus the New World Spanish and Portuguese which is spoken near the Caribbean has been restructured, at least to some extent.

Obviously the above is only part of the story. Needed is detailed linguistic work on the varieties of the different languages and language families most involved in the Atlantic contact setting.

## References

- Aikhenvald, A.Y. 2002. *Language contact in Amazonia*. Oxford: OUP.
- Aikhenvald, A.Y. & Dixon, R.M.W. 2001. Introduction. In *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, A.Y. Aikhenvald & R.M.W. Dixon (eds.), 1–26. Oxford: OUP.
- Ammann, A. In preparation. The Mediterranean as a linguistic area: Synchronic structures and diachronic processes. PhD dissertation, University of Antwerp.
- Bakker, P. 1987. A Basque nautical pidgin: A missing link in the history of fu. *Journal of Pidgin and Creole Languages* 2: 1–30.
- Bakker, P. 1989. ‘The language of the coast tribes is half Basque’. A Basque-Amerindian pidgin in use between Europeans and Native Americans in North America, ca. 1540-ca. 1640. *Anthropological Linguistics* 31: 117–147.
- Bakker, P. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Oxford: OUP.
- Ballard, J.A. 1971. Historical inferences from the linguistic geography of the Nigerian Middle Belt. *Africa* 41: 294–305.
- Barz, R.K. & Siegel, J. (eds.). 1988. *Language transplanted: The development of overseas Hindi*. Wiesbaden: Harrassowitz.
- Bereznak, C. 1995. The Pueblo region as a linguistic area. PhD dissertation, Louisiana State University, Baton Rouge LA.
- Boas, F. 1911. Introduction. In *Handbook of American Indian Languages* [Bureau of American Ethnology, bulletin 40, part 1: 183]. Washington DC: U.S. Government Printing Office.
- Britain, D. 2002. Space and spatial diffusion. In *The handbook of language variation and change*, J.K. Chambers, P. Trudgill & N. Schilling-Estes (eds.), 603–637. Oxford: Blackwell.
- Callaghan, C.A. 1964. Phonemic borrowing in Lake Miwok. In *Studies in Californian linguistics*, W. Bright (ed.), 46–53. Berkeley CA: University of California Press.
- Campbell, L. 1985. Areal linguistics and its implications for historical linguistic theory. In *Proceedings of the sixth international conference on historical linguistics*, J. Fisiak (ed.), 25–56. Amsterdam: John Benjamins.
- Campbell, L. 1997a. *American Indian languages: The historical linguistics of native America*. Oxford: OUP.
- Campbell, L. 1997b. Typological and areal issues in reconstruction. In *Linguistic reconstruction and typology*, J. Fisiak (ed.), 45–72. Berlin: Mouton de Gruyter.
- Campbell, L. 1998. *Historical linguistics. An introduction*. Edinburgh: EUP.

- Constenla, A. 1991. *Las lenguas del Area Intermedia: Introducción a su estudio areal*. San José, Costa Rica: Editorial de la Universidad de Costa Rica.
- Creissels, D., Dimmendaal, G.J., Frajzyngier, Z. & König, C. 2007. Africa as a morphosyntactic area. In *Africa as a Linguistic Area*, B. Heine & D. Nurse (eds.). Cambridge: CUP.
- Dahl, Ö. 2001. Principles of areal typology. In *Language typology and language universals: An international handbook*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), Vol. 2: 1456–70. Berlin: Mouton de Gruyter.
- Derbyshire, D. 1987. Morphosyntactic areal characteristics of Amazonian languages. *International Journal of American Linguistics* 53: 311–326.
- Emeneau, M.B. 1980. *Language and linguistic area*. Stanford CA: Stanford University Press.
- Enfield, N.J. 2003. *Linguistic epidemiology. Semantics and grammar in mainland South-East Asia*. London: Routledge.
- Foley, W.A. 1986. *The Papuan languages of New Guinea*. Cambridge: CUP.
- Foster, G.M. 1960 *Culture and conquest: America's Spanish heritage* [Viking Fund Publications in Anthropology 27]. New York NY: Wenner-Gren Foundation for Anthropological Research.
- Gerhardt, L. & Wolff, E. 1977. Interferenzen zwischen Benua-Kongo- und Tschad-Sprachen. *Zeitschrift der Deutschen Morgenländischen Gesellschaft Supplement III*, 2: 1518–1543.
- Goodman, M. 1985. The origin of Virgin island Creole Dutch. *Amsterdam Creole Studies* 8: 67–106.
- Goodman, M. 1987. The Portuguese element in the American creoles. In *Pidgins and Creole languages: Essays in memory of John E. Reinecke*, G. Gilbert (ed.), 361–405. Honolulu HI: University Press of Hawaii.
- Greenberg, J.H. 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- Greenberg, J.H. 1983. Some areal characteristics of African languages. In *Current Approaches to African Linguistics*, Vol. 1. [Publications in African Languages and Linguistics vol. 1], I.R. Dihoff (ed.), 3–21. Dordrecht: Foris.
- Güldemann, T. 2007a. The 'Macro-Sudan belt': Towards identifying a linguistic area in northern Sub-Saharan Africa. In *A linguistic geography of Africa*, B. Heine & D. Nurse (eds.). Cambridge: CUP.
- Güldemann, T. 2007b. Linguistic areas without evidence for contact. Presented at the workshop on Language contact and the dynamics of language: Theory and implications. Leipzig: Max Planck Institute for Evolutionary Anthropology, May 10–13, 2007.
- Gumperz, J.J. & Wilson, R. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In *Pidginization and creolization of languages*, D. Hymes (ed.), 151–168. Cambridge: CUP.
- Haas, M. 1976. The northern California linguistic area. In *Hokan studies: Papers from the first conference on Hokan languages*, M. Langdon & S. Silver (eds.), 347–359. The Hague: Mouton.
- Haas, M. 1979. Southeastern languages. In *The languages of native America: Historical and comparative assessment*, L. Campbell & M. Mithun (eds.), 299–326. Austin TX: University of Texas Press.
- Haspelmath, M. 1998. How young is Standard Average European? In *Areal typology*, P. Ramat (ed.), 271–287. Oxford: Pergamon.
- Haspelmath, M. 2001. The European linguistic area: Standard Average European. In *Language typology and language universals. An international handbook*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), Vol. 2: 1492–1510. Berlin: de Gruyter.
- Haspelmath, M., Dryer, M.S., Gil, D. & Comrie, B. (eds.). 2005. *World Atlas of Linguistic Structures*. Oxford: OUP.

- Hesseling, D.C. 1933. Papiamentu en Negerhollands. *Tijd* 52: 265–288. Reprinted in English in Hesseling, D.C. 1971. *On the origin and formation of creoles: A miscellany of articles*, 47–61. Ann Arbor MI: Karoma.
- Hinton, L. 1991. Takic and Yuman: A study in linguistic convergence. *International Journal of American Linguistics* 57: 133–157.
- Hoerder, D. 2002. *Cultures in contact: World migrations in the second millennium*. Durham NC: Duke University Press.
- Holm, J. 1988. *Pidgins and Creoles*. 2 Vols. Cambridge: CUP.
- Hudson, R.A. 1996. *Sociolinguistics*. 2nd edn. Cambridge: CUP.
- Jahr, E.H. & Broch, I. (eds.). 1996. *Language contact in the Arctic. Northern pidgins and contact languages*. Berlin: Mouton de Gruyter.
- Kaufmann, T. 1973. Areal linguistics and Middle America. In *Diachronic, areal, and typological linguistics*, T. Sebeok (ed.), 459–483. The Hague: Mouton.
- Klein, H.E.M. 1992. South American languages. In *International encyclopedia of linguistics*, W. Bright (ed.), 4:31–35. New York NY: OUP.
- Koptjevskaja-Tamm, M. & Wälchli, B. 2001. The Circum-Baltic languages. An areal-typological approach. In *Circum-Baltic Languages*, Vol. 2: *Grammar and typology*, Ö. Dahl & M. Koptjevskaja-Tamm (eds.), 615–750. Amsterdam: John Benjamins.
- Kroeber, A.L. 1939. *Cultural and natural areas of native North America*. Berkeley CA: University of California Press.
- Leer, J. 1991. Evidence for a Northern Northwest Coast language area: Promiscuous number marking and periphrastic possessive constructions in Haida, Eyak, and Aleut. *IJAL* 57: 158–193.
- Leslau, W. 1945. The influence of Cushitic on the Semitic languages of Ethiopia: A problem of substratum. *Word* 1: 59–82.
- Lipski, J. 1994. *Latin American Spanish*. London: Longman.
- Masica, C.P. 1976. *Defining a linguistic area: South Asia*. Chicago CA: University of Chicago Press.
- Massey, D. 1984. Introduction: Geography matters. In *Geography matters*, D. Massey & J. Allen (eds.), 1–11. Cambridge: CUP.
- Massey, D. 1985. New directions in space. In *Social relations and spatial structures*, D. Gregory & J. Urry (eds.), 9–19. London: Macmillan.
- Matras, Y., McMahon, A. & Vincent, N. (eds.). 2006. *Linguistic areas. Convergence in historical and typological perspective*. London: Palgrave.
- Michaelis, S. & Haspelmath, M. 2003. Ditransitive constructions: Creole languages in a cross-linguistic perspective. *Creolica* 23/4/2003. ([www.creolica.net](http://www.creolica.net))
- Migge, B. & Smith, N. (eds.). 2007. Substrate influence in Creole formation. Special issue of *Journal of Pidgin and Creole Languages* 22:1.
- Mufwene, S.S. (ed.). 1993. *Africanisms in Afro-American language varieties*. Athens GA: University of Georgia Press.
- Muysken, P.C. 1981. Halfway between Quechua and Spanish: The case for relexification. In *Historicity and variation in creole studies*, A. Highfield & A. Valdman (eds.), 52–78. Ann Arbor MI: Karoma.
- Muysken, P.C. 1997. Media Lengua. In *Contact languages: A wider perspective*, S.G. Thomason (ed.), 365–426. Amsterdam: John Benjamins.
- Muysken, P.C. & Smith, N. (eds.). In preparation. The trans-Atlantic sprachbund: Benin and Surinam.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago IL: The University of Chicago Press.



- Payne, D.L. 1990. Some widespread grammatical forms in South American languages. In D.L. Payne (ed.), 75–87.
- Payne, D.L. 1990a. Morphological characteristics of Lowland South American Languages. In D.L. Payne (ed.) 1990b, 213–241.
- Payne, D.L. (ed.) 1990b. *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin TX: University of Texas Press.
- Sherzer, J. 1973. Areal linguistics in North America. In *Linguistics in North America*, T.A. Sebeok (ed.), 749–795. The Hague: Mouton.
- Sherzer, J. 1976. *An areal-typological study of American Indian languages north of Mexico*. Amsterdam: North-Holland.
- Stolz, T. 2002. No *Sprachbund* beyond this line. On the age-old discussion of how to define a linguistic area. In *Mediterranean languages. Papers from the MEDTYP workshop, Tirrenia, June 2000*, P. Ramat & T. Stolz (eds.), 259–281. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Thomason, S.G. 2001. *Language contact. An introduction*. Edinburgh: EUP.
- Tosco, M. 2000. Is there an Ethiopian language area? *Anthropological Linguistics* 42: 329–65.
- Trubetskoj, N.S. 1928. Proposition 16. In *Actes de 1er Congrès international de linguistes*, 17–18. Leiden: A.W. Sijthoff's Uitgevermaatschappij.
- Winford, D. 2003. *Contact linguistics*. Oxford: Blackwell.
- Wolf, H.E. 2000. Language and society. In *African languages: An introduction*, B. Heine & D. Nurse (eds.), 298–347. Cambridge: CUP.





# The problem of the Caucasian Sprachbund

Viacheslav A. Chirikba  
Leiden University

*In memory of Helma van den Berg,  
a fine Caucasian linguist and a dear friend.*

The aim of this paper is to demonstrate the fact that by many parameters, which involve all levels of linguistic structure (phonology, morphology, syntax, lexical semantics and lexicon), the Caucasus constitutes a linguistic area. The totality of numerous features shared by the two mutually unrelated indigenous Caucasian linguistic families (North Caucasian and Kartvelian) could be interpreted in terms of the Caucasian Sprachbund; this does not preclude the postulation of parallel smaller Caucasian Sprachbünde.<sup>1</sup>

## 1. Linguistic area or Sprachbund

Languages spoken in geographic proximity over a considerable period of time always interact and develop common traits. This is most visible in the diffusion of local geography- and culture-specific lexicon, as envisaged, for example, in the Carpathian region. Such language clusters will not however qualify for a language union, unless the diffusion involves substantial layers of linguistic structure, most importantly grammar. The term “linguistic area” refers to a situation when two or more languages, genetically distant or unrelated, exhibit a specific set of shared features on phonological, morphological, syntactic and lexico-semantic levels, or on several of these levels, which are explainable not in terms of genetic inheritance, universal tendencies or “elementary kinship”<sup>2</sup>, but in terms of language contact and convergence (borrowings, calques, contact-induced developments, substrate).

---

1. I would like to thank A. Aikhenvald, B.G. Hewitt, F. Kortlandt, R. Smeets and B. Tikkanen for valuable comments and suggestions on the text of this paper.

2. Here the words belonging to baby talk (*mama, papa*, etc.), or onomatopoeic (descriptive) vocabulary are meant, which look similar in many unrelated languages across the globe.

Another term used to designate a secondary linguistic community (as opposed to the “primary” community based on genetic relationship) is “language union”, often called by its German prototype *Sprachbund*, itself a calque from Russian *jazykovoĭ sojuz* “language union”. It seems that it was Nikolai Trubetzkoy who first introduced this term in his 1923 paper published in a Russian émigré magazine.<sup>3</sup> But even earlier, in his 1904 article, the Russian-Polish linguist I.A. Baudouin de Courtenay stressed the distinction between the affinities which are explained by the genetic relationship of languages (*rodstvo*) and those which are caused by convergence (*srodstvo*). As an example of the latter he referred to “the common features of the languages of the Balkan peninsula, the languages of the Caucasus, etc., completely irrespective of the presence or absence of their ancient genetic relationship. In view of all this, beside the genetic relationship of languages we have to accept also their convergent relationship [*svojstvo*, “*porodnenie*”] as a result of the mutual influence. . .”<sup>4</sup>

Like his elder colleague Baudouin de Courtenay, Trubetzkoy also cited the Balkan languages as a striking example of a language union in Europe. Furthermore, he specified that a language union can be formed not only between individual languages, but also between language families, leading thus to a “union of linguistic families” (*sojuz jazykovyx semejstv*). As examples of bigger unions he cites “the union of Uralo-Altaic language families”, consisting of unrelated Uralic, Turkic, Mongolian and Manchurian families, “the union of Mediterranean language families”, based on the nominal distinction of grammatical gender and the ablaut phenomena, which unite the Indo-European, Semitic, Hamitic and North-Caucasian families, as well as some extinct languages of the Mediterranean basin. Besides, according to Trubetzkoy, one and the same language or language family can belong at the same time to two different language unions, or fluctuate between the two unions, such as Indo-European, fluctuating between the Mediterranean and Uralo-Altaic unions (Trubetzkoy 1995: 333).

In his 1928 paper, written in German, Trubetzkoy defines the *Sprachbund* (here the term first appears in its German form) in the following way: when groups of languages display (1) considerable similarity in syntax; (2) similarity in principles of morphological structure; (3) a big number of common cultural words, and sometimes also (4) surface similarity in sound systems, whereby, however, they do not display any systematic sound correspondences, any correspondences in phonetic shape of morphological elements, and any common elementary vocabulary – such language groups can be called Language Unions (“*Sprachbünde*”).<sup>5</sup>

Sometimes the postulation of a language union is based not on several parameters of the linguistic structure, but on a single isogloss only. Within this approach one can speak about such cases as a phonological union (cf. N. Trubetzkoy’s notion of

3. Trubetzkoy (1923); reprint in Trubetzkoy (1995: 333).

4. Baudouin de Courtenay (1904); cited from Šaradzenidze (1983: 70–71).

5. Trubetzkoy (1928: 18), cited from Tuite (1999: 2).

the “Caucasian phonological union”<sup>6</sup> based on the presence of glottalized consonants, R. Jakobson’s phonological union of languages from southern Alaska up to Central California, based on the distribution of the correlation of glottalization, the “palatalizing” Eurasian union<sup>7</sup>, the distribution of retroflex consonants over unrelated languages on the Indian subcontinent)<sup>8</sup>, a morphological union (common morphological traits as the primary feature of the Balkan union), a syntactic union (cf. mainly syntactic features of Standard Average European<sup>9</sup>). A more standard approach to a language union presupposes, however, a set of specific to a particular area linguistic isoglosses/traits that relate to several layers of the linguistic structure.

The presence of grammatical (morphological, syntactic) similarities between distantly related or unrelated languages spoken in geographic proximity is crucial for the postulation of a language union. On the other hand, it would probably be a too limited approach to regard only common grammatical traits as indicative of a Sprachbund, to the exclusion of phonetics/phonology, lexicon and phraseology. Though the Balkan Sprachbund is being defined in nearly exclusively grammatical terms, in India the area-defining common traits do include phonological features (specifically, the presence of retroflex consonants) among the basic indications, beside morphology and syntax, of the pan-Indian linguistic area (cf. Masica 1976: 187). In fact, it is often phonology which is initially affected by the convergence process (cf. Aikhenvald and Dixon 2001: 17), especially intonation patterns.

From the point of view of postulating a linguistic area, it is not necessary that a certain linguistic trait be a unique property of this particular zone not found beyond its boundaries. More important is that this trait, even if not unique in itself, is specific enough to make a meaningful contrast with languages outside this area, and that, together with other traits, it forms a set or cluster of features shared by the languages of this area. These common specific features can be both positive and negative, as contrasted to the presence vs. absence of such features in the languages outside the area.

Generally speaking, related languages, due to their origin from an ancestral idiom spoken in a concrete location, tend to cluster in close geographic areas, with the exceptions explained by migrations. But even if people migrate, various groups speaking related idioms, after some period, can meet in a new territory and again become neighbours. In both cases this can lead to the situation when the related languages form a secondary linguistic community, irrespective of their actual relationship. Cf., especially, the spread of the IE languages in Europe, where different distantly related idioms, after

---

6. Trubetzkoy (1975: 393–394).

7. Jakobson (1985: 102–104).

8. Cf. also Martinet’s (1949: 26) definition of a “phonological area” as an area “in which genetically different or divergent languages show similar phonological features”.

9. Aikhenvald and Dixon (2001: 12).

their split from the proto-language and a period of separate development, again came into contact on a new territory and formed secondary communities in new configurations (e.g., Baltic and Slavic, Celtic and Germanic, Greek, Albanian and Slavic, etc.).

A secondary linguistic community can be genetically homogeneous and include only related languages, as is, in general, the case in the Balkans, or be genetically heterogeneous, like the languages in India and, as I shall argue below, in the Caucasus. In the second case it is much easier to determine the areal nature of the shared linguistic features: if languages are known to be unrelated, explanations for their structural similarities may lie primarily in the contact. If the neighbouring languages are related, it is much more difficult to distinguish between the genetically inherited structural traits from those acquired via later contact and parallel development. This is exactly the situation within the Abkhazo-Adyghean, Daghestanian and Kartvelian-speaking areas. All languages inside these groups have developed in millennia-long contact. The case of the Balkans is easier to diagnose in this respect, as some languages forming the Sprachbund, although related, have their sister-languages outside this area, which facilitate the task of sorting out the contact-induced and genetically retained phenomena.

The situation is even more complex when we register similarities between neighbouring languages or families with uncertain genetic relations between them. In this case possible Sprachbund features, including lexical and phonological similarities, can create a wrong impression of family-like affinity, as happened, for example, in the history of the Caucasian linguistics. Thus, the authors of “The Structural Affinities Between the Caucasian Languages” (Klimov 1978), because of the unclear nature of relations between the three native Caucasian groupings, were hesitant whether to attribute the numerous structural affinities observed between them to (distant) kinship, or simply to universal properties and tendencies.

As to the types of convergence patterns between the languages or language families, there can be cases of unilateral convergence, when one language (family) develops in the direction of the other, bilateral convergence, when two languages or families influence each other, and multilateral convergence, when more than two languages or families converge, developing in parallel directions.

One can talk also of a centre and periphery of a linguistic union, whereby certain traits concentrate within the core language(s) of the area, but become rarer towards its periphery. This should not be understood in purely geographical terms: one and the same geographic area can contain languages which belong to the core of the union and those belonging to its periphery, while still others might not belong to this cluster at all. This often happens in “residual zones” (cf. Nichols 1992), for instance, in the Caucasus, when new languages are being added to the long-existing linguistic communities.

We can now turn to the central topic of this paper: what can be said of the indigenous languages of the Caucasus? Do they comprise a single genetic taxon, or simply a group of unrelated language families? And, related or not, do they form an areal unity? Intriguingly, every one of these options has its proponents and opponents. These are the issues which we will be dealing with in the following parts of this paper.

## 2. The languages spoken in the Caucasus

Only three genetic groupings are known to be indigenous and were spoken in the Caucasus area well before the appearance here of Indo-European, Turkic or Semitic languages: Abkhazo-Adyghean, or N(orth-)W(est) C(aucasian), spoken in the north-western Caucasus and western Transcaucasus; Nakh-Daghestanian, or N(orth-)E(ast) C(aucasian), spoken in the north-central and north-eastern Caucasus, and Kartvelian, or S(outh) C(aucasian), spoken in the southern Caucasus.

Apart from these indigenous Caucasian idioms, there are also Indo-European Ossetic, Armenian, Tat, Talysh, Kurdish and Russian and Turkic Azeri, Turkish, Karachay-Balkar, Kumyk and Nogay. Besides, one can add here also Indo-European Greek, and Semitic Neo-Aramaic (called in the Caucasus Aysor).

Table 1. Languages spoken in the Caucasus

Language family	Branch	Group	Language(s)
Kartvelian (or South Caucasian)		East Kartvelian	Georgian
		West Kartvelian	Megrelian, Laz
		Svan	Svan
	North-West Caucasian	Abkhaz	Abkhaz, Abaza
		Circassian	Adyghe, Kabardian
	North-East Caucasian	Ubykh	Ubykh
		Avar	Avar
		Andi	Andi, Akhvakh, Karata, Botlikh, Godoberi, Bagvala, Chamala, Tindi
		Tsez	Tsez, Khvarshi, Hinukh, Bezhta, Hunzib
		Lak	Lak
		Dargi	Dargi (with Kubachi, Megeb, etc.)
		Lezgi	Lezgi, Tabasaran, Aghul, Rutul, Tsakhur, Udi, Kryz, Budukh, Archi
		Khinalug	Khinalug
Indo- European	Iranian	Nakh	Chechen, Ingush, Bats
		North-Eastern	Ossetic
		North-Western	Talysh, Kurdish
		South-Western	Tat
	Armenian		Armenian
	Slavic	East Slavic	Russian
	Greek	Northern	Pontic Greek, Tsalka-Alaverdy (< Cappadocian)

(Continued)

Table 1. Continued

Language family	Branch	Group	Language(s)
Turkic		North-Western	Karachay-Balkar, Kumyk, Nogay
		(Kypchak)	
		South-Western	Azeri, Turkish
Afroasiatic	Semitic	(Oghuz)	
		West-Central	Neo-Aramaic (Aysor)

The antiquity of the autochthonous Caucasian languages in the areas they are spoken now is beyond any doubt. It is generally assumed that “Northwest Caucasian, North-east Caucasian and Kartvelian all apparently arose in or near their present territories” (Nichols 1992: 14). Another generally accepted view is that these languages have been in close contact with each other for a very long period of time. Though migrations and language change did take place in one or another direction, the main configurations of the three groups have been relatively stable over millennia, which is corroborated by linguistic, archaeological and anthropological data. For example, historians find it possible to connect the ancestors of the two North Caucasian groups with such early Bronze Age archaeological cultures as the dolmen and Maykop cultures in the western Caucasus and the Kuro-Arax culture in the southern Caucasus (second half – the end of the 3rd millennium BC; cf. Fedorov 1983: 28–31, 41–42).

3. Multilingualism in the Caucasus

Before the introduction of Russian, the Caucasus has never had a single dominating language or *lingua franca* as far as is known. The current nearly universal knowledge of Russian is a recent phenomenon, starting from the time the whole Caucasus was incorporated into the Russian Empire (around 1864). Before that, in many parts of the Caucasus it was multilingualism that served the communications needs of neighbouring heterolingual communities. Multilingualism was the norm in many Caucasian communities, though its scope was not the same in various parts of the Caucasus. Daghestan is a classical example of a multilingual polyglottal area. In other parts of the Caucasus multilingualism was not that widespread, and the population at large, especially in its core territory, was monolingual. Thus, in western Georgia, until the beginning of the 20th century, the level of knowledge of Georgian was rather low in Megrelia and Svanetia, and both Megrelians and Svans became actively bilingual (native+Georgian) only with the introduction by the Soviets of universal education, which in Megrelia and Svanetia was conducted in Georgian. Likewise, major parts of the mainland Adyghe, Kabardian, Abkhaz, Georgian, Azeri, and Armenian populations were traditionally monolingual, though in mixed communities and in borderlands bilingualism (Ubykh-Abkhaz,

Ubykh-Circassian, Abkhaz-Megrelian, Svan-Megrelian, Svan-Georgian, Megrelian-Georgian, Bats-Georgian, Ossetic-Georgian, Georgian-Armenian, Georgian-Turkish, Armenian-Azeri, Daghestanian-Azeri, etc.) was more common.

Several situations obtain:

1. Symmetrical bilingualism/multilingualism, when members of the neighbouring communities are competent both in their own and in their neighbour(s)'s language(s).
2. Unilateral bilingualism/multilingualism, when a community speaking a certain language is also competent in its neighbour(s)'s language(s), but not vice versa.
3. A third language competence, or a *lingua franca* situation, when the neighbouring heterolingual communities do not speak each other's languages but use instead a third language as a means of mutual communication.

Symmetrical bilingualism is more typical for the territories where two larger languages are being spoken; cf. in Daghestan bilateral Avar-Dargi, Avar-Lak, Dargi-Lak, Tabasaran-Lezgi bilingualism.

As to unilateral bilingualism, typically it is a minority group that is conversant in the language of its more numerous neighbours, not vice versa. For instance, the speakers of Archi, which belongs to the Lezgi group of Daghestanian, live in the Avar-populated Charodin district of Daghestan, which borders with the Lak area. Consequently, the Archis, who are a very small minority, speak, beside their native tongue, also Avar, Lak and Russian. The Megeb Dargis, who live in the Avar-populated Gunib district, speak Dargi, Avar, Lak and Russian (Madiëva 1991: 48). In the western Caucasus, many Abazas, a minority population in the Karachay-Cherkes Republic, speak, beside Abaza and Russian, also the distantly related Kabardian, and some also Turkic Karachay, the languages of their more numerous neighbours. The Ubykhs, who were numerically smaller than neighbouring Circassians and Abkhazians, were either bilingual with Circassian or Abkhaz as a second language, or tri-lingual Ubykh-Circassian-Abkhaz. In the Muslim Georgian area of Adzharia in the southern Caucasus adjacent to Turkey, Turkish has traditionally been the second language of the population, and even today many (older) Adzharians are fluent in Turkish; besides, nearly all Adzharians speak Russian. The tiny Laz and Abkhaz communities in Adzharia, beside their native tongues, speak also Georgian and Russian, and many also Turkish, and are thus quadrilingual.

As to the third communication pattern, in the Daghestan area, for instance, Avar was often used in the past by various Daghestanian communities as a *lingua franca*, and even a special form of Avar, called *bolmac̣* ('public language') had developed to facilitate communication. To some extent Turkic Kumyk was also used in some North Caucasian communities as a kind of *lingua franca*, a role now overwhelmingly taken over by Russian.

It has been noted that multilingualism in the Caucasus has a certain vertical dependency: the population of the mountains knew the language(s) of their heterolingual neighbours in the foothills and lowlands, but not vice versa.



In general, the Caucasus can be characterized, according to Johanna Nichols' (1992: 14, 16) definition, as a *residual zone*: "languages accumulate and survive in the Caucasus. . . Intrusive languages . . . do not replace other languages or families but are added to them. Thus the Caucasus tends to increase in genetic and typological diversity over time".

#### 4. The Caucasus as a socio-cultural setting

The Caucasus represents a geographic area with a variety of contrasting landscapes and climatic zones: from subtropical in Adzharia (in Georgia), Abkhazia and Lenkoran (in Azerbaijan) to perennial snows and glaciers in the mountains of the Great Caucasus Range; from the extremely fertile lowlands of the Kuban valley to the arid hilly landscapes of southern Georgia and Armenia. The traditional economies were based on agriculture, animal husbandry and hunting. Crafts were also a part of the native economy, famous centres being situated in Daghestan and parts of Azerbaijan. Baku (in Azerbaijan), Tiflis (in Georgia) and Derbent (in Daghestan) were important trade centres and cross-roads of cultures. The intense contact between neighbouring peoples contrasted with rather weak links between distant parts of the Caucasus.

Similar patriarchal feudal systems were typical for most parts of both the North and South Caucasus. Despite some differences in economic, cultural and geographic environment, the Caucasian peoples, irrespective of the languages they speak, are characterized by many common traits in social and cultural life, factors that facilitate their communication and strengthen their sense of belonging to a common Caucasian culture.

From the religious perspective, Georgia and Armenia are ancient Christian countries, while Azerbaijan is predominantly Shia Muslim. Abkhazia is nominally partly (Orthodox) Christian, partly (Sunni) Muslim, while Ossetia, apart from small Muslim enclaves, is mostly Orthodox Christian. The Iranian-speaking Tats or Mountain Jews (Daghestan, Azerbaijan and some other areas) profess either Judaism (the speakers of Northern Tat), Islam (in Azerbaijan) or Christianity (the Armenian Tats); the latter two groups are speakers of the southern dialect (Grjunberg 1997: 141–142). The rest of the traditional population groups in the Caucasus (Adyghe, Kabardians, Karachays, Balkars, Abazas, Chechens, Ingushs, Avars, Dargis, Tabasarans, Aghuls, Tsakhurs, etc.) are Sunni Muslims; in Chechnya, Ingushetia and parts of Daghestan various schools of the Sufi orders are also active.

Some Caucasian peoples, especially Abkhazians, Circassians, Ossetians, Svans and mountain Georgian groups still preserve remnants of traditional, i.e., pre-Christian and pre-Muslim, religious practices. In general, religion rarely functioned as an antagonizing factor in relations between the peoples of the Caucasus. In the cosmopolitan urban centres like Tiflis (Tbilisi), Baku or Derbent peoples of various ethnic backgrounds and confessions easily mixed and lived peacefully side by side.

## 5. The indigenous Caucasian languages: one, two, or three families?

A controversial issue in Caucasian linguistics is the genetic relationship between the three indigenous Caucasian groupings. One school of thought regards all three groups as undoubtedly related. Others speak in terms of the North Caucasian vs. Kartvelian division. The third group of specialists speaks of three independent Caucasian families, attributing the affinities between them to their long-term contacts.

The so-called Ibero-Caucasian<sup>10</sup> school of Caucasology regards the Caucasian linguistic material as too specific to be governed by systemic phonetic rules as observed in the other, for example, Indo-European languages; hence the rejection of the classical methods of comparative linguistics. The proponents of this school (of which the most prominent figures were A. Čikobava, K. Lomtadze, and G. Rogava) take the relationship of the three groups of the Caucasian languages for granted, not bothering much about providing solid proof for this alleged kinship. Another peculiarity of this school is the predominant interest in the relationship between the individual NC groups or even individual languages with Kartvelian, most prominently Georgian, without giving much thought to the demonstration of the relationship between the NWC and NEC groups, their kinship being regarded as self-evident. This is in obvious contradiction to the opinion expressed in 1922 by Trubetzkoy, that “the comparative grammar of the Caucasian languages – if all Caucasian languages do indeed form a single linguistic family – cannot be created, unless, on the one hand, a comparative grammar of Kartvelian, and on the other hand, the comparative grammar of North Caucasian languages have been created, and the creation of each of these grammars must be regarded as independent tasks.” And more definitely: “Unless similar correspondences [as those revealed between the IE languages – V.Ch.] are established between “Kartvelian” phonemes and phonemes of the North Caucasian languages, we have no right to speak about the Caucasian linguistic community, and any theory assuming this community as given, should be declared a fantasy.”<sup>11</sup>

The obvious failure of the Ibero-Caucasian school (which, on the other hand, scored important achievements in the description of many Caucasian languages and dialects) to prove convincingly the relationship between the three indigenous Caucasian groups can be seen as a primary source of profound skepticism on a part of Caucasologists, who took the view that even if all three groups are eventually derived from a common ancestor, this must have happened so long ago that nearly all traces of such a relationship

---

10. The part “Ibero-” means “Georgian/Kartvelian” and is derived from the ancient name of modern Eastern Georgia, Iberia (which has nothing to do with the Spanish Iberia). The second part, “Caucasian”, refers to the North-Caucasian (NWC and NEC) languages. The main purport of the Ibero-Caucasian school is thus the substantiation of the kinship between the Kartvelian and the (North) Caucasian languages.

11. Trubetzkoy (1922); cited from the Russian edition: Trubetzkoy (1987: 234).

must have long since disappeared, so that this relationship was, though probable, in fact indemonstrable.

Other authors reject, mostly on typological grounds, the possibility of genetic links between the three Caucasian groups, explaining similarities between them either by geographic proximity, or by universal tendencies.

The most impressive results, in my view, have been achieved by the linguists working in the direction of the North Caucasian relationship. Nikolai Trubetzkoy and George Dumézil laid the basis for the North Caucasian comparative linguistics, having established numerous correspondences between the NWC and NEC languages in phonetics, lexicon and morphology. These results convinced a number of specialists in the field of Caucasian linguistics of the existence of a North Caucasian family. Thus, G. Deeters (1931: 290) wrote that the relationship between the NWC and NEC languages was proven by Trubetzkoy, and that the SC languages do not seem to be related to this family. In another paper, Deeters (1955: 26) asserts, referring to the works by Trubetzkoy, that there are undoubted lexical similarities between the NWC and NEC groups. K.-H. Schmidt (1972: 25) wrote that the genetic relationship between the NEC and NWC languages, after the famous 1930 article by Trubetzkoy “Nordkaukasische Wortgleichungen”, must be regarded as proven. A similar idea was expressed somewhat more cautiously by the Dutch Caucasologist A. H. Kuipers (1963: 315): “The existence of a genetic relationship between NW and NE Cauc.[asian] is probable; the relations of S Cauc.[asian] to this N group so far remain unclear . . . This appraisal of the possible genetic relationships between the three groups is based on the number of reasonable etymologies that have been proposed, cf. especially N.S. Trubetzkoy, “Nordkaukasische Wortgleichungen.”

The Trubetzkoy/Dumézil tradition (cf. Trubetzkoy 1922; 1930; Dumézil 1933; 1937) was continued, after a long pause, by the late Kabardian scholar Auez Abdokov (cf. Abdokov 1976; 1981; 1983) and most notably by the Moscow linguists Sergei Nikolayev and Sergei Starostin (Nikolayev and Starostin 1994; cf. also Starostin 1999), who produced, in my view, compelling evidence for the NWC and NEC relationship.<sup>12</sup>

I shall present here just a few examples of lexical correspondences between the NWC and NEC languages from the basic vocabulary<sup>13</sup>, in comparison with forms reconstructed by Klimov (1964) for P(roto-)K(artvelian):

‘leg, foot’: NWC \*λ’a, PN \*lar, PAvar-Andi \*łoli, PTsez \*łolé, PLez \*lāl; cf. PK \*bark’-(al), \*berq-, \*k’warc<sub>1</sub>χl’-;

12. Among the proponents of the idea of the North Caucasian linguistic unity are such scholars as M. Kumaxov, S. Kodzasov, M. Alekseev, V.V. Ivanov, A. Šagirov, Y. Testelec, et al.

13. NEC proto-forms are from Nikolayev and Starostin (1994), NWC are mine (cf. Chirikba 1996), which in many cases coincide with the NWC proto-forms elaborated in Nikolayev and Starostin (1994).

- ‘walk, go’: NWC *\*k’<sup>o</sup>a*, PN *\*-k’-* (‘lead, drive’), PAvar-Andi *\*k’Vb-//\*-ik’<sup>o</sup>(Vn)-* (‘drive’, ‘urge, direct, (re)turn’), PTsez *\*ʔek’-*, PDar *\*-irk’-* (‘drive’, ‘urge’), PLez *\*ʔak’i-* (‘walk, go’, ‘drive’, ‘urge’); cf. PK *\*gwal-*, *\*wid-*;
- ‘road, path’: NWC *\*məʔ<sup>o</sup>a*, PN *\*niq*, PAvar-Andi *\*miq*<sup>i</sup>, PDar *\*daq*, PLez *\*rāq*; cf. PK *\*gza-*
- ‘tooth’: NWC *\*ca*, PN *\*ca*, PAvar-Andi *\*colu*, PTsez *\*sīl*, PDar *\*cula*, PLez *\*sīl*, Khin *\*culoz*; cf. PK *\*k’b-il-*;
- ‘tongue’: NWC *\*bz/za*, PN *\*mot’t*, PAvar-Andi *\*mič<sup>i</sup>*, PTsez *\*mic*, Lak *maz*, PDar *\*meč*, PLez *\*melč*; cf. PK *\*nena-*;
- ‘house’: NWC *\*ʔ<sup>o</sup>ana*, PN *\*phē* (‘village’), PTsez *\*q<sup>o</sup>in* (‘farmstead’), PTsez-Khvar *\*q<sup>o</sup>ən* (‘village’), PHunz-Bezh *\*qun* (‘village’), PLez *\*muq*<sup>(a)</sup> (‘village’, ‘farmstead’); cf. PK *\*(s)a-xl-*;
- ‘name’: NWC *\*(p)c’a*, PN *\*c’a*, PAvar-Andi *\*c<sup>o</sup>iri*, PTsez *\*c<sup>o</sup>ā*, Lak *c’a*, PDar *z<sup>o</sup>e*, PLez *\*c<sup>o</sup>er*, Khin *\*c’u*; cf. PK *\*ʒax<sub>1</sub>e-*;
- ‘rat, mouse’: NWC *\*q<sup>o</sup>ənə//H<sup>o</sup>ənə*, PAvar-Andi *\*hink<sup>o</sup>o* (Avar *ʔunk’*), PTsez *\*ʔāq<sup>o</sup>V*, Lak *uk’u-lu*, PLez *\*nVq<sup>o</sup>e-l*, Khin *nuk’ur*; cf. PK *\*(s<sub>1</sub>)tagw-* ‘mouse’, Geo *vir-txa* ‘rat’;
- ‘meat, flesh’: NWC *\*L’ə*, PN *\*dilxu*, PAvar-Andi *\*riḫ<sup>i</sup>*, PTsez *\*riḫ*, Lak *dik*, PDar *\*diḡ*, PLez *\*jaḫ*, Khin *lika*; cf. Geo *\*qorc-*;
- ‘grass’: NWC *\*wəxə*, PN *\*būc*, PAvar-Andi *\*bici*, PLez *\*wVcVn* (‘burdock’); cf. PK *\*tib-*;
- ‘soil, clay’: NWC *\*nəs<sup>o</sup>ə*, PAvar-Andi *\*ʔonš<sup>o</sup>i*, Lak *arši*, PDar *\*heš/š*; cf. PK *\*tiqa-*;
- ‘louse’: NWC *\*c’a*, PN *\*macē*, PAvar-Andi *\*noč<sup>i</sup>*, PTsez *\*nəcə*, Lak *nac*, PDar *\*nez*, PLez *\*nāč*, Khin *nimc*; cf. PK *\*t’iz<sub>1</sub>-*;
- ‘moon’: NWC *\*məʒa*, PN *\*butt*, PAvar-Andi *\*birč<sup>o</sup>i* (Avar *moč*), PTsez *\*bocV*, Lak *\*barz*, PDar *\*bač*, PLez *\*wač*, Khin *\*wac*; cf. PK *\*du(s<sub>1</sub>)te-*;
- ‘star’: NWC *\*c<sup>o</sup>a*, PN *\*t’fari*, PAvar-Andi *\*c<sup>o</sup>arhi*, PTsez *\*c<sup>o</sup>ā*, Lak *c’u-ku*, PDar *zuri*; cf. Geo *varskvlav-*;
- ‘ice’: NWC *\*məLə*, Lak *mik*, PDar *\*miḡ*, PLez *\*merḫ<sup>o</sup>*, Khin *mik*; cf. Geo *q’inul-*;
- ‘ashes’: NWC *\*tq<sup>o</sup>a* (also ‘grey’), PAvar-Andi *\*rVqV*, Lak *lax* (also ‘dust’), PLez *laḫ* (‘grain peelings’); cf. PK *\*t’ut’a*;
- ‘new’: NWC *\*c’a*, PN *\*c’in*, PAvar-Andi *\*c’in<sup>h</sup>V*, PTsez *\*-ic’Vn-*, Lak *c’u-*, PDar *\*c’i-*, Lez *\*c’enjä-*; cf. Geo *axal-*;
- ‘day’: NWC *\*məḷ<sup>o</sup>a*, PN *\*mälx* (‘sun’), PAvar-Andi *\*mili* (‘day, sun’); cf. PK *\*dye-*;
- ‘year’: NWC *\*s<sup>o</sup>a/ə*, PAvar-Andi *\*rišin*, Lak *šin*, PLez *\*sān*; cf. PK *\*za-*, *\*cēl-*;
- ‘I’: NWC *\*sa*, PN *\*sō*, PLez *\*zo-n*, Khin *zə*; cf. PK *\*me(n)-*;
- ‘thou’: NWC *\*wa*, PN *\*waj*, PAvar-Andi *\*mi-n* (< *\*wi-n*, by assimilation), PTsez *\*mə* (< *\*wə-n*), PLez *\*uo-n*, Khin *wə*; cf. PK *\*sen-*;
- ‘you (pl.)’: NWC *\*s<sup>o</sup>a*, PN *\*šu*, PAvar-Andi *\*b-iš-//\*ʔuš-*, PTsez *\*mižə*, Lak *zu*, PDar *\*nu-ša*, PLez *\*ʒ<sup>o</sup>[e]-n*, Khin *zu-r* (oblique base *su-r*); cf. PK *\*(s<sub>1</sub>)tkwen-*;

- ‘one’: NWC \**za*, PN \**cha*, PAvar \**ci*-, PTsez \**hə̌š*, Lak *ca*, PDar \**ca*, PLez \**ṣa*, Khin *sa*; cf. PK \**ert*-;
- ‘two’: NWC \**dG°ə*, PAvar-Andi \**k’i*-, PTsez \**q’i-nV*, Lak *k’i-a*, PDar \**k’i*, PLez \**q’ä*, Khin *k’u*; cf. PK \**jor*-;
- ‘three’: NWC \**χə*, PAvar-Andi \**χob*-, PTsez \**χ̌*, PDar \**hab*-, PLez \**λep̄i*-; cf. PK \**sam*-;
- ‘four’: NWC \**p’χ’ə*, corresponds to NEC ‘eight’:<sup>14</sup> PN \**barλ*, PAvar-Andi \**biχ̌i*-, PTsez \**beχ̌-(ň)*, PLez \**menχ̌ä*-; cf. PK \**o(s<sub>l</sub>)tx(w)*- ‘four’, \**arwa*-‘eight’.

Most importantly, systematic comparison of the vocabulary of the NWC and NEC languages (or rather, of their reconstructed proto-languages) reveals patterns of regular sound correspondences (cf. Trubetzkoy 1930; Abdokov 1983: 46–72; Nikolayev and Starostin 1994: 40–91). This is contrasted with the failure to establish any comparable system of regular phonetic correspondences between the NC groups and Kartvelian, as the history of the Ibero-Caucasian hypothesis testifies. A thorough analysis of the lexical material of the NWC, NEC and SC languages will lead any competent historical linguist to the conclusion that we deal here with two mutually unrelated linguistic families: North Caucasian (with its Western and Eastern branches) and Kartvelian.

If the NC and SC languages represent genetically unrelated taxons, to what factors should we attribute then the existence of numerous common traits observed between these two families? It would be only logical to seek the sources of these similarities in language contact.

## 6. Polyglottal Caucasus: a geographic notion, a Sprachbund, or what?

The second major controversy around the Caucasian languages, beside the problem of their genetic relations, is whether these languages, related or not, form a language union. Some authors do not doubt the existence of a “common Caucasian linguistic type” and regard its existence as self-evident. Others dispute and sometimes even categorically reject the existence of any pan-Caucasian similarities, except for the sole phonological trait of glottalization.

As noted by Klimov (1986a: 172), the most positive views on the existence of a common Caucasian type were held by “outsiders”, i.e., not specialists in the Caucasus in *sensu strictu*. Thus, Baudouin de Courtenay emphasized the existence of common

14. Cf. Nikolayev and Starostin (1994: 315), Chirikba (1996: 406)

features between the Caucasian languages “without any regard to the presence or absence of their ancient genetic kinship”.<sup>15</sup> Ernst Lewy (1961: 597) asserted that the Caucasian languages do form a unit, even in case they are not genetically related. Václav Polák (1950: 400) went even further, claiming the existence of a Caucasian language union: “Les langues du Caucase, caucasiennes d’origine ou non, ont pris le type “caucasien” dû – me semble-t-il – à la pression des modèles typologiques de l’union de langues caractéristique pour le territoire en question. C’est pourquoi il n’est pas nécessaire – me semble-t-il – de reconstruire la langue commune caucasienne pour expliquer certains traits communs. Il suffit d’y voir des phénomènes dûs à la pression du modèle structurel de l’union des langues en question”. V. Pisani also asserted that “here two or three language groups, absolutely different genetically, became so close between themselves that one can speak about ‘Caucasian languages’ as languages possessing many common elements, which give them the appearance of a language family”.<sup>16</sup> In another work Pisani suggested that in the Caucasus we can observe a language union in the process of formation (“Sprachbund im Werden”) (Pisani 1959: 85). A somewhat different approach was taken by the Georgian linguist G. Cereteli (1968: 14), who proposed the term “allogenic relationship”, which unites separate languages of the Caucasus, or even groups of these languages, on the basis of certain affinities between them, which can hardly be explained by a simple borrowing.

The notion of a Caucasian linguistic area or of a common Caucasian language type has firmly penetrated the non-Caucasological literature and became, as Tuite (1999: 1) notes, a commonplace.<sup>17</sup> Ironically, Caucasologists themselves are divided on this important issue. As Klimov (1986b: 129) noted on that account, “Paradoxical as it may seem, the Caucasian area studies have so far failed to ascertain a “Sprachbund” within the Caucasus among the autochthonic and Indo-European (Armenian and Ossetic) languages. While to outsiders its existence is beyond doubt . . . , to Caucasologists it is at best a hypothetical assumption”. Others were quite categorical in rejecting any idea of structural unity of the Caucasian languages. Thus, the German Caucasologist Gerhard Deeters in his 1931 article remarked: “In general, the structures of the three groups are so different, that there can be no question of a common “Caucasian” language type”.<sup>18</sup> A similar view was expressed recently by Tuite (1999: 5): “When it comes to the typology of NWC, NEC and SC, linguists who know these languages well find genuinely

---

15. Baudouin de Courtenay (1963: 112); cited from Klimov (1986: 172).

16. Pisani (1956: 54).

17. Cf. for example, Emeneau (1980: 1): “long-recognized linguistic areas such as the languages of the Caucasus or of the Balkans”. Cf. also Bloomfield (1933: 468–471).

18. Deeters (1931: 290); but in his later paper Deeters (1957: 13) does speak of the possibility of the Caucasian Sprachbund.

pan-Caucasian traits hard to come by. There is certainly nothing comparable to what can be described for the Balkans or other well-established Sprachbünde: no pan-Caucasian patterns of clause linkage, nominal categories (such as definiteness), or verbal categories.”

But other specialists in the Caucasus do assume the existence of certain pan-Caucasian traits. Thus, Mačavariani (1966: 8) noted the presence of numerous and deep structural-typological similarities between the Kartvelian and the North Caucasian languages. In his survey on the languages of the Caucasus, J. Catford (1977) produced a table, which lists relevant phonological, morphological and syntactic features common to all three groups. Although Catford concludes that there are not many traits common to all Caucasian languages, he remarks that “anyone working with these languages receives a strong impression of ‘family likeness’ running through all of them.”

Such deep disagreement between the specialists on the crucial issue of the areal relationships of the Caucasian languages can only be compared with an even greater discord on the issue of the genetic relationship between these languages. The reason for such polarity in both cases is no doubt the insufficient level of research. Besides the volume edited by Klimov in 1978, a chapter in Klimov’s survey of the Caucasian languages (1986a) and a number of articles on the theme, very little work has actually been done on the issue of the Caucasian areal linguistics in a satisfactory scope and depth. This reflects, as Klimov himself emphasized, the Caucasologists’ preoccupation mainly with genetic and grammatical issues, not areal ones. Admitting that “the areal connections of the autochthonous languages of the Caucasus . . . remain to this day poorly worked out”, Klimov (1986b: 123) provides insights into why the Sprachbund idea did not find its supporters among the specialists in the Caucasus: “This situation seems, nevertheless, natural, since the processes of convergence in the Caucasus have been little investigated because of the bias of part of the specialists against the “Sprachbund” notion itself.” In another paper Klimov (1991: 8) even notes that some Caucasologists regarded the Sprachbund theory as “anti-historical”.

Against this background, the recent article by Tuite (1999) is a welcome exception, although it is devoted primarily to only one aspect of the problem, namely, the nature and the functioning of the ergative construction in the three Caucasian groups. The main aim of this paper was to knock out the second major stone (alongside glottalization) at the basis of, in Tuite’s words, the “myth” of the Caucasian linguistic union. As he writes on p. 23 of his paper, “Whereas the pan-Caucasian distribution of glottalization is doubtless due to local diffusion, the (nearly) pan-Caucasian distribution of ergativity must have some other explanation, one that may go far back into the past, and which must be explored separately in each Caucasian language family . . . If this is so, there remains little to link the Caucasus together as a linguistic area save a single phonetic feature (glottalization), and the general impression we outsiders have that it is somehow exotic and different”.

The publication in 1978 of an important volume (Klimov 1978) specifically devoted to the discussion of the structural similarities between the three Caucasian groups was



a major contribution to the topic. Noting “a *considerable number* (my italics. – V.Ch.) of common Caucasian structural parallels found in lexicon, syntax, morphology and phonetics” (p. 127), the authors of this book were, however, quite cautious not to attribute these similarities to areal diffusion, preferring to interpret them by general typological factors or, alternatively, as the probable evidence for their ancient genetic links. The general conclusion of the book quite eloquently highlights this deliberately indecisive approach: “the modern state of Caucasian linguistics enables one to say that the majority of the listed parallels are due to a typological stimulus”, and “it is very doubtful that the observed common Caucasian parallels are due to any sort of areal interaction, either to ancient interaction of the autochthonous Caucasian languages or the affect of the common substrate, or any others. Indeed, in the vast majority of cases the origin of common Caucasian structural parallels cannot be interpreted by linguistic contacts” (p. 127, 130). Besides: “It is not improbable that quite a number of the Caucasian isoglosses can be accounted for genetically as the original legacy of a common ancestor” (p. 128–129).

One is thus confronted with a dilemma: either to regard the “numerous” common Caucasian structural traits as explainable by general typological factors or, alternatively, as not random and explainable by genetic retention. Yet, despite the hesitant approach of its authors, this monograph did succeed in demonstrating the major pan-Caucasian structural traits, which distinguish the Caucasian linguistic area from the adjacent Indo-European, Turkic, or Semitic areas.

The late Georgy Klimov, the celebrated Russian specialist in the Caucasian languages, too was cautious to call on areal factors for explaining inter-Caucasian similarities, hesitating between purely typological explanations and the possibility of very ancient genetic ties. Despite his generally negative attitude to the possibility of the areal nature of certain pan-Caucasian traits, he nevertheless noted numerous borrowings and calques on various levels of linguistic structure between the three Caucasian groups, and on the lexical level a noticeable layer of pan-Caucasian cultural isoglosses. He also noted that certain facts might be explained by substrate or by language replacement (Klimov 1991: 8).

Another authority in Caucasian linguistics, the Norwegian scholar Hans Vogt (1988: 502, 504), in contrast, did not show any prejudice to the idea of the Caucasian areal union. He wrote: “Whatever their genetic historical links, such a situation [long-term close neighbourhood of various Caucasian languages, regular contacts and multilingualism. – V.Ch] creates ideal conditions for mutual linguistic diffusion, . . . for the spread of innovations, for borrowings, in short, for convergence and parallelism in the development of these languages, for the creation of the language union (Sprachbund). The influence of the neighbouring languages is amazing even in the case of such relatively recent “newcomer” as the Ossetic language (. . .) Analyzing isoglosses which we meet in the Caucasian area, we would discover that a certain number of them is found beyond the borders of the Caucasus and link the Caucasian languages with the neighbouring non-Caucasian idioms. As is well-known, Ossetic and some Armenian



dialects, spoken in Georgia, have many Caucasian features both in morphology and syntax". Some other specialists in the Caucasus also regard the existence of the Caucasian linguistic area, irrespective of genetic relationships between the three groups, as probable or even self-evident. Cf., beside Vogt, also Deeters (1957: 13), despite his earlier negative view on that account (in his 1931 paper), and Čikobava (1970: 52), who regarded the Kartvelian languages as belonging to "one typological class" with the North Caucasian languages. Shimomiya (1978: 202, 209) speaks directly about the Caucasian Sprachbund as an obvious fact; cf. also Gabunia and Guzman Tirado (2002: 80–82).

It is interesting to analyze the position of Trubetzkoy, the instigator of the concept of the language union. As remarked by Tuite (1999: 3), "As both one of the premier Caucasologists of his day and inventor of the term Sprachbund, Trubetzkoy, of all people, ought to know whether the Caucasus qualified as a Sprachbund." From this Tuite concludes that "the fact that he never applied the term he invented to a region to which he had devoted years of study and fieldwork implies very strongly that he did not believe the Caucasus constituted a Sprachbund". Indeed, Trubetzkoy, the author of the term "Sprachbund", never spoke of the Caucasian language union, though he did mention the "Caucasian phonological union" in a letter dated 20 May 1937 to Roman Jakobson.<sup>19</sup> But at the time when Trubetzkoy was working on the NC languages, comparative Caucasology was at its initial stages, and it was not known what kind of relations – genetic or only areal – the NC family, whose existence was postulated by Trubetzkoy, had with the SC family. Trubetzkoy regarded even posing the question on the nature of relations between the NC and SC languages as premature, until the comparative grammars of both families had been independently created, and referred to his 1922 article as a *first step* towards a NC comparative grammar, expecting parallel research to be undertaken on the side of the Kartvelian linguistics. Being not in a position to ascertain the nature of the relations and common traits between the NC and SC languages, Trubetzkoy understandably chose the Balkan languages, with clear genetic affiliations, as a more obvious example of a Sprachbund.

## 7. The Caucasus as a linguistic area

### 7.1 The diagnostic features

There have been several attempts to make an inventory of features common for all three groups of the Caucasian languages (cf. Lewy 1961: 596–597; Charachidzé 1967; Catford 1977; Klimov 1986a: 172–173). In defining the linguistic area, it is useful to range the isoglosses from the point of view of non-trivial typological features. The following table gives a summary of such important ("weighty") traits which can be regarded as defining the Caucasian Sprachbund.

---

19. Cf. Trubetzkoy (1975: 393–394); cit. from Tuite (1999: 3).

Table 2. The diagnostic features of the Caucasian Sprachbund

Phonology	Morphology	Syntax	Lexical semantics	Lexicon
rich consonantism	agglutination polysynthetism predominance of prefixal conjugation	identical word order (SOV, Attr-N)	stative vs. dynamic verbs	common cultural terms not found outside the Caucasus
ternary contrast of stops and affricates		ergative construction	inversive verbs	common phraseology specific to the area
glottalization		inversive construction	ambitransitive (labile) verbs	common semantic patterns
rich sibilant systems	predominance of postpositional constructions	the possessor constituent precedes the possessed one	suppletive verbs for singular and plural arguments	
rich postvelar (uvular, pharyngeal and laryngeal) systems	masdar (verbal noun) morphological marking of causative			
similarly built harmonic clusters				
presence of schwa	category of evidentiality			
lack of phonemic diphthongs	category of potential			
lack of vocalic clusters	attachment of coordination markers to each conjunct			
ablaut	directional and orientational preverbs group inflection a three-grade deictic distinction vigesimal numeral system			

Though some of these traits can be explained by universal tendencies and independent parallel development, what is important here is that there exists a *set* of common features found on various levels of linguistic structure which is specific to the Caucasus and which renders this linguistic area different from any other linguistic community outside the Caucasus.

### 7.2 Phonology

Phonological systems of all Caucasian languages are characterized by substantial similarities which allow one to speak of a common Caucasian phonetic type (cf. SO 85; Catford 1977) or of “the Caucasian phonological union” (Trubetzkoy 1975: 393–4). Beside the existence of shared positive traits, the Caucasian languages share important negative features as well. Thus, for all their richness and diversity, all Caucasian systems lack velar or uvular nasals, interdental fricatives, glottalized resonants and voiced aspirated consonants. A general impression of the expanded consonant systems in the Caucasus can give the example of the Ubykh consonantism.

Table 3. A chart of Ubykh consonants

	Stops			Affricates			Fricatives		Resonants	Glides
	vd	vls	glot <sup>20</sup>	vd	vls	glot	vd	vls		
Labial	<i>b</i>	<i>p</i>	<i>p'</i>					<i>f</i>	<i>m</i>	<i>w</i>
pharyngealized	<i><u>b</u></i>	<i><u>p</u></i>	<i><u>p'</u></i>				<i><u>v</u></i>		<i><u>m</u></i>	<i><u>w</u></i>
Dental	<i>d</i>	<i>t</i>	<i>t'</i>	<i>ʒ</i>	<i>c</i>	<i>c'</i>	<i>z</i>	<i>s</i>	<i>n</i> <i>r</i>	
labialized	<i>d<sup>o</sup></i>	<i>t<sup>o</sup></i>	<i>t<sup>oo</sup></i>							
Middle				<i>ʃ</i>	<i>č</i>	<i>č'</i>	<i>ʒ</i>	<i>ś</i>		
labialized				<i>ʃ<sup>o</sup></i>	<i>č<sup>o</sup></i>	<i>č<sup>oo</sup></i>	<i>ʒ<sup>o</sup></i>	<i>ś<sup>o</sup></i>		
Back				<i>ʒ̣</i>	<i>č̣</i>	<i>č̣'</i>	<i>ʒ̣</i>	<i>ṣ́</i>		
labialized							<i>ʒ̣<sup>o</sup></i>	<i>ṣ́<sup>o</sup></i>		
palatalized				<i>ʒ̣'</i>	<i>č̣'</i>	<i>č̣'</i>	<i>ʒ̣'</i>	<i>ṣ́'</i>		
Palatal										<i>j</i>
Lateral						<i>ʎ'</i>	<i>L</i>	<i>λ</i>		
Velar							<i>ġ</i>	<i>ġ̣</i>		
labialized	<i>g<sup>o</sup></i>	<i>k<sup>o</sup></i>	<i>k<sup>oo</sup></i>							
palatalized	<i>g'</i>	<i>k'</i>	<i>k''</i>							
Uvular		<i>q</i>	<i>q'</i>				<i>ɣ</i>	<i>x</i>		
labialized		<i>q<sup>o</sup></i>	<i>q<sup>oo</sup></i>				<i>ɣ<sup>o</sup></i>	<i>x<sup>o</sup></i>		
palatalized		<i>q'</i>	<i>q''</i>				<i>ɣ'</i>	<i>x'</i>		
Pharyngealized		<i>q̣</i>	<i>q̣'</i>				<i>ɣ̣</i>	<i>x̣</i>		
uvulars										
labialized		<i>q<sup>o</sup></i>	<i>q<sup>oo</sup></i>				<i>ɣ<sup>o</sup></i>	<i>x<sup>o</sup></i>		
Laryngeal										<i>h</i>

20. vd – voiced, vls – voiceless, glot – glottalized.

### 7.2.1 *Consonant-type languages*

Syntagmatically, all (indigenous) Caucasian idioms can be called “consonant-type languages”, with more consonants in a speech sequence than vowels (cf. SO 104). The same term (“consonantal languages”) can be applied to them paradigmatically as well, all Caucasian languages being notorious for the richness of their consonantal inventories, versus restricted or very restricted vowel systems. The most moderate consonant system is presented in Kartvelian – ca. 30 consonants. Some other Caucasian languages display much richer inventories, which belong to the richest systems of the world, cf. Archi with its 69 consonants, Ubykh with its 80 or 81 consonants and Sadz Abkhaz with its 110 or so consonants.

The Kartvelian consonantism, modest by Caucasian standards, can be regarded as representing the core of any Caucasian system, to which additional features, such as labialization, palatalization, strength, pharyngealization, etc. have been added. Remarkably, the same core system is attested also in the non-Caucasian languages of the area, Armenian (in the dialects close to the Caucasus) and Ossetic. Historically, the relative “simplicity” of the Kartvelian consonant system is secondary, as compared to the reconstructed Proto-Kartvelian inventory. In contrast, Iranian Ossetic obviously complicated its phonemic system by acquiring an additional series of “Caucasian” glottalized consonants.

### 7.2.2 *The structural organization of the obstruent system*

The Caucasian obstruent system is represented by one model, which can be described in the words of J.C. Catford (1977: 288): “(i) *stops* articulated at *labial, dentalveolar, velar*, and *uvular* locations (types *p t k q*); (ii) *affricates* at two locations (types *ts tʃ*); *fricatives* at *alveolar, postalveolar*, and *uvular* locations (types *s ʃ x*).” In all languages there is a sharp dichotomy between obstruents and resonants. According to Kodzasov (SO 115), though the dental, alveolar and uvular stops and fricatives are attested in many languages outside the Caucasus, the combination of both stops and fricatives of all three of these series, as attested in the Caucasus, is extremely rare.

Some non-Caucasian languages of the area share with the neighbouring Caucasian languages not only phonological but phonetic characteristics as well. Thus, both in Caucasian languages and in Ossetic and Azeri, voiced obstruents are characterized by incomplete voicedness and a relative tenseness of articulation. Similar is also the articulation of voiceless aspirated stops in Ossetic and in Caucasian languages (cf. Melikišvili 1983: 221).

Whereas the Ossetic obstruent system shows undoubted Caucasian features, Armenian displays a more complicated picture. Classical Armenian had a ternary system, with voiced, voiceless aspirated and voiceless unaspirated members (e.g., *b - p - pʰ*). Those East Armenian dialects spoken in Georgia or in its near vicinity (Tiflis, Artvin, Kars, etc.) have the ejective series as the second member of the row (i.e., *b - p' - pʰ*). The glottalized obstruents are also found in some forms of West Armenian (cf. Pisowicz 1997: 217–219), whereas other western dialects have a binary obstruent system, resembling that of Turkish.

### 7.2.3 *The ternary opposition of laryngeal features*

Nearly all Caucasian languages have a ternary opposition of laryngeal features, including voiced, voiceless aspirated and glottalized correlates, instead of a binary system (voiced vs. voiceless), more widespread in the other areas of the world, including the contiguous areas. In all neighbouring non-Caucasian languages where such a ternary system occurs (in dialects of Armenian, Kumyk, Azeri, in Ossetic), it is nearly universally explained by the Caucasian influence or substrate (with some reservations for Armenian, see below).

Some (North) Caucasian languages expanded the basic ternary system by adding an additional contrast in strength, realized phonetically either as lax vs. tense, geminated vs. non-geminated or aspirated vs. non-aspirated consonants; such contrasts exist in the NWC Bzhadugh and Shapsygh Adyghe, and in NEC – in Bats and in many Daghestanian languages. A similar quaternary system of stops is attested in Ossetic dialects as well, assumed to be due to influence from neighbouring Caucasian languages.

### 7.2.4 *Glottalized consonants*

The presence of glottalized obstruents in all Caucasian and in some non-Caucasian languages, such as Indo-European Armenian and Ossetic, the dialects of Turkic Kumyk, Azeri, Karachay-Balkar and Turkish, and in Afroasiatic Neo-Aramaic dialects (which have developed glottalized consonants from emphatics) renders it an incontestable Caucasian areal feature. The glottalization is ejective and involves predominantly (voiceless) stops and affricates. The glottalized fricatives in Circassian, Abkhaz dialects, and in the Andi languages are rare exceptions and result from later developments. There are no glottalized resonants, which also renders it a typically Caucasian feature, as such sounds are not uncommon, for example, in North American indigenous languages. Several factors conspire here: a) the presence of glottalization in all Caucasian idioms and their diffusion into a number of non-Caucasian (Indo-European, Turkic or Afroasiatic) languages of the area; b) the nature of glottalization and the distributional limitations; c) the near complete absence of ejective glottalization elsewhere in Eurasia.<sup>21</sup>

Of the non-Caucasian languages, glottalization in the Caucasus is attested in Ossetic, the East Armenian dialects (e.g., Tiflis, Artvin), in dialects of Kumyk (e.g., Kaytag), northern Azeri (e.g., Zakatala-Kakh) and Karachay-Balkar (e.g., Malkar). Ossetic has glottalization not only in words borrowed from the Caucasian languages, but also in the native IE vocabulary (e.g., *st'aly* 'star'). The same is typical for the Malkar dialect of Karachay-Balkar (e.g., *k'ordum* 'I saw'). In the Kumyk dialects glottalization is explained by the NEC substrate.<sup>22</sup> The glottalized consonants are

21. The supposed presence of glottalization in both Afroasiatic and Indo-European reconstructed systems (though Indo-European is disputed) could in principle suggest a very early (prehistoric) "phonological union" between these latter and the two Caucasian families.

22. The population of several Kumyk villages, whose ancestors are known to be originally Avar speakers, represents a separate Kumyk dialect, displaying such non-Turkic traits as the presence of glottalized consonants, the violation of the vowel harmony rules, etc. (cf. Mikailov 1954: 12–13).

present in words of NEC origin in some northern Azeri dialects, e.g., in Zakatala-Kakh, which is based on the Tsakhur substrate, in the Tabasaran sub-dialects of Azeri, based on Tabasaran, etc. Armenian shows variegated systems, some of its dialects having glottalized consonants, some (including Classical Armenian) having a ternary opposition of voiced vs. voiceless aspirated and voiceless unaspirated. Some authors reject the idea of the areal origin of Armenian glottalics, tracing them back to Common Armenian and further back to the Proto-Indo-European glottalized stops (cf. Kortlandt 1978: 13, 15; Gamkrelidze and Ivanov 1984: 41), while others (cf. Pisowicz 1997: 217–219) regard them as the result of a secondary development. According to G. Cereteli, “It is wrong to think that the glottalized consonants are present only in those Armenian dialects, which are under the immediate influence of Georgian (e.g., Tiflis, Artvin, etc.). They are equally typical for all East Armenian, including the Literary Armenian language” (cf. Gamkrelidze and Mačavariani 1965: 46, fn 2).

The fact remains that phonological glottalization is found exclusively in those Indo-European languages spoken in the Caucasus, which is why their attestation here is attributed by many authors to the influence of the Caucasian languages (cf. for instance Vogt 1988: 458). It is not quite clear whether some Caucasoid substrate language was responsible for the evolution of emphatics into glottalized stops and affricates ( $p'$ ,  $t'$ ,  $k'$ ,  $č'$ ) in the Eastern Neo-Aramaic dialects of Urmia, Van and Mosul (cf. Cereteli 1976: 229), geographically close to the Caucasus.

#### 7.2.5 *Rich sibilant systems*

All Caucasian languages are characterized by extreme richness of their sibilant<sup>23</sup> systems, which can be regarded, along with glottalization, both as a pan-Caucasian trait and as one of the main distinctive features of the Caucasian phonological type. This is contrasted with a relative paucity of sibilants in other, especially contiguous, areas.

**Table 4a.** The number of sibilants in the Caucasian languages

30+	Sadz Abkhaz
29	Bzhadugh Adyghe
27	Bzyp Abkhaz, Ubykh
20	Abzhywa Abkhaz
18	Tabasaran, Burshag Aghul
16	Andi
15	Proto-Kartvelian, Akhvakh, Botlikh
14	Avar
13	Tsakhur
12	Lak, Chamala, Archi

(Continued)

23. By “sibilants” I mean front ([s] and [ts]-type), middle/hissing-hushing ([ç] and [tç]-type) and back ([ʃ] and [tʃ]-type) fricatives and affricates.

Table 4a. Continued

11	Chechen, Tindi, Dargi, Fite Aghul, Khinalug, Bagvala
10	Lezgi, Rutul, Kryz, Udi, Ingush, Bats, Georgian, Megrelian, Svan
9	Richa Aghul, Budukh
8	Inkhokvari Khvarshi, Tsez, Hinukh, Hunzib

The situation with sibilants in the non-Caucasian languages of the area and in the contiguous languages is as follows:

Table 4b. The number of sibilants in the non-Caucasian and contiguous languages

Areal non-Caucasian languages		Contiguous languages	
10	East Armenian	7	Kurdish
8	West Armenian, Ossetic	6	Talysh, Turkish, Persian
7	Kumyk, Karachay-Balkar	5	Neo-Aramaic
6	Nogay, Azeri, Tat	4	Greek

Though some areal and contiguous languages also display a developed system of 6-7 sibilants, the sibilant systems of Ossetic and especially Armenian are obviously richer, which may well be attributed to the areal factor.

Some non-contiguous Indo-European and Uralic languages of Europe also boast developed sibilant systems, though other Eurasian systems are more moderate and in the majority of them the number of sibilants vacillates between 6 and 4, while the smallest systems contain 2 or even 1 sibilant:

Table 5. The number of sibilants in the languages of Eurasia

16	Lithuanian (IE/Baltic)
14	Wakhi (IE/Iranian)
12	Burushaski (isolate)
11	Ishkashim (IE/Iranian), Albanian (IE/Albanian)
10	Rushan (IE/Iranian), Russian (IE/Slavic), Permiak, Komi-Zyrian (Uralic)
9	Karelian and Mordvin (Uralic)
8	Arabic (Afroasiatic), East Baluchi and Roshorv (IE/Iranian), Latvian (IE/Baltic), English (IE/Germanic), Udmurt, Veps and Hungarian (Uralic), Karaim (Turkic), Mongolian (Mongolic)
7	Karakalpak (Turkic), Moldovan (IE/Romance)
6	Krym-Tatar, Bashkir, Gagauz (Turkic), West Baluchi, Dari (IE/Iranian), Basque (isolate), Kalmyk (Mongolic)
5	German (IE/Germanic), Hindi (IE/Indoarian), Kyrgyz (Turkic), Korean (isolate)

(Continued)

Table 5. Continued

4	French (IE/Romance), Greek (IE/Greek), Icelandic (IE/Germanic), Kazakh (Turkic), Ket (isolate), Nenets, Nganasan (Uralic/Samoyed), Itelmen (Chukchi-Kamchatkan), Aleut (Eskimo-Aleut)
3	Oroch, Orok, Nanay, Negidal (Tunguso-Manchu), Welsh (IE/Celtic)
2	Dutch (IE/Germanic), Mari (Uralic/Finnic), Ainu, Nivkh (isolates)
1	Kerek, Koriak (Chukchi-Kamchatkan), Yakut (Turkic), Finnish (Uralic/Finnic), some forms of Dutch.

The areality in the distribution of sibilant systems in Eurasia is quite obvious, the Caucasus manifesting the region with the exceptionally developed sibilant systems.

### 7.2.6 *Rich postvelar consonant systems*

The presence of postvelar (uvular, pharyngeal and laryngeal) consonants is mentioned by some authors as another pan-Caucasian feature (cf. Catford 1977: 308; Klimov 1986a: 172). Though Tuite (1999: 5) regards this feature as non-essential, arguing that these sounds can be found also outside the Caucasus, it is the striking richness of the system of postvelars that renders it a specific feature of the Caucasian linguistic area, which has no analogues among the languages of Eurasia.

Table 6a. The number of postvelars in the Caucasian languages

26+	Proto-West-Caucasian
21	Ubykh
16	Abaza
14	Bzyp Abkhaz
13	Burshag Aghul
12	Abzhywa Abkhaz
11	Akhvakh, Kryz, Bzhadugh Adyghe, Kabardian
10	Dargi, Archi, Budukh, Khinalug, Abadzakh Adyghe
9	Bats, Chamala, Tindi, Lezgi, Tsakhur, Temirgoy Adyghe
8	Chechen, Ingush, Avar, Andi, Inkhokvari Khvarshi, Tsez, Hinukh, Bezhta, Hunzib, Lak, Tabasaran, Rutul
6	Proto-Kartvelian, Svan, Old Georgian, some modern Georgian dialects, Udi
5	Standard Georgian, Megrelian, Laz

The basic Caucasian systems include five postvelars: uvular stops  $q$   $q'$ , uvular fricatives  $\text{ɣ}$   $\text{χ}$ , and pharyngeal/laryngeal  $H/h$ . This is what we basically have in Proto-Kartvelian, Svan, Old Georgian and some modern Georgian dialects, although Standard Georgian, Megrelian and Laz lost  $q$ , and in Megrelian and Laz dialects  $q'$  is substituted by a glottal stop. In the NC languages the systems of postvelars are at least twice as rich as in SC, while the NWC languages again hit the record. Typical for many



Caucasian languages is the absence of the voiced counterpart of *q*, i.e., *G*. This consonant is found only in such NEC languages as Andi, Dübek Tabasaran, Rutul, Tsakhur, Kryz, Budukh and Khinalug, though it is reconstructed for both NC groups and for Proto-NC. It has been observed that “in Caucasian languages if there is only *one* type of dorsal fricative it is always uvular, not velar” (Catford 1977: 288). The same principle holds for both Ossetic and Armenian.

The non-Caucasian languages of the region and the contiguous languages show variegated postvelar systems:

Table 6b. The number of postvelars in the non-Caucasian language

Areal non-Caucasian languages	Contiguous languages
6 Iron Ossetic	4 Persian, Neo-Aramaic
4 Kumyk, Tat	3 Kurdish
3 Armenian, Digor Ossetic, Nogay, Azeri, Talysh, Karachay-Balkar	1 Turkish
	0 Greek

The following table demonstrates the universal rarity of developed postvelar systems among the languages of Eurasia, with the few exceptions represented by Arabic or isolated Nivkh.

Table 7. The number of postvelars in the languages of Eurasia

7	Arabic (Afroasiatic/Semitic)
6	Nivkh (isolate)
5	Karakalpak (Turkic), Ket (Yenissey), Burushaski (isolate)
4	Itelmen (Chukchi-Kamchatkan), Kazakh (Turkic), Parya (IE/Indoarian), Tajik (IE/Iranian)
3	Wakhi, Baluchi, Rushan (IE/Iranian), Bashkir, Krym-Tatar, Karaim (Turkic), Koriak (Chukchi-Kamchatkan)
2	Ainu (isolate), Aleut (Eskimo-Aleut), Karelian (Uralic/Finnic), Kyrgyz (Turkic), Negidal (Tunguso-Manchu)
1	Vod, Finnish (Uralic/Finnic), Welsh (IE/Celtic), Albanian (IE), German, English, Dutch, Icelandic (IE/Germanic), Hindi (IE/Indoarian), Latvian (IE/Baltic), some Basque dialects (isolate), Korean (isolate), Nenets, Nganasan (Uralic/Samoyed), Yakut (Turkic)
0	French (IE/Romance), Lithuanian (IE/Baltic), Greek (IE/Greek), Russian (IE/Slavic), Hungarian (Uralic/Ugric), Mongolian (Mongolic), Mordvin, Udmurt, Permiak (Uralic/Finnic), Oroch, Orok (Tunguso-Manchu)

7.2.7 Simple resonant systems

Similarities in the general make-up of the obstruent system continue in the system of resonants. All Caucasian languages are characterized by a simple system of resonants: “two nasal (*m n*), a labial semivowel or fricative (*w/v*), a palatal semivowel (*j*),

an apical trill (*r*), and all but the Adyghean languages have a lateral approximant (*l*)” (Catford 1977: 288). Ubykh should be added to Adyghean in lacking the lateral resonant (historically a secondary loss in both languages); besides, Standard Georgian lacks *j*. Resonants are not syllabic (with a possible exception of semivowels), cannot be glottalized, and normally (with few exceptions) lack any other secondary features, thus presenting a sharp contrast to the class of obstruents.

In other linguistic zones of Western Eurasia the systems of resonants are also rather simple, so that this feature is not specific to the Caucasus. What is typical for the Caucasus is the lack of any secondary features for resonants, unlike obstruents, as well as the lack of the velar nasal resonant (*ŋ*), which is a typical phoneme in Turkic and in many Indo-European languages. The neighbouring Kumyk, Nogay and Karachay-Balkar have *ŋ*, and Armenian and Kurdish have two kinds of rhotics.

### 7.2.8 *Vocalic systems*

In contrast to consonants, the vocalic systems in the Caucasian languages are not uniform at all. The NWC vocalism is characterized by a minimal number of vowels, again hitting world records: 3 in Circassian and Ubykh (historically 2) and 2 in Abkhaz and Abaza. In Kartvelian the number of vowels is quite moderate: 5 phonemes in Georgian, Megrelian and Laz; a somewhat richer vocalism characterizes the Ingilo and Fereydan dialects of Georgian. The number of vowels in Svan, due to secondary developments, amounts to 18. In NEC the picture is quite variegated, from the moderately developed vocalism in languages like Avar, Lezgi, Aghul (5 vowels), Tabasaran (6), Lak (7) to richer systems of Rutul (8 vowels), Budukh, Kryz (each 9), Botlikh, Karata (each 10), Hinukh, Archi, Udi (each 11), and still richer in Godoberi (13), Tsakhur (14), Khvarshi (16), Chamala (17), Akhvakh, Tindi (both 20), Bezhta (23), Tsez, Hunzib (both 24), Bats (15 vowels + diphthongs), Ingush (28) and Chechen (15 vowels, 15 phonemic diphthongs).

Among the vocalic traits, which can tentatively be listed as common Caucasian, one can mention one positive and one negative:

1. The presence of schwa (or a schwa-like phoneme). It is attested in all NWC languages. In Kartvelian it is present in Megrelian and Svan, both bordering on Abkhaz, and is reported for some forms of Laz. This phoneme is unknown in Chechen or Bats, but is attested in Ingush, where it is of later origin. Phonemic schwa is present in 10 Dagestani languages (Khvarshi, Hunzib, Archi, Rutul, Tsakhur, Kryz, Budukh, Khinalug, Nidzh Udi, Kimil Lezgi), but not attested in 17 others (Avar, Andi, Akhvakh, Chamala, Tindi, Botlikh, Tsez, Godoberi, Karata, Hinukh, Bezhta, Lak, Dargi, Lezgi, Tabasaran, Aghul, Vartashen Udi). At least in some languages the presence of (phonetic) schwa can be attributed to areal factors (e.g., in Megrelian – under the Abkhaz influence, in Nidzh Udi, Kryz, Budukh, Khinalug, Kimil Lezgi – under the influence of Azeri).

Outside the Caucasian languages, in Iron Ossetic schwa has developed from other vowels, which may suggest an areal stimulus, and in Armenian it was present already in the classical period (i.e., since 5th c. AD).

2. A common negative feature is the lack of monophonemic diphthongs. The exceptions are rare and involve mostly the Nakh languages.

### 7.2.9 *Phonostatistics*

Certain phonostatic traits can be regarded as specific to the Caucasus. A most characteristic feature here is the difference in frequency of voiced stops, affricates and fricatives: in stops the frequency of the voiced ones is higher than of the voiceless (aspirated or glottalized), whereas in affricates and fricatives the situation is reverse (cf. Melikišvili 1976: 159; SO 103–104).

Other common Caucasian phonostatic traits are:

1. In the speech sequence, consonants dominate over vowels.
2. The rarity or, in many Daghestanian languages, absence of the glottalized  $p'$ .<sup>24</sup>
3. The rarity or complete absence of the voiced uvular stop  $G$ .
4. The rarity or complete absence of labial fricatives  $v$ ,  $f$ ; if present, nearly everywhere they are of secondary origin.

### 7.2.10 *Phonotactics*

#### *Consonants*

In the consonant systems, the following common traits can be mentioned:

1. The presence of harmonic clusters, which constituents possess identical laryngeal features (voiced, glottalized, voiceless) and which can be regarded as unitary segments; besides, a distinct preference for decessive clusters (labial+dental, dental+velar, etc.) over the acessive (velar+dental) ones (cf. Catford 1977: 292–3).
2. The absence or rarity of the initial  $r$ - (with some exception of SC).
3. The weak devoicing of voiced consonants in final position.

#### *Vowels*

In the vocalic system, the typical Caucasian phonotactical traits are:

1. The absence or rarity of vocalic sequences (hiatus).
2. The prohibition of vowel-initial syllables, which means the impossibility of vocalic Anlaut (SO 104), with few exceptions.<sup>25</sup>

---

24. Kodzasov (SO 92) explains this rarity by the universal-phonetic unnaturalness and articulatory difficulty of  $p'$ , which is doubtful:  $p'$  is much easier articulated than many other Caucasian consonants; its rarity, as observed by Trubetzkoy (1987: 55), has probably more to do with the fact of it being the highly marked member of the correlation, as compared with the unmarked voiced or aspirated counterparts, which conditions its structural weakness.

25. Abkhaz has, due to later developments, a number of roots with initial  $a$ -; as to the hiatus, in Georgian and Laz the combinations of vowels are more or less normal; by contrast, in

### 7.2.11 *Basic syllable and morpheme structure*

In the majority of Caucasian languages both open and closed syllables are possible (with very few exceptions, like Northern Akhvakh) (cf. SO 105). The common features in the morphemic structure are:

1. The preponderance of mono- or disyllabic nominal roots.
2. Affixal morphemes are mostly represented by the structures like V, C, CV.

### 7.2.12 *Prosody*

Typical for the majority of the Caucasian languages is the presence of dynamic word stress, with some (Daghestanian) languages displaying also probably tonal contrasts.

## 7.3 *Morphophonology: Ablaut*

Ablaut as a morphological mechanism is used in both NC groups and in the SC languages. In NWC ablaut serves to differentiate transitive and intransitive stems (in Circassian), as well as extravert vs. introvert verb forms. In the Nakh languages ablaut differentiates tense and iterative aspect forms, in Daghestanian it is used to create aspectual forms and imperatives, as well as (e.g., in Dargi, Avar, Hunzib) to differentiate singular and plural nominal forms. In Lak the apophony *a ~ u* is used in the creation of the so-called oblique stems. In Old Georgian, ablaut was used to express transitivity/intransitivity (cf. SO 65, 109), while in modern Georgian the apophony *e/i, ø/a* characterizes the opposition of Present vs. Aorist forms (cf. Shimomiya 1978: 112).

## 7.4 *Morphosyntax*

The morphological similarities between the Caucasian languages can be summed up as follows:

1. Agglutination as a major structural type, with some elements of fusion. Turkic languages are also agglutinative, while Armenian and Ossetic are fusional, with elements of agglutination; Kurdish and Neo-Aramaic are also fusional. What is interesting here, is that Armenian (in nominal plural paradigms) and Ossetic (in its case system) have developed elements of agglutination. If we recall that Turkish is actually a new language in the area (from the 13th–14th centuries), which replaced the predominantly fusional Greek and Armenian, the agglutinative character of the Caucasian languages, partially spread to some basically fusional languages of the area, acquires a strong areal significance.

---

Megrelian and Svan vocalic sequences are often avoided by the insertion of epenthetic sounds; a vocalic anlaut can be avoided in Svan by means of prothetic resonants and in Megrelian – by the glottal stop.

2. All Caucasian languages are characterized by a high degree of synthesis (polysynthesism), having its peak in NWC, with more moderate forms in NEC, the SC languages being somewhere in between (cf. Klimov 1986a: 147).
3. A weak formal distinction between nouns and verbs, and between nouns and adjectives (SO 66).
4. A well developed system of verbal tempora (SO 68).
5. Diachronically, greater antiquity of aspectual over temporal distinctions: the latter have been formed on the basis of the former (SO 71–72).
6. The presence of a morphologically expressed category of evidentiality attested in the majority of the indigenous Caucasian languages,<sup>26</sup> as well as in Armenian, though lacking in Ossetic and diminished (under the influence of Persian) in Azeri. Tuite (1999: 5), briefly mentioning evidentiality as a shared feature of the Caucasus area, regards it rather as a ‘Circumpontic’ or Balkano-Caucasian linguistic feature. Even though evidentiality is indeed attested, besides the Caucasus, in the Balkans (Bulgarian, Macedonian, Albanian) and in Anatolia (Turkish), this still allows us to regard this category as another important pan-Caucasian morphological trait. As a distinct morphological category, evidentiality is not universally attested in the languages of the world but is regarded as “extremely prone to diffusion”, and “the emergence and loss of evidentiality systems is often due to intensive language contact” (cf. Aikhenvald 2003: 21), which is one of the explanations of its spread in the Balkans (presumably, under the influence of Turkish, which means that it is historically a late innovation there; let us remember, that Turkish is also a comparatively new language in the area, and its predecessor in Anatolia, Greek, did not have this category). The widespread attestation of evidentiality across the Caucasus and the possibility of its contact-induced origin in at least some languages of the area render it an important feature of the Caucasian linguistic area.
7. The presence in all three groups of a morphologically expressed category of potential, i.e., morphological means to express ability to produce an action. Examples:
  - a. in NWC: Abkh *də-z-tʰa-wa-m* ‘(s)he-POT-sit-PRES:DYN-NEG = (s)he cannot sit down’;
  - b. in NEC: Chech *kxossa-vala* ‘to jump-POT = to be able to jump’, Bagv *ašt-a-š* ‘listen-POT-FUT1 = will be able to listen’;
  - c. in SC: Geo *ar m-e-čʼm-eb-a* ‘not I/me-IOV-eat-PASS-X = I cannot eat X’, Megr *a-čʼar-e(-n)* ‘IOV-write-PASS-X = (s)he can write it’.
8. The morphological marking of causative can also be listed among the pan-Caucasian traits.

---

26. Tuite rightly notes the absence of the evidential category in Ubykh, but it is certainly present in Circassian. It seems that evidentiality is also lacking in Udi (W. Schultze).

9. The predominance of postpositional constructions as a common Caucasian trait. Examples:

- a. in NWC: Abkh *lara l-ax'* 'she her-to = to her';
- b. in NEC: Chech *govr-ana k'èlah* 'horse-DAT under = under the horse', Udi *k'òž-in beš* 'house-GEN before = before the house';
- c. in SC: Geo *mta-ze* 'mountain-at = on the mountain'.

As to the non-Caucasian languages, in Armenian, of numerous prepositions in the classical period (around 6 c. AD) only few survived by the present time, whereas the modern language has developed a great number of postpositions, cf. *jeṛk'-i meṭ* 'hand-GEN/DAT in = in the hand'. Likewise, modern Ossetic, originally prepositional, preserves only two prepositions (*ænæ* 'without', *æd* 'with') and has developed instead numerous postpositions, cf. *xæžar-y sær* 'roof-GEN head = on the roof', as contrasted to Kurdish *ber deri* 'by the door'. The development of postpositions in both Ossetic and Armenian is attributed by specialists to Caucasian influence (cf. Abaev 1995: 488, 498–499).

10. The attachment of coordination markers to each conjunct as a pan-Caucasian trait, cf. in NWC: Abkh *war-g'ə sar-g'ə* 'you-and I-and = you and me'; in NEC: Chech *vaša a jiša a* 'brother and sister', Avar *gazet-al-gi žurnal-al-gi* 'newspapers and magazines'; in SC: Megr *ma-ti si-ti*, Svan *m-i s-i* 'I-and you-and = I and you'.<sup>27</sup>

11. The presence in the majority of the Caucasian languages of one model of nominal declension: base (root) + plural marker + declension marker, e.g., Ad *labže-xe-mč'e* 'claw-PL-INSTR = by/with claws', Lez *ruš-ar-iz* 'girl-PL-DAT = to the girls', Geo *muxl-eb-it* 'knee-PL-INSTR = by/with knees' (cf. Klimov 1986a: 145). The same agglutinative model is manifested in otherwise fusional Ossetic (*k'ux-t-ə* 'hand-PL-GEN = of the hands') and Armenian (*ban-er-i* 'thing-PL-GEN/DAT = of the things'). The development of the agglutinative declension in Armenian could have been stimulated by Caucasian or/and Hurro-Urartian influence (cf. Hurri *šavala-na-až-a* 'year-ART:PL-PL-LOC = in the years', Urartu *huradi-na-we* 'soldier-ART:PL-DAT = to the soldiers'). Interestingly, Classical Armenian had the reverse order of morphemes, cf. *ban-i-w-k'* 'thing-decl.marker-INSTR-PL = by the things' as opposed to modern Armenian *ban-er-ow* 'thing-PL-INSTR'. The declensional model in modern Armenian is thus closer to that of the Caucasian languages and Hurri-Urartian (and, for that matter, Turkish, cf. Tu *köy-ler-in* 'village-PL-GEN = of the villages'), than in Classical Armenian. In Ossetic, this development is attributed to Caucasian influence (Abaev 1995: 496).

27. Outside the Caucasus, a similar model is found in Latin, Sanskrit, some Uralic and Dravidian languages, Burushaski and Japanese (B. Tikkanen, p.c.), but not in the contiguous languages, which underlines its areal character in the Caucasus.

12. The importance and in some groups predominance of prefixal conjugation; in NWC and SC prefixes express person (in Abkhaz also nominal class), version, potential and other grammatical functions, in NEC prefixes mostly express nominal classes, cf. Abkh *də-q'a-w-p'* '3P.SG.HUM-be-STAT:PRES-FIN = (s)he is', Chech *b-aha* '4class:NHUM-carry = it carries', Avar *j-igo* '2class-be = is, exists', Geo *v-ar* '1P.SG.-be = I am'.

13. Another pan-Caucasian feature is the use of directional and orientational preverbs, attested in all three groups. The diffusibility of preverbs is especially clearly seen in the West Caucasian-Kartvelian zone of interaction. The number of preverbs in the West Kartvelian languages is much higher than in Georgian, Laz boasting 50 preverbs, while Megrelian, directly bordering on Abkhaz (which is characterized by a highly elaborated system of preverbs), has as many as 92 simple and compound preverbs (cf. Čikobava 1977: 15–16). Some of the Megrelian preverbs are directly borrowed from Abkhaz, such as *k'əla-* 'through', cf. Megr *k'əla-sxap'ua*, Abkh *a-k'əl-pa-ra* 'to jump out' (cf. Chirikba 1998: 139).

14. Like the presence of glottalized consonants in the phonological system, in morphology it is ergativity, which is generally regarded as the trademark of the Caucasian linguistic type. Both NWC and NEC have more or less unambiguous expression of ergative alignment, though even within these groups there can be differences in its formal realization: conjointly by nominal class, person markers plus cases (as in Bats, Dargi, Lak, Tabasaran), by class markers plus cases (as in the majority of NEC languages), by person markers plus cases (as in Circassian and Ubykh), only by cases (as in Lezgi and Aghul) and only by class/person markers (as in Abkhaz). In both NWC and NEC groups ergativity is tense-neutral (i.e., no split).

Against this background, there are serious disagreements among specialists as to the ergative character of Kartvelian, cf. e.g., Hewitt (1995), who defends the traditional view regarding Georgian as manifesting ergativity, and Harris (1990), who treats this language as manifesting active, rather than ergative typology; still others regard Kartvelian as mixed, combining elements of nominative and active systems, with the predominance of the former (cf. Klimov and Alekseev 1980: 299). In general, even if we regard Georgian as manifesting ergative alignment, it represents a split ergative type, realizing its ergative strategy in the past tenses, which differs sharply from what we see in the two NC groups, where ergativity is tense-neutral.

Apart from the indigenous languages of the Caucasus, ergative-like perfect constructions are attested in Armenian, which resemble the Georgian-type split ergativity system and which appeared, as suggested by Meillet (1936: 95), due to the Caucasian, probably SC influence. The split ergativity system in the past-tense transitive constructions is present also in Kurdish and Talysh, which resembles Georgian more closely. Some authors even ascribe the development of the ergative strategy in Kurdish to the influence of some Caucasian idioms (like Hurrian or Georgian), though this explanation will obviously not hold for the origin of split ergativity systems in other Indo-Iranian languages.

15. Group inflection, i.e., the morphological marking of only one constituent of a phrase, is another common pan-Caucasian morphosyntactic trait, cf. Ad *biraq* *płǝž-xǝr* 'flag red-PL-ABS = red flags', Lez *čulav yal-ar* 'black thread-PL = the black threads', Megr *da do žima-s* 'sister and brother-DAT = to the sister and brother', Laz *didi žal-epe-s* 'big tree-PL-DAT = to the big trees' (cf. SO 56). Of the neighbouring languages, the same is observed in Ossetic, Armenian and Tat (cf. Abaev 1995: 285; Grjunberg 1997: 146).

## 7.5 Word Formation

In word-formation, three features can be singled out as having a pan-Caucasian character:

1. The pan-Caucasian parallelism of compounds of the type *bahuvrihi*, formed by the combination of noun and its adjective, e.g., Abkh *a-g<sup>o</sup>ǝ-ž<sup>o</sup>p<sup>a</sup>* 'heartless, callous', Lak *dard-du-ssa* 'sad', Geo *gul-gril-i* 'indifferent' (cf. Klimov 1986a: 155; SO 21).
2. Reduplication in all Caucasian groups. Especially interesting are Caucasian parallels in the use of reduplication of a part of the verbal root, the whole root or the root consonant for morphological purposes, noted already by Trubetzkoy (1987: 56–57), who compared it with similar mechanisms in PIE. Cf. the following examples:
  - a. in NWC: Abkh *a-h<sup>o</sup>a-ra* 'to speak' vs. *a-h<sup>o</sup>h<sup>o</sup>a-ra* 'to shout', *a-p+č-ra* 'to break' vs. *a-p+čč-ra* 'to break in many pieces';
  - b. in NEC: Avar *kānc'-ize* 'to jump' vs. the durative form *kānc'-kānc'-ize* 'to jump incessantly', *bek-ize* 'to break' vs. a partial reduplication in *bek-er-k-ize*, Bagv *eta* 'to fly' vs. *et-ita* 'to fly (multiple action)', *b=uq'u* 'to cut' vs. *b=uq'uq'u* 'to crumble', Tsakh imperfective *ikān* 'to want, love' vs. iterative/perfective *ikī kin*;
  - c. in SC: Geo *titxn-*, Megr *txitxon-* 'to make (oneself) dirty', etymologically derived from Geo *tixa* 'clay', Megr *dixa* 'earth, land, place', Geo *ǝg(\*-)-er-* 'to drum' vs. *ǝig-ǝig-* 'to shake, tremble' (cf. Testelec 1992: 161–162).

Of the other languages of the area, cf. similar constructions in Armenian: *ost-n-el* 'to jump', *ost-ost-el* 'to jump repeatedly'.<sup>28</sup>

3. Another interesting areal feature is the popularity of echo-compounding, i.e., the combination of two words or roots, the second of which is the phonetically modified form of the first one. Especially popular is the use of the resonant *m-* as the substitute of the initial sound of the second member of such compounds, cf. Abkh *a-k<sup>ʰo</sup>ša-mək<sup>ʰo</sup>ša* 'all around', Hunz *has-mus* 'horizon', Geo *axlo-maxlo* 'near at hand', Megr *erti-morti* 'manners', etc. This *m*-reduplication actually represents a much wider

28. I am grateful to Hrach Martirosyan for providing me with this and some other Armenian examples.



areal feature, including also Azerbaijan, Armenia, Turkey, Iran, India and even the Balkans (from Turkish ?). But, apart from *m*-, other consonants can also be used for such a purpose. A variant of such type of reduplication is the change of the root vowel in the second constituent, which is especially typical for SC and NEC.

## 7.6 Syntax

In syntax common Caucasian traits are:

1. A verbo-centric sentence structure in all three Caucasian groups, the finite predicate being the syntactically dominant part of the sentence (Klimov 1986a: 152; SO 36).
2. The presence in all groups of the so-called “affective” or “inversive” construction: the verbs belonging to the group of *verba sentiendi* or *verba habendi* present the experiencer NP in the Dative (in the Daghestanian languages sometimes also in the Locative) case, while the source NP is in the nominative (absolutive) case. Examples (cf. SO 44–46):

- a. in NWC: Ad *s<sup>o</sup>əzə-m* (ERG/OBL) *q<sup>o</sup>ă* (ABS) *jə-ʔă* ‘the woman (ERG/OBL) has a son (ABS)’;
- b. in NEC: Ing *so-na* (DAT) *ford* (ABS) *bainab* ‘I (DAT) saw the sea (ABS)’, Avar *insuje* (DAT) *žindirgo limer* (ABS) *boʃʷula* ‘the father (DAT) loves his child (ABS)’;
- c. in SC: Geo *bavšv-s* (DAT) *deda* (ABS) *uqʷvars* ‘the child loves the mother’.

3. All three groups are characterized by relatively free word order. Here the most general principle is the preposing of the governed word before the governing one. The subject NP has a tendency to occupy the leftmost position, the predicate – the final (rightmost) position. The basic word order is thus SOV.

4. Another pan-Caucasian trait is the presence with a group of semantically transitive verbs of indirect, rather than direct objects, whereby a would be direct object is being expressed by an indirect object. Examples (cf. SO 58–59):

- a. in NWC: Abkh intransitive *d-lə-sə-jtʰ* ‘he (ABS)-her(OBL)-hit = he hit her’, as opposed to the transitive *də-z-ba-jtʰ* ‘her (ABS)-I (ERG)-saw = I saw her’, Kab *ʔăx<sup>o</sup>ă-r* (ABS) *žāmə-m* (ERG/OBL) *baš-čă* (INSTR) *jăwa* ‘the shepherd (ABS) hit the cow (ERG/OBL) with the log (INSTR)’;
- b. in NEC: Chech *vaša-s* (ERG) *žalie-na* (DAT) *yaž* (ABS) *jetta* ‘the brother (ERG) hit the dog (DAT) with the log (ABS)’;
- c. in SC: Geo *gvel-ma* (ERG) *ukʰbina bavšv-s* (DAT) *pex-ze* ‘the snake (ERG) bit the child (DAT) at his leg’, Svan *dede* (NOM) *lexqʷajan bobš-s* (DAT) *aqʷaži* ‘the mother (NOM) kissed the child’s (DAT) cheek’.

5. In genitive constructions of the structure N+N, the possessor constituent precedes the possessed one. Cf. in NWC: Ub *a-məzə-n ɣə-čʰə*, Abkh *a-čʰkʷən jə-čə* ‘the boy’s horse’; Ing *nana koč* ‘mother’s dress’, Lez *čuban kitab* ‘girl’s book’; Svan *mare-m-iš kor*,

Laz *k'ōč-iš oxori* 'man's house'. The same order is found in both Ossetic and Armenian (cf. Rogava 1987: 9).

## 7.7 Lexical semantics

### 7.7.1 *The nominal system*

The common features in the nominal system are the presence of semantic categories of animate vs. non-animate, on the one hand, and of humans vs. things, on the other hand, found in all groups of Caucasian languages and expressed in various classes of nouns and verbs (cf. SO 7).

### 7.7.2 *The verbal system*

In the semantic organization of the verbal system all three groups manifest the following common traits:

1. The dichotomy between stative (e.g., 'be sleeping', 'be seated', 'be lying') vs. dynamic (e.g., 'to fall asleep', 'to sit down', 'to lie down') verbs. This dichotomy manifests itself, in particular, in the fact that stative verbs lack some (sometimes the majority) of the temporal forms which the dynamic verbs possess; often there are also different affixes to mark stative and dynamic verbs. In many languages the nominal base can be used for forming stative verbs.
2. The presence in all three groups of ambitransitive or labile verbs, which, unlike fundamentally transitive or intransitive verbs, are, depending on the context, either transitive or intransitive, with implications for morphology and syntax. These verbs can be agentive, of the type of Abkh *s-pa-wa-jt'* 'I-knit-PRES-DYN:FIN = I am (busy with) knitting' (ITR) vs. *jə-s-pa-wa-jt'* 'it-I-knit-PRES-DYN:FIN = I am knitting it' (TR), or patientive, cf. Abkh *jə-p+čə-ø-jt'* 'it-break-AOR-DYN:FIN = it broke' (ITR), as opposed to the transitive *jə-pə-s-čə-ø-jt'* 'it-PREV-I-break-AOR-DYN:FIN = I broke it'.
3. The presence in all groups of suppletive verbs for singular and plural arguments (in NWC this phenomenon is limited to Ubykh). The examples are from (SO 17):
  - a. NWC: Ub *s(ə)*-(SG) vs. *ž°a*-(PL) 'to sit', *t°*-(SG) vs. *xa*-(PL) 'to stand, be somewhere', *t°a*-(SG) vs. *q'a*-(PL) 'to give';
  - b. NEC: Bats *daar*(SG) vs. *daxk'ar*(PL) 'to come', *lallar*(SG) vs. *laxk'ar*(PL) 'to drive away', Lak *ivč'an*(SG) vs. *lit'un*(PL) 'to kill', Archi *k'is*(SG) vs. *xvis* (PL) 'to die';
  - c. SC: Geo *ždoma* (SG) vs. *sxdoma* (PL) 'to sit', *vardna* (SG) vs. *cvena* (PL) 'to fall'.
4. The presence of the lexico-grammatical class of deverbal nouns or masdars. Masdars resemble the English gerunds, manifesting both nominal and verbal features, and can be formed, with rare exceptions, from nearly any verb. They are usually not formally differentiated on the principle 'transitive-intransitive', with the exception (sometimes) of Georgian and Adyghe. Unlike SC and NWC, NEC languages have both masdars and infinitives (cf. SO 18), though Old Georgian did have an infinitive.

7.7.3 *The pronominal system*

In the pronominal system three common features can be pointed out:

1. The distinction between the interrogative pronouns ‘who’ and ‘what’ in conformity with the category of “human ~ non-human”, cf. examples from (SO 19):
  - a. in NWC: Abkh *d-arban* ‘who?’ vs. *j-arban* ‘what?’, Ad *âât?* vs. *səd?*;
  - b. in NEC: Chech *mila?* vs. *hun?*, Lak *cu?* vs. *ci?*, Tab *fuž?* vs. *fu?*;
  - c. in SC: Geo *vin?* vs. *ra?*, Svan *jär?* vs. *maj?*.

This distinction is of course not specific to the Caucasus, though it is not universal either. In the neighbouring non-Caucasian languages (Ossetic, Azeri, Kumyk, Karachay-Balkar, Armenian) and in the contiguous areas (Turkish, Kurdish) the situation is similar.

2. A three-grade deictic distinction of the type of Latin *hic-iste-ille* in all three Caucasian groups (SO 19). A similar ternary system of deixis is present also in Kumyk (*bu – şo – o*) and Armenian (*ays – ayd – ayn*) and, outside the Caucasus, in Turkish (*bu – şu – o*) (cf. also the Balkan languages), whereas Karachay-Balkar, Ossetic, Kurdish, Tat and Talysh have a binary system (‘this’ – ‘that’).

3. A similar ternary organization of the pronominal locative adverbs: ‘here’ (close to the speaker) ~ ‘there’ (further) ~ ‘over there’ (far away) (cf. SO 20). The same three-way opposition of locative adverbs exists in Armenian (*aysteł- aydteł- aynteł*).

7.7.4 *The numerical system*

1. An important common Caucasian feature is the vigesimal model of derivation of numerals, whereby, starting from ‘thirty’, the numeral ‘twenty’ lies at the basis of numerals:  $30 = 20 + 10$ ,  $40 = 2 \times 20$ ,  $50 = 2 \times 20 + 10$ ,  $60 = 3 \times 20$ ,  $70 = 3 \times 20 + 10$ ,  $80 = 4 \times 20$ ,  $90 = 4 \times 20 + 10$ . In some other Caucasian languages a decimal system is used, based on the numeral ‘ten’. The distribution of different systems in the Caucasus can be shown in the following table:

Table 8. Numeral systems in the Caucasus

Languages	Vigesimal system	Decimal system
NWC	Abkhaz, Abaza, Adyghe, Kabardian dialects, Ubykh	Kabardian, partly Abaza
NEC		
Nakh	Chechen, Ingush, Bats	
Avar	Avar	
Andic	Akhvakh	Andi, Botlikh, Godoberi, Karata, Bagvala, Tindi, Chamala

(Continued)

Table 8. cont'd

Languages	Vigesimal system	Decimal system
Tsezic	Tsez, Khvarshi, Hinukh	Bezhta, Hunzib
Lezgi	Lezgi, Aghul dialects, Tabasaran dialects, Udi, Kryz, Budukh, Khinalug	Tabasaran dialects, Aghul dialects, Rutul, Tsakhur, Archi
Lak-Dargi		Lak, Dargi
SC	Georgian, Megrelian, Laz, Bal and Lentekh Svan	Upper Bal and Lashkh Svan
IE	Ossetic	Armenian, Ossetic shepherds' jargon and Literary Ossetic, Kurdish, Tat, Talysh, Greek
Turkic	Karachay-Balkar	Karachay-Balkar, Kumyk, Nogay, Turkish
Afroasiatic		Neo-Aramaic

Some languages have parallel decimal and vigesimal systems, and some mixed systems. Thus, in Kabardian dialects, sometimes even within one speech community, one can find an archaic vigesimal system, a more popular decimal system and a mixed decimal-vigesimal system (cf. Turčaninov and Cagov 1949: 66–67). Interestingly, in the Kuban dialect of Kabardian, whose speakers have been living since the end of the 18th century amongst the closely related Adyghe, the decimal system is being replaced, under the influence of Adyghe, by the parallel vigesimal one (Keraševa 1977: 35). In Adyghe, which is basically vigesimal, '30' has two parallel forms, vigesimal (20-and 10-and) and decimal ( $3 \times 10$ ). In Abaza, parallel to a more standard vigesimal system, a decimal system was formed under the influence of Kabardian. In Inkhokvari Khvarshi, beginning from 50, the original vigesimal system is preserved alongside a parallel decimal one borrowed from an Andi language (Bokarev 1967: 427). Archi, an archaic member of the Lezgi subgroup, stands apart, as it forms its numerals 40, 70, 80, 90 by means of the metathesis of the corresponding simple numerals 4, 7, 8 and 9 (cf. Xajdakov 1967: 615). In the dialects of Tabasaran and Aghul one can find the older vigesimal, the decimal and the mixed vigesimal-decimal systems. In some of these dialects the numeration up to 'sixty' is based on the decimal system, and above 'sixty' – on vigesimal, and sometimes both decimal and vigesimal systems are used in parallel (Magometov 1970: 168).

The majority of the Caucasian languages thus use the vigesimal system, and those which use the decimal one either form isolated enclaves against the background of predominantly vigesimal sister languages, like NWC Kabardian and SC Upper Bal and Lashkh Svan, or small contiguous areas, like Andic (with the sole exception of Akhvakh), a part of the Tsezic and Lezgi languages, and the Lak-Dargi cluster. Interestingly, the dialects of Kabardian, Tabasaran, Aghul and Svan still preserve the more archaic vigesimal systems. We can thus see a distinct areality in the distribution of both systems

in the Caucasus: the vigesimal West-Caucasian continuum (Adyghe-Ubykh-Abkhaz-Abaza-archaic Kabardian dialects), the South-Caucasian continuum (Bal and Lentekh Svan-Megrelian-Laz-Georgian), and the Nakh continuum (Ingush-Chechen-Bats).

In Daghestan the situation is more complex: here we have the central (Avar) and south-eastern (Lezgi) vigesimal areas versus north-western (Andic, partially Tsezic) and north-eastern (Lak-Dargi and a part of Lezgi) areas being decimal. The Andic languages form a decimal continuum, with the exception of vigesimal Akhvakh, which is contiguous to vigesimal Avar; the neighbouring decimal Bezhta and Hunzib separate the vigesimal Tsez and Hinukh from Avar; the Bagvala-Chamala-Karata-Botlikh-Godoberi cluster is decimal, and so is the Tsakhur-Rutul-Archi-Lak-Dargi-Aghul-Tabasaran continuum. Finally, the Lezgi-Khinalug-Kryz-Budukh-Udi continuum, despite the decimal Azeri, is vigesimal. The clustering is significant in the sense that it runs across the genetic boundaries: the decimal Lak-Dargi clusters with a part of Lezgi languages, the rest of Lezgi being vigesimal, whereas the vigesimal Akhvakh clusters with the geographically adjacent Avar, the rest of its sister-languages being decimal.

The spread of decimal systems in the Caucasus seems to be contact-induced and secondary in comparison with the older vigesimal systems. Thus, Proto-Lezgi (Alekseev 1985: 68), Proto-Circassian (Kumaxov 1989: 114) and Proto-Abkhaz were vigesimal; the same is true for Proto-Kartvelian. The decimal north-eastern area was in close contact with the decimal Turkic languages (Kumyk, Nogay, Azeri), as well as with Iranian Tat. For the north-western cluster the possible source of the decimal system is not that obvious. For Kabardian, the contacts with the decimal Old Ossetic and Kypchak Turkic languages could have been decisive for the development of the decimal system, and the same explanation holds for the decimal Svan dialects (though later Ossetic itself shifted to a vigesimal system, under the pressure of the neighbouring Caucasian languages).

All non-Caucasian languages of the area (Azeri, Kumyk, Nogay, Armenian, Tat) show fidelity to the original decimal system, except for two languages: Karachay-Balkar and Ossetic, most heavily affected by the neighbouring Caucasian languages. The first has developed a parallel vigesimal system, whereas Ossetic preserved the Iranian decimal system only in counting cattle (it has also been adopted in Literary Ossetic), the "Caucasian" vigesimal system having taken over. In both Karachay-Balkar and Ossetic the development of the vigesimal system of counting is undoubtedly contact-induced, under the influence of the neighbouring Caucasian languages.

Outside the Caucasus, the majority of European languages are decimal, with such vigesimal islands as Basque, old Celtic languages (the vestiges of vigesimal counting are preserved also in some modern Celtic languages and in French), Danish, Faroese, South Italian (Calabrian, Sicilian) dialects; there are some traces of this system in Albanian. In Asia, the few vigesimal languages are Burushaski (isolate), several Indo-Iranian languages of the Pamir and Hindu Kush area (Kati, Pashai, Yazguliemi, Yaghnobi, Baluchi), some Tibeto-Burman languages (Dzongkha, Tamang), Munda (Austroasiatic, India), Ainu (isolate, Japan), and some others. In some of these languages

the emergence of the vigesimal system is attributed to the influence of substrate (cf. the supposed Burushaski substrate for some of the mentioned Indo-Iranian languages, the pre-Indo-European substrate in insular Celtic and some Basque/Aquitania-type substrate in Gallo-Romance) or language contact (e.g., the vigesimal count brought into Sicily and Southern Italy by the Normans).<sup>29</sup> Outside Eurasia, the vigesimal systems are known in some African (e.g., Yoruba), Amerindian (e.g., Maya, Chol, Nahuatl) and Oceanic (Alamblak, Papua New Guinea) languages.

The strong areal character of the Caucasian vigesimal system is obvious: the contiguous non-Caucasian languages have decimal systems, and at least some of the non-Caucasian languages of the area (Karachay-Balkar and Ossetic) have adopted the "Caucasian" vigesimal system.

2. Similar derivation of distributive numerals by means of reduplication of simple forms of the corresponding numerals, cf. Abkh *pš'ba-pš'ba* 'four each', Chech *iss-iss* 'nine each', Avar *šu-šu* 'five each', Geo *at-ati* 'ten each' (cf. SO 23). This reduplication type is used also in Azeri (*iki-iki* 'two each'), Armenian (*čors-čors* 'four each'),<sup>30</sup> Kurdish (*dödö-dödö* 'two each') and may have an areal origin.

#### 7.7.5 Other cases

The semantic structure of many Caucasian words has similar models. I shall cite here just a few such examples, which have a pan-Caucasian distribution.

1. In all three groups the internal semantic structure of the word 'parents' is based on a compound of the type 'mother+father', with the left place usually occupied by the word for 'mother' (cf. SO 23). Examples:

- a. in NWC: Abkh *anə-j=abə-j* (mother-and-father-and);
- b. in NEC: Bats *nan-da* (mother-father), Lez *dide-buba* (mother-father), Rut *nin-did* (mother-father), Bud *dide-ada* (mother-father);
- c. in SC: Geo *ded-mama*, Svan *dī-mū* (mother-father).

Of the non-Caucasian languages cf. also Kar-Balk *ata-ana* 'parents', with a different order of constituents ('father-mother').

2. A pan-Caucasian isogloss is a distinction between pairs of adjectives meaning 'thick' and 'thin', the choice of which depends on whether the referent is flat (e.g., paper, leaf, skin) or roundish (as finger, log, etc.; cf. SO 23–24).

29. Cf. Blažek (1999: 333–334).

30. Another model in Armenian is suffixal: *čors-akan* 'four each'; cf. Turkish *dörd-er* 'id'.

Table 9. Suppletive forms for ‘thin’ and ‘thick’ in the Caucasus

	‘thin’		‘thick’	
	flat	roundish	flat	roundish
Abkhaz	<i>a-c’aya</i>	<i>a-p’a</i>	<i>a-ž’p’a</i>	<i>jət’əw</i>
Kabardian	<i>p’aš’e</i>	<i>psəy’ə</i>	<i>l’əv</i>	<i>γ’əm</i>
Georgian	<i>txel-i</i>	<i>c’wili</i>	<i>skeli</i>	<i>msxwili</i>
Svan	<i>dətxel</i>	<i>necin</i>	<i>sgel</i>	<i>mengre</i>
Dargi	<i>buk’usi</i>	<i>heriṣi</i>	<i>buzsi</i>	<i>burṣusi</i>

The same principle is attested in Ossetic. Outside the Caucasus such semantic correlation is found also in some Indo-Iranian languages, and outside Eurasia – in Quechua-Aymara in South America (SO 23–24; Abaev 1995: 504–505).

3. The combination of meanings ‘to drown’ (of animate referents) and ‘stifle’ in one verb, as opposed to a specific verb or a descriptive formation referring to drowning of inanimate referents (SO 25–26).

4. In the majority of Caucasian languages there are different verbs for ‘to grow (of animate referents)’ and ‘to grow (of inanimate referents)’ (SO 26).

5. A common Caucasian model of the verb ‘to name’ is the combination ‘name+put’ (SO 26); the same model is attested in Armenian and Azeri.

6. The connection of the idea of remembrance (or forgetfulness) with the somatic term ‘heart’: in NWC: Abkh *a-g<sup>o</sup>+a+la+š<sup>o</sup>a-ra* ‘(to fall into heart)’, Ad *g<sup>o</sup>ə-m jəwəbətān* ‘to remember’; in NEC: Lez *rikəl xūn*, Dar *urč’ale bixxare*, Cham *jak’več’ bidla* (lit. ‘to keep on heart’); in SC: Geo *gul-mavic’q’-i* ‘forgetful’ (with forgetful heart); cf. also Oss *zærdyl daryn* ‘to remember’ (‘to keep on heart’) (cf. Abaev 1995: 503).<sup>31</sup>

7.8 Lexical shapes

Lexicon, even if it shows obvious areal colouring, generally is not regarded as central in the definition of linguistic areas. Just to make a picture complete, we shall briefly discuss here some pan-Caucasian lexical items.

1. A common layer of “oriental” words of Arabic, Persian or Turkish origin which have penetrated all idioms spoken in the Caucasus with the spread of Islam, irrespective of the Muslim or Christian affiliation of the speakers. These vocabularies are characterized by a relative uniformity in form and meaning and can be regarded as belonging to the common pan-Caucasian lexical heritage.<sup>32</sup>

31. Outside the Caucasus, cf. similar expressions in Burushaski (B. Tikkanen, p.c.).

32. Cf. the analysis of such terms in the NC languages in Provasi (1981).

2. A common layer of cultural terms, which are not borrowed from Arabic, Persian and Turkish, but are of local Caucasian origin. Klimov (1986a: 190–194) discusses several dozens of such pan-Caucasian cultural lexemes, found in all three groups and in non-Caucasian languages of the area, whose spread is plausibly explained by areal diffusion. Here are just a few of such items:

a. ‘cherry’: in NWC: Kab *balij*; in NEC: Chech *bal*, Ing *boal*, Avar, Hunz, Lez, Aghul *baʃli*; in SC: Geo *bal-i*, Megr, Laz *bul-i*; cf. also Oss *bal(i)*, Kar-Balk *balij*, Arm *bal*. Outside the Caucasus the word is known in Persian: *bālū*. The lexeme is quite old in the Caucasus: cf. the regular sound correspondence between Georgian and Megrelian, underlined by Klimov. According to him, in some of the languages the centre of spread was probably Georgian (at least for Nakh and Ossetic, and possibly also for Kabardian).

b. ‘chicken, hen’: in NWC: Ad *č'etə*, Kab *žed* (< PCirc *\*kaṭə*), Abkh *a-k'o't'*; in NEC: Bats *kotam*, Chech, Ing *kuotam*, Rut *kāt'*, Tsakh *kāt'e*, Bezg *gudö*, Hunz *gudo*; in SC: Geo *katam-i*, Megr *kotom-i*, Svan *kata-l*. For the Nakh languages Klimov supposes a Georgian source.

c. ‘sister/daughter-in-law, bride’: in NWC: Ad, Kab *nəse*; in NEC: Bats, Chech *nus*, Avar, Andi, Botl *nusa*, Dar *nus-* (in a word for ‘doll’), Archi *nus-dur*; in SC: Geo (17th century) *nusa-dia* ‘the wife of uncle’, Megr *nosa//nisa*, Laz *nusa//nisa*. The probable source of this word is IE *\*snuso-s* ‘bride, daughter-in-law’, which reflects, according to Klimov, some very ancient IE-Caucasian contacts.

d. ‘plough’: in NWC: Bzyp Abkh *a-k'atana*, Abzhywa Abkh *a-k'atan//a-g'otan*; in NEC: Bats *guta* ‘plough’, Chech *guota*, Ing *guta* ‘ploughing (equipment)’, Hunz *gotani*, Tsez, Hin, Dar *gutan*, Avar, Lak *kutan*, Lez *köten*, Udi *kötän*; in SC: Geo, Megr *gutan-i*, Laz *kotani*. Cf. also Oss *güton/goton*, Balk *gaton*, Arm *gutan*, Azeri *kotan*. Klimov supposes Armenian to be the source for Georgian and Megrelian, Georgian as the source for Nakh and probably Dargi forms, and Azeri as the source for the rest of the Daghestanian languages. The word is known also in the contiguous Turkish (*gutan* ‘a big plough’) and Kurdish (*kotan*). Whatever the actual source and the etymology,<sup>33</sup> the centre of spread is clearly the Caucasus. Turkish could have borrowed this cultural lexeme from Armenian or Georgian, while Kurdish, by its form, is closer to Azeri, Udi and Laz.

e. ‘melon’ or ‘cucumber’: in NWC: Abkh *a-naša*, Kab *našă* ‘cucumber’, Ad *naš* ‘melon’; in NEC: Chech *närs*, Ing *nars*, Hunz *neso/u*, Bezg *nesi* ‘cucumber’; in SC: Geo *nesv-i*, Svan *k'vax-nesg* ‘melon’. Cf. also Oss *nesi* ‘melon’, Kar-Balk *naša* ‘cucumber’.

f. ‘bandit, brigand’: in NWC: Abkh *abrag'*, Kab *abrăž* ‘armed fugitive, brigand’; in NEC: Chech *oburg*, Ing *äbarg*, Avar *aburik* ‘bandit’; in SC: Geo *abrak'-i*, *abrag-i*, *ap'areka*, Megr *abragi*, Svan *ambreg*. The word is known also in Ossetic: *abyræg/abæreg*

33. On a possible (Middle) Iranian origin see Êdel'man (2007).



and in Balkar: *abrek* ‘bandit’, and penetrated southern Russian dialects, and from there the Literary Russian. Abaev (1958: 25) seeks an Iranian etymology for this word, connecting it with the hypothetical Middle Persian *\*āparak*, cf. Pahlavi *āpartan* ‘to rob’, *āpar* ‘robbery’, Persian *āvāra* ‘vagabond’. Whatever the etymology, this pan-Caucasian word has a strong cultural colouring, reflecting the Caucasian tradition of blood revenge or ‘noble banditry’.

g. ‘head of the feasting-table’, ‘old man’: in NWC: Abkh *a-tahmada*, Kab *themade* ‘old man’; in NEC: Avar, Lak, Lez, Udi *tamada*, Dar *tamada* ‘old man’; in SC: Geo, Megr *tamada* ‘head of the feasting-table’. According to Klimov (1986a: 194), the centre of spread of this word in the Caucasus could be Abkhazo-Adyghean languages, but the word itself, in his view, could have been borrowed from Persian//>Turkish *damad*//*damat* ‘son-in-law’. The lexeme has spread into the non-Caucasian languages of the Caucasus as well (Oss *tamada*, Balk *tamata* ‘head of the feasting-table’), and found its way also into Russian (*tamada* ‘head of the feasting-table’).

7.9 Common traits in phraseology

More than anything, phraseology (on a par with the cultural lexicon) tends to show strong diffusibility. In the Caucasus, too, it is one of the most visible elements of the Caucasian linguistic area. Many of the Caucasian phraseological units, according to Vogt (1942: 250), are calques from one language into another, as testified by the spread of the same models in Ossetic and Armenian (cf. also Klimov 1991: 9).

One of the common traits in phraseology is the use of somatic terms ‘heart’, ‘head’, ‘eye’, ‘hand’, ‘nose’, ‘mouth’, ‘face’, etc. as kernels of phraseological units. Klimov (1986a: 195) gives the statistical data as to the relative frequency of the use of these terms in Kabardian (NWC), Lezgi (EC) and Georgian (SC) expressions, which can be summarized in the following chart.

Table 10. Somatic terms used in sayings

Somatic terms	Kabardian	Lezgi	Georgian
‘heart’	218	174	248
‘head’	58	167	145
‘eye’	111	160	55
‘hand’	52	111	60
‘soul’	130	?	45

8. The contours of the Caucasian “Club” of languages

In the preceding paragraphs I have attempted to summarize the established structural similarities between the three Caucasian groups. As these groups belong to two unrelated families (NC and SC), it is probable that some of these similarities can be

explained by their long-term contact and diffusion, i.e., can have areal nature. On the basis of these commonly shared features it is possible to postulate the existence in the Caucasus of the linguistic union of two unrelated families, the North Caucasian and Kartvelian. In what follows I shall outline the main contours of this Sprachbund, its core and periphery.

### 8.1 The core members

The core members of the Caucasian Sprachbund are the unrelated North Caucasian (with its Western and Eastern branches) and Kartvelian language families. Megrelian and Svan have been in millennia-long contact with the West Caucasian idioms, whereas the Georgian dialects and probably also Svan have been in contact with the East Caucasian languages. This contact, as well as a possible NC substrate in Kartvelian (see below), can account for some of the important structural isoglosses observed between these two families. On the other hand, the Kartvelian influence on the NC dialects was only marginal and manifests itself mainly in the spread of cultural lexicon, not involving, with the exception of heavily Georgianized Bats, grammatical structure.

### 8.2 The peripheral members

The peripheral members of the Caucasian language union are Indo-European Ossetic and Armenian, which share at least some of the features pertaining to the Caucasian linguistic area, first of all in phonology, lexicon and phraseology, and partially in morphosyntax. In both cases specialists speak not only of contact-induced phenomena, but also of a Caucasian substrate. Because of the great time-span of the attestation of both languages in the Caucasus, it is of course very difficult to separate the substrate phenomena from those caused by contact with the neighbouring Caucasian idioms. Besides, both languages have experienced Turkic influence, which is especially noticeable in Armenian. Specialists note striking similarities in results of the Caucasian influence on these IE languages, such as the acquisition of glottalized consonants, the disappearance of formal accusative case, the agglutinative character of nominal declension, the development of postpositions, etc. (cf. Vogt 1988: 281; Abaev 1995: 485–488).

#### 8.2.1 *Armenian*

Armenian, whose contact with the Caucasian languages goes back some 2,5 millennia, has undergone considerable Caucasian influence (cf. Klimov 1986a: 198–199; Greppin 1981: 501). Especially in phonology we observe a remarkable accommodation of Kartvelian and Armenian systems, both in consonants (three series of stops and affricates) and in vowels (simplification of the putative originally richer system, due to the loss of

long vowels).<sup>34</sup> The following features of Armenian can be adduced as manifesting its being an, albeit peripheral, part of the Caucasian linguistic area:

- phonology: the ternary system of obstruent stops and binary system of fricatives; glottalized consonants (in dialects situated in the Caucasus or close to it); a rich system of sibilants; the loss of quantitative distinctions in vowels; the presence of phonological schwa.
- morphology: the absence of the morphologically marked accusative, which coincides with the nominative (with inanimate referents), and with dative (with animate referents); the development of agglutinative (plural) declension; the category of evidentiality; the split ergative-like perfect constructions; the development of preverbs; the loss of grammatical gender; group inflection; the structure of comparative expressions, close to that of Kartvelian and deviating from IE (cf. Shimomiya 1978: 196).
- syntax: the development of declinable postpositions; the SOV word order.

### 8.2.2 *Ossetic*

The evidence of the Caucasian influence on this Iranian language, which has been spoken in the Caucasus at least since ca. V c. AD, is abundant (cf. Klimov 1986a: 200; Abaev 1995: 481–509):

- phonology: the ternary or quaternary system of obstruent stops and the binary system of fricatives; glottalized consonants; a rich system of sibilants; a rich system of uvulars; the presence of labialized consonants; the presence in the dialects of hissing-hushing (middle) sibilants; the contrast of aspirated vs. non-aspirated consonants; the development of the “schwa” phoneme; the loss of quantitative distinctions in vowels;
- morphology: the absence of morphologically marked accusative, which coincides either with nominative or with genitive; the development of agglutinative nominal declension; the vigesimal system of numerals; the development of orientational preverbs; group inflection; the development of a series of local cases;
- syntax: the development of declinable postpositions; the SOV word order.

---

34. Cf. Vogt (1988: 179–180); in Kartvelian only some archaic Svan dialects still preserve the quantitative distinctions in vowels.

### 8.3 The marginal languages

Here belong Turkic Karachay-Balkar, Kumyk, Azeri, Nogay, some forms of Anatolian Turkish in the South Caucasus and in the adjacent areas of north-eastern Anatolia, as well as Iranian Tat. All these languages sometimes bear, apart from numerous lexical items shared with the indigenous Caucasian languages, their phonological or grammatical features. Thus, some of the Kumyk (e.g., the Kaytag dialect) and Azeri (Zakatala-Kakh) dialects, which have NEC substrate, possess glottalized consonants. The Azeri Vartashen and Kutkashen dialects have phonetic features which bear clear evidence of Udi influence or substrate. In the case of Kaytag Kumyk, there is evidence of Dargi influence, and in Zakatala-Kakh Azeri – of Tsakhur substrate. In the latter Azeri dialect there are also pharyngealized vowels found not only in words of Tsakhur origin, but also in Turkic lexemes, as well as a palatalized *l'*, qualified as the evidence of the Caucasian (Tsakhur) substrate: in Tsakhur this resonant is pronounced as phonetically palatalized (cf. Gadžieva 1979: 157–158, 159, 160). Besides, in the northern areas of Azerbaijan, under the influence of the neighbouring NEC languages, there are cases of the violation of vocalic harmony, as, for instance, in the Nukh dialect; this occurs also in Kumyk (in the Kaytag dialect) and in the dialects of Karachay-Balkar. Some Karachay-Balkar dialects (Baksano-Chegem, Upper Balkar) have glottalized consonants (Gadžieva 1979: 27–29, 57). Phonetically, like the voiced stops in the Caucasian languages (and in Ossetic), Azeri voiced stops are also characterized by weak voice, explained by the Caucasian influence (cf. Gadžieva 1979: 55).

Karachay-Balkar has a parallel vigesimal numerals system, developed under the influence of the Caucasian languages. The formation of present forms in *-a* with the help of the copula in some northern Azeri dialects (Kuba, Derbent, Tabasaran) is thought to have been stimulated by, or even calqued on, the NEC languages. Besides, in Zakatala-Kakh Azeri there are many idiomatic expressions representing calques from Tsakhur. Moreover, some northern Azeri dialects, adjacent to the Caucasian languages, show parallels to the ergative construction of the Caucasian languages (cf. Gadžieva 1979: 57, 81–82, 160).

Outside the Caucasus, some adjacent north-eastern Turkish dialects show traces of Georgian (the population of the historical Georgian provinces of Shavsheti, Artvin, Kars, Artanudzhi, Artaani) or Laz substrates (the areas between Hopa and Rize), manifested in a specific accent, the presence of a number of Kartvelian lexical items and even reportedly (cf. Gadžieva 1979: 57) of glottalized consonants. In phonology, the Rize-Trabzon Turkish dialect is characterized by the delabialization of vowels, which is explained by the Laz influence (cf. Gadžieva 1979: 26).

### 8.4 The contiguous languages

These are the languages spoken either in the Caucasus itself or in the immediate vicinity of the Caucasus area: Indo-European Talysh, Kurdish, Persian, historically Pontic Greek, Turkic Trukhmen (close to Turkmen and spoken in the North Caucasus) and

Anatolian Turkish, Semitic Neo-Aramaic. The majority of these languages have been spoken in areas contiguous to the Caucasus for a considerable period of time. Though the level of interaction of these languages with the Caucasian idioms was minimal, they sometimes manifest phonetic, grammatical, or lexical features reminiscent of those in the Caucasus. Thus, apart from the “Caucasian” lexical items, Kurdish and Talysh have a split ergative system, whereas eastern Neo-Aramaic dialects have developed glottalized consonants in place of the original emphatics. Pontic Greek in the areas of Trabzon, Hopa, Rize, etc. contacted for a long time with Georgian, Laz and Armenian, which traces can be seen in its vocabulary.

### **9. Multi-tier alliances**

The Caucasus is a broken terrain where, against the background of the pan-Caucasian regional macro-alliance, which represents a Sprachbund of the unrelated NC and SC families (similar to the one attested in the Indian linguistic area), the languages within the three Caucasian groups too form smaller, sub-regional Sprachbünde (which resembles the situation in the Balkan). The NWC, NEC and SC groups have occupied compact geographical zones for a very long period of time, characterized by an intensive internal contact and the spread of regional *lingua francas*. In the Western Caucasus we have a close areal alliance of related Adyghe, Kabardian, Ubykh, Abkhaz and Abaza. Another areal grouping, in the central Caucasus and Daghestan, unites the Nakh-Daghestanian languages. Finally, in the South Caucasus we have an areal union of the related Kartvelian languages: Georgian, Svan, Megrelian (and historically Laz). The geographically neighbouring related languages and dialects also engage thus with each other in a Sprachbund-type relationship.

Within each of the sub-regional groupings there used to be one or, more rarely, two commonly used languages or *lingua francas*. In the Western Caucasus Circassian played a dominant role, serving as a means of communication for Ubykhs and Abazas with their neighbours; as to the Abkhazians, only speakers of some mountain Abkhaz communities might have used Circassian to communicate with their North-Caucasian neighbours. To a certain extent, Circassian did influence Abkhaz in lexicon and phraseology, whereas its influence on Ubykh and Abaza was overwhelming. Within the Nakh-Daghestanian group in many parts it was Avar which was used as a *lingua franca*. Besides Avar, this role was played to some degree also by Turkic Kumyk and Nogay. Within the Kartvelian area it was Georgian which was the dominant language and which has influenced both Megrelian and Svan, and in earlier periods also Laz.

### **10. The lower-level alliances**

Finally, there are smaller areal units, formed by the immediately neighbouring idioms which have developed, over a period of a long-term contact, a considerable amount

of traits not present or not prominent in sister-languages or dialects outside these 'mini-unions'. One of such local alliances is Georgian-Armenian interaction, partially, as some suggest, based on the Kartvelian substrate in some Armenian dialects.<sup>35</sup> Another example of a close local alliance is the Megrelian-Abkhaz interaction in the Gal (historical Samyrzaqan) area of southern Abkhazia.

In what follows I shall briefly describe the sub-regional and local alliances in the Caucasus, underneath the pan-Caucasian linguistic area. In defining such lower-level areal alignments the relevant terms are the core, peripheral and marginal languages, the dominant idiom and the common superstrate language(s).

## 10.1 Sub-regional alliances

### 10.1.1 *The North-Western Caucasus contact area*

<i>number of languages in contact</i>	<i>core languages</i>	<i>peripheral languages</i>	<i>marginal languages</i>	<i>dominant languages or lingua francas</i>	<i>superstrate languages</i>
12	Abaza, Abkhaz, Circassian (Adyghe, Kabardian), Ubykh <sup>36</sup>	Balkar, Karachay, Nogay, (Old) Ossetic	Ingush, Megrelian, Svan	Circassian (for Abaza and Ubykh), Abkhaz (for Ubykh)	Arabic, Turkish, Persian

The NWC languages developed in close millennia-long contact and interaction, leading to remarkable structural uniformity, the conservation of archaisms and parallel innovations. Adyghe and Kabardian had a considerable impact on, respectively, Ubykh and Abaza, and marginally also on Abkhaz, which is especially seen in the layer of Circassian lexical borrowings common to these three languages. Abaza has also interacted with Karachay and Nogay and historically also with Alanian (Old Ossetic) and Megrelian, while Kabardian had contact with Abaza, Karachay-Balkar, Nogay, Ossetic, Ingush and Svan. Adyghe had contact with Ubykh, historically with Byzantine Greek and Ottoman Turkish. Abkhaz had contact with Ubykh, Megrelian, historically with Byzantine Greek, Old Ossetic (Alanian) and Ottoman Turkish. All five idioms have a common layer of Muslim (Arabic, Turkish, Persian) words.

35. Shimomiya (1978: 209) speaks of two parallel language unions in the Caucasus: the (pan-)Caucasian and the Armenian-Georgian one.

36. The entire Ubykh population was expelled in 1864 by Russians to Turkey, where they lost their language.

10.1.2 *The North-Central Caucasus and Daghestan contact area*

<i>number of languages in contact</i>	<i>core languages</i>	<i>peripheral languages</i>	<i>marginal languages</i>	<i>dominant languages or lingua francas</i>	<i>superstrate languages</i>
34+	Nakh, Daghestanian	Ossetic, Kumyk, mountain Georgian dialects	Nogay, Azeri, Tat	Avar, Lezgi, Azeri, Kumyk, Nogay (northern Daghestan)	Arabic, Turkish, Persian

The Nakh-Daghestanian languages have developed in close contact over a considerable period of time, which prevented them from growing apart at a faster pace. Besides, they had contacts with the Turkic languages and marginally with Ossetic and Georgian dialects. On a lexical level, there exists a specific layer of “Daghestanisms”, i.e., areally diffused social and cultural terms, such as *lay//luk* ‘slave’, *šamxal*, *nucal*, *ucmi* (titles of rulers), *manyuš* ‘herald’, *č’assag* ‘date (fruit)’, *mahi* ‘ivory’, *bataya* ‘fishing’, etc. Given the processes of convergence observed in Daghestan, some authors deem it possible to discuss this region within the framework of the theory of “language union” (cf. Abdullaev 1991: 40).

10.1.3 *The Kartvelian contact area*

<i>number of languages in contact</i>	<i>core languages</i>	<i>peripheral languages</i>	<i>marginal languages</i>	<i>dominant languages or lingua francas</i>	<i>superstrate languages</i>
15+	Georgian, Svan, Megrelian Laz, Bats, Kistin Chechen, South Ossetic	Armenian, Udi, neighbouring Daghestanian languages/ dialects, Abkhaz, some forms of Turkish	Balkar, Abaza, Azeri, Kurdish, Neo-Aramaic	Georgian, Turkish (in areas of southern Georgia)	Arabic, Turkish, Persian

Like NWC and NEC, the Kartvelian languages have been in mutual contact for a long period of time. Because of the numerical size of its speakers, as well as the political, military and economical weight of Georgia, Georgian has become a dominant language in the area since the early Middle Ages, and it has remained as such until now. Of non-Kartvelian languages, Georgian was in contact with Armenian, Azeri, Ossetic, Nakh and Daghestanian languages, as well as with Turkish (in southern Georgia). Historically, Middle (Pahlavi) and Modern Persian were of major influence on Georgian, as well as Arabic and Ottoman Turkish. As to Megrelian, beside Georgian,

it had contacts only with Svan and Abkhaz, and historically with Byzantine Greek and Ottoman Turkish. Svan had contacts with Georgian, Megrelian, Abkhaz and Karachay, marginally with Kabardian and historically with Old Ossetic (Alanian) and probably also with some NEC idioms (see below).

## 11. Zones of influence

Parallel to sub-regional areal alliances, there are languages which have exerted considerable influence on the neighbouring idioms, forming zones of influence, which is most clearly seen in the spread of certain lexical items and sometimes structural traits. Some of the source languages are not members of the Caucasian Sprachbund. The zones of influence can be shown in the following tables.

**Table 11.** Zones of influence of the Caucasian languages

donor language	recipient languages
Adyghe	Ubykh, marginally Abkhaz
Kabardian	Abaza, Karachay-Balkar, Nogay, Ossetic, marginally Abkhaz and Ingush
Abkhaz	Ubykh, Megrelian
Ingush	Ossetic
Chechen	Ingush, some Kumyk groups, neighbouring Daghestanian and mountain Georgian dialects
Avar	Western Daghestan: Archi, Megeb Dargi, Andi, Tsez, Bezhta, Hunzib, Karata, Tindi, Botlikh, Akhvakh, Godoberi
Lezgi	Aghul, Rutul, south Tabasaran, Kuba Azeri
Dargi	historically: Archi, Aghul, Tabasaran
Lak	Aghul, Archi, Kubachi, Megeb and Chirakh Dargi
Bezhta	Hunzib
Tsez	Hinukh
Tsakhur	Avar dialects of Upper Samur, Zakatala and Belokan
Georgian	Megrelian, Svan, Laz, Bats, Tsez, Andi, Bezhta, Hinukh, Hunzib, Udi, Antsukh Avar, neighbouring Chechen dialects, Meskhetian Turkish, South (Kudar Iron) Ossetic, Tiflis and Artvin Armenian, marginally Abkhaz and North Ossetic
Megrelian	Southern Abkhaz, Svan, western Georgian dialects

Comments to the table:

- The Chechen influence on the isolated Kumyk enclaves is visible on both lexical and syntactic levels.
- Before the spread of Russian, in Western Daghestan the lingua franca was Avar, to be more precise, the Avar *koine* called 'bolmats' (*bolmac̣*'), spread over a considerable



- part of Daghestan and used by the speakers of Andi, Botlikh, Godoberi not only in communication with Avar, but also between themselves. It has been even reported that popular folklore genres of these peoples (songs, fairy-tales, fables) were often produced in Avar, not in their own languages (cf. Ibragimov 1991: 50).
- c. There has been a considerable Avar-Tsakhur interaction in the upper Samur region in Daghestan and in the Zakatala and Belokan regions of Azerbaijan, which resulted in many Tsakhur loans in Avar (cf. Dibirow 1991: 119).
- d. At present time the Lezgi influence has been weakened by the spread of Russian and Azeri, both serving as alternative lingua francas in this area. In the past the influence of Lezgi was considerable, as indicated by numerous innovational isoglosses uniting this particular grouping of the Lezgi languages (Alekseev 1991: 31).
- e. According to Alekseev (1991: 31), a number of isoglosses in Archi, Aghul and Tabasaran, which formerly were qualified as archaisms, can be reinterpreted as having a possible areal character, with Dargi as the probable center of spread.
- f. Among the languages of the Tsez subgroup (Tsez, Bezhta, Hinukh, Hunzib, etc.) the knowledge of Georgian was widespread among the men, as the trade, economic and other contacts connected them more closely with the neighbouring Georgia than with Avars and other Daghestanian peoples (Ibragimov 1991: 50). Georgian influence (predominantly lexical) is noticeable in both Udi and in the Antsukh dialect of Avar (Kiazimov and Musaev 1991: 112). Georgian exerted a major influence on the related Megrelian, Svan and to a somewhat less degree Laz. Its lexical borrowings are numerous in Armenian, Abkhaz, Ossetic, Bats, Meskhetian Turkish in Southern Georgia (whose speakers were deported by Stalin to Central Asia), the neighbouring Chechen and Daghestanian dialects. The development of hypotactic constructions and complementizers in Bats and Kistin (Pankisi) Chechen is also attributed to contact with Georgian. The Kartvelian influence on some Armenian dialects goes deeper than simple lexical borrowing, being noticeable in phonology and grammar as well (cf. Klimov 1986a: 198–199). As to the Tiflis and Artvin dialects of Armenian, the system of stops, according to Vogt (1988: 122), is completely assimilated to that of Georgian.

**Table 12.** Zones of influence of non-Caucasian languages

donor language	recipient languages
Azeri (Turkic: Oghuz)	NC: Lezgi, north Tabasaran, Kryz, Budukh, Khinalug, south Avar, Rutul, Tsakhur, Udi; SC: Ingilo Georgian; IE/Iranian: Tat, Talysh
Persian (IE: Iranian)	historically: Georgian, Armenian, Udi, some Daghestanian idioms
Ossetic (IE: Iranian)	NC: Ingush, Chechen; marginally – Kabardian, Abaza, Abkhaz; SC: Georgian, marginally Svan; Turkic: historically – Karachay-Balkar

(Continued)

Table 12. Continued

donor language	recipient languages
Turkish (Turkic: Oghuz)	NC: Abkhaz, Adyghe, Ubykh, Kabardian, Abaza; SC: Laz, Georgian, Megrelian, marginally Svan; IE: Armenian, Pontic and Tsalka-Alaverdy Greek
Armenian (IE: Armenian)	NC: Udi; SC: Georgian; Turkic: Azeri, Turkish dialects close to Armenian
Tat (IE: Iranian)	Budukh, Kryz, Udi
Pontic Greek (IE)	SC: Laz; Turkic: some north-eastern Turkish dialects

The peoples of South Daghestan used Azeri for communication not only with Azeris, but also among themselves and with the speakers of Tat (Ibragimov 1991: 51). In the Zakatala dialect of Avar there appeared, presumably under Azeri influence, elements of both class and personal agreement, as opposed to purely class conjugation in the North Caucasian Avar dialects (cf. Nurmagomedov 1991: 76). In Udi, Tsakhur (Sabunchin dialect), Lezgi (Kimil sub-dialect), Budukh, Kryz, Khinalug and the Ingilo Georgian spoken in Azerbaijan, beside many loanwords from Azeri, the evidence of the Azeri influence on their phonemic systems is the presence of the umlaut vowels *ä*, *ö*, *ü*, and in Lezgi, Udi, Budukh and Khinalug also the schwa (*ə*) phoneme (Magometov 1982: 179–180). In the Avar spoken in Azerbaijan (the Kakh, Zakatala and Belokan regions), the lateral obstruents, pharyngealization and labialization have been lost; in Zakatala Avar and Tsakhur there is also sporadic de-glottalization, and in Azerbaijanian Tsakhur also the loss of pharyngealization (cf. Dibirov and Čeerčiev 1998: 176–177). In Udi, Azerbaijani Tsakhur and Ingilo Georgian there is a tendency, under Azeri influence, towards vocalic harmony (Gadžieva 1979: 164–165; Magometov 1982: 180). In northern Tabasaran dialects the loss of the majority of nominal classes (except for two), the development of personal conjugation, the tendency towards vowel harmony and the presence of the umlaut vowel *ü* are attributed to the Azeri influence (Magometov 1977). Under the Azeri influence, there is tendency to delabialization in Tsakhur, Khinalug, Kryz, Budukh and those Lezgi dialects which neighbour on Azeri. There is also the substitution in Tsakhur of the phoneme *ʒ* by *z*, which, except for several words, happens also in Udi. Under Azeri influence, Udi has substituted its glottalized consonants by the voiceless non-aspirated ones. Besides, in the Mirzabeyli speech of the Nidzh dialect of Udi the ergative construction of the transitive verbs has been replaced by the nominative one, with the disappearance of the ergative case affix (cf. Gadžieva 1979: 165–167).

Table 13. Languages used for external communication in the Caucasus

Avar	by Archi, Andi, Akhvakh, Botlikh, Megeb Dargi, etc.
Lak	by Aghul, Archi, Kubachi Dargi, Megeb Dargi, Chirakh Dargi, etc.
Chechen	by Andi, Botlikh, Godoberi, etc.

(Continued)

Table 13. Continued

Georgian	by Megrelian, Svan, South Ossetic, partially Armenian, Bezhta, Hunzib, Tsez
Circassian	by Ubykh, Abaza, Karachay-Balkar, partially Ossetic
Azeri	by Akhvakh, Aghul, Lezgi, Tsakhur, Rutul, Tabasaran, Udi, Khinalug, Budukh, Kryz, Tat, partially Avar and Lak
Kumyk	by Avar, Lak, Dargi

Table 14. Languages used at present or in the past as lingua franca in the Caucasus

Azeri	in Southern Daghestan
Kumyk	in Northern Daghestan
Avar	in Western Daghestan
Nogay	in Northern Daghestan
Circassian	In Western Caucasus
Russian	across the Caucasus (since the second half of the 19th c.)

Until the beginning of the 19th century Turkic Kumyk, beside Avar and Azeri, served as one of the lingua francas in foothill and lowland Daghestan, whereas in Northern Daghestan this role was sometimes played by Nogay (cf. Ibragimov 1991: 50). There existed a tradition of sending 8–10 year-old Dargi, Avar and even Russian boys to Kumyk villages to live in families and learn the Kumyk language (Ibragimov 1991: 50–51). In general, seasonal migrant workers from Daghestan (both Daghestani- and Tat-speakers) usually knew one of the Turkic languages (Azeri or Kumyk, sometimes Nogay) in order to be able to communicate with the speakers of various languages of Daghestan. In the 16th–17th centuries Azeri started to be used as a *lingua franca* in southern Daghestan by such peoples as Budukh, Kryz, Khinalug, Udi, Tsakhur, Rutul, Tabasaran, Lezgi, partially Avar and Lak. In the Samur valley of Daghestan, until the 1960s, the Rutuls, Tsakhurs, Lezgis and Laks have used a local *Azeri koine* for mutual communication (cf. Džidalaev 1990: 8–9).

12. The dynamics of historical areal alliances

Sub-areal and local alignments have been shifting throughout history, contacts between certain languages becoming stronger or weaker, the donor-languages becoming the recipients and vice versa. Thus, in the Abkhaz-Megrelian alliance in the earlier epochs, approximately between the 7th to 12th centuries, it was Abkhaz which played a dominant role and served as a donor, while towards the present time it was Megrelian which was becoming increasingly influential over Abzhywa Abkhaz, especially over its Samyrzaqan variety. In the 7th–12th centuries Alanian (Old Ossetic) lexical elements have penetrated many neighbouring idioms: Chechen, Ingush, Abkhaz, Abaza,

Kabardian, Karachay-Balkar, Georgian, Svan, Daghestanian, marginally also Megrelian. After the destruction of the Alanian state by the Mongols (in the 13th c.), their habitat shrunk to several gorges in the Central Caucasus and direct links with the majority of the mentioned languages, except for Ingush, Georgian and Kabardian, had been severed. From a donor language, Ossetic became a recipient, borrowing many elements from Ingush, Kabardian and Georgian.

Some historical alliances ceased to exist due to the disappearance of one of the participants. Thus, old Ubykh-Adyghe and Ubykh-Abkhaz alliances, due to the total exodus of the speakers of Ubykh to Turkey, are now non-existent. The traditional Laz-Pontic Greek alliance ceased to exist due to a near complete disappearance of Greek from what is now north-eastern Turkey. The Laz-Georgian contact, though currently attested only within tiny Laz enclaves in Georgia (especially in Adzharia), was historically more prominent, in the light of the supposed Laz (and Megrelian) substrate in the Western Georgian Gurian, Imeretian and Adzharian dialects (cf. Megreldze 1938) and of the later Georgian-Laz contacts in north-eastern Asia Minor, before the Turkification of Anatolia and some time after. Besides, it has been suggested that in the past, approximately up to the 5th–6th centuries AD, Laz directly contacted with West Caucasian languages, now separated from Laz by Georgian and Megrelian. Apart from some lexical borrowings from West Caucasian (e.g., Laz *nusaya*, *nisaq'a* ~ Circ *nəsay*<sup>o</sup>, Ubykh *nəsay* 'bride', Laz *obyə* ~ Circ *aby*<sup>o</sup>*e*, Ubykh *abya* 'nest', etc.), there are a number of interesting structural borrowings as well, cf. Laz *k'ai oc'k'ert'u* ~ Abkh *bzəja dəjbon* 'he liked, loved him/her' ('to love' = 'to see well', the same model in Circassian and Ubykh).

**Table 15.** Some historical alliances in the Caucasus

historical alliances	dominant idiom(s)	period
Laz-Pontic Greek	Pontic Greek	? – early 20th c.
Laz-Georgian	Georgian	ca. 5th c. – 17th c.
Old Ossetic-	Old Ossetic	ca. 6th c. – 13th c.
Karachay-Balkar		
Caucasian Albanian-	Georgian, Armenian	ca. 4th c. – 11th c.
Georgian-Armenian		

### 13. Substrate and superstrate in the Caucasus

Quite a large number of words of Arabic, Iranian and Turkish origin constitutes a specific "oriental" superstrate layer common for all Caucasian idioms, as well as for the traditional Indo-European and Turkic languages of the Caucasus. These words started to appear in the Caucasus with the spread of Islam and the establishment of Arabic, Persian and later Crimean Tatar and Ottoman Turkish hegemony over parts of

the Caucasus. The relative share of this “oriental” lexicon is different in various languages, and many of such words have entered individual languages not through the direct contacts with Arabs, Persians or Turks, but via the neighbouring idioms. Given that a considerable number of these words, well adapted and regarded by the speakers as an integral part of their native lexicon, are spread across the Caucasus, we can speak about a common “oriental” superstrate in the languages of the Caucasian linguistic area.

Though processes of language shift did not occur in the Caucasus on a large scale, in certain cases this undoubtedly did take place. When languages cease being used by their speakers, they rarely disappear without trace. More often they leave their traces in the languages which replace them: in intonation and stress patterns, often also in morphological and syntactic traits; such traces can especially be visible in toponyms and in local geography- and culture-specific lexicon. The totality of these residual phenomena is referred to as substrate.

In the Caucasian context, specialists speak about the Nakh substrate in Ossetic, in the mountain dialects of Georgian and in sub-dialects of Terek Kumyk, the Alanian (Old Ossetic) substrate in Nakh and in Karachay-Balkar, the Udi substrate in the Ingilo dialect of Georgian (Klimov 1986a: 83, 187–188, 202), the Daghestanian substrate in some Kumyk and Azeri dialects, the Georgian substrate in Meskhetian Turkish and some north-eastern dialects of Anatolian Turkish, the Megrelo-Laz substrate in the Imeretian, Gurian and Adzhar dialects of Georgian, the Svan substrate in the Lechkhum dialect of Georgian, the Hurro-Urartian, and partially presumably Kartvelian (cf. Vogt 1988: 116–133) and (e.g., in the Karabakh/Artsakh area) NEC substrate in Armenian, etc.

In Georgia, the general dynamics of the expansion of East Kartvelian-speaking groups to the west can be seen in the replacement of Laz and Megrelian by Georgian in the modern provinces of Guria, Imeretia and Adzharia, as well as in the replacement of Svan by Georgian in the north-western province of Lechkhumi. This is indicated not only by the substrate Megrelo-Laz or Svan lexemes preserved in these dialects, but also by a significant layer of Megrelo-Laz toponymics in Guria, Imeretia and Adzharia (cf. Megreliдзе 1938; Kakabadze 1931) and Svan toponymics in the Lechkhumi province of Georgia (the very element *le-* in *lečxum-i* is a usual Svan prefix).

The idea of the NC substrate in Kartvelian was once rather popular in the Kartvelological literature.<sup>37</sup> There is indeed evidence that a part of the NEC-speaking population of Caucasian Albania (Gardabanians in the province of Gardabani, Herians in the province of Hereti, and Tsanarians in the province of Kakheti) underwent linguistic Georgianization by the late Middle Ages. The phonetic, grammatical and lexical peculiarities of the Ingilo dialect of Georgian (now in Azerbaijan) are assumed to be the result of the shift of Udi speakers to Georgian (cf. Klimov 1986a: 187–188). Besides, there are strong indications of a Nakh substrate in the mountainous Georgian

---

37. Cf. Džavaxišvili (1950: 247–248), Schmidt (1952: 10, 13), Čikobava (1977: 14–16).

dialects, such as Tush, Pshav, Mtiul, Khevsur and partially Kakhetian (cf. Uturgaidze 1966; Klimov 1986a: 184–186), as well as in the central Georgian province of Dvaleti, now populated by the speakers of (South) Ossetic and Georgian (cf. Gamrekeli 1958). The assimilation of the relic NEC idioms in Eastern Georgia can be explained by the economical, political and military hegemony of Georgians, strengthened by the consolidating role of the Georgian church. The other part of the Caucasian Albanian tribes has shifted either to Azeri, or to Armenian. Bats, Udi, Kryz, Budukh and Khinalug, which form isolated relic NEC islands in the Kartvelian- or Turkic-speaking areas, can be regarded as the last surviving vestiges of the indigenous NEC population of the South Caucasus.<sup>38</sup>

On the territory populated now by speakers of West Kartvelian Megrelian and Laz, some authors envisage traces of ancient West Caucasian substrate languages, which is reflected in the toponymics (e.g., the river Supsa in Western Georgia, supposedly containing the Adyghe formant *-psă* ‘water, river’), as well as in the Megrelian lexicon, morphology (e.g., the prolific system of preverbs, some of which were directly borrowed from Abkhaz) and syntax; a part of these features can be traced also in Laz (cf. Čikobava 1942; Hewitt 1992; Chirikba 1998). Vogt (1988: 466) regarded the ancient contact of the speakers of SC with the NWC (Circassian) population on the Black Sea coast, in the area between Trebizond (Trabzon) and Phasis (modern Rioni) in the period between the 7th and 4th centuries BC, as “undoubtedly” the source of innovation for Megrelo-Laz.

As to the fourth member of the Kartvelian family, Svan, some earlier authors even regarded it as a mixed language (cf. Marr 1916: 15; Deeters 1957: 12). According to Vogt (1988: 467), though Svan neatly preserves Kartvelian morphology and syntax, and less so its phonemic system, its vocabulary differs substantially from that of the sister-languages and can be the result of contact with the other Caucasian languages, in particular with Circassian. One of the examples of such contacts is the Svan pronoun *xedi* ‘which’, presumably a loan from the Circassian interrogative and relative pronoun *šet* ‘who’. It has also been assumed that the Svan ergative suffix *-m* might have been borrowed from Circassian. Furthermore, in Svan, as in Circassian, the ergative case form can be used as a basis for other cases (the so-called “principle of two stems”), a trait alien to other Kartvelian languages, but present both in Circassian and in the NEC languages, cf. Svan (Upper Bal) *mār-ēm-iš* ‘man-ERG-GEN = of the man’, Ad *c’əfə-m-č’e* ‘man-ERG-INSTR = by the man’ (cf. Čikobava 1977: 15). Another possible Circassian borrowing, noted already by H. Schuchardt, is the 3rd person verbal plural suffix *-x* in the forms like *leg-x* ‘they are standing’, which has no Kartvelian parallels, cf. the Circassian plural nominal and verbal suffix *-x̂(e)* (e.g., *ma-k<sup>o</sup>e-x̂* ‘they are going’).

---

38. The appearance of Kists, or Pankisi Chechens in Eastern Georgia is the result of a later (19th century) Chechen migration from the Central Caucasus.

Despite these and some other interesting morphological and lexical parallels between Svan and West Caucasian (specifically, Circassian), it is clear now that the early claims of a major Circassian influence on Svan morphology or of the mixed (Kartvelian-West Caucasian) character of Svan have been exaggerated. On the other hand, some authors now assert the presence in Svan of a lexical layer of the NEC origin, which can partially be attributed to Nakh, and partially to Daghestanian source. According to Fähnrich (1986: 32), who suggests a NEC substrate in Svan, the contact of Svan with the NEC languages must have occurred before the settlement of Ossetes in their present habitat in Georgia.

It should be noted that, indeed, some typological traits render Svan closer to Nakh and Daghestani (specifically, Ando-Tsez languages), rather than to the related Georgian and Megrelo-Laz and, for that matter, the contiguous Abkhaz or Circassian. These features include:

- a. in phonology – the elaborate system of vowels (simple, long and umlauted) and diphthongs, which sharply distinguishes Svan from both the rest of the Kartvelian languages and from NWC; interestingly, at least partially these features are found also in those northern Georgian dialects – Tush, Khevsur, etc. – which have undergone a substantial Nakh influence or are even thought to be based on a Nakh substrate;
- b. in morphology – the presence of the exclusive-inclusive distinction in personal pronouns, also lacking in Kartvelian and in the NWC languages, but present in Nakh and Daghestanian. The Svan decimal numeral system connects it directly with Ando-Tsez languages, whereas Georgian, Megrelo-Laz and NWC (with the exception of Kabardian) are strictly vigesimal.

Though the idea of the Circassian origin of the Svan ergative suffix is not popular in the modern Kartvelological literature, the declensional model of Svan according to the “principle of two stems” can well be areal, resulting from the influence of the NEC languages, in which the “second” (oblique) stem is based on the ergative or genitive. The presence of these similarities between Svan and Nakh-Daghestanian languages, which occupy geographical zones now remote from each other, can indicate the existence of an ancient Central-Caucasian union between Svan and some Nakh and Daghestanian dialects.

In Azerbaijan, the (North) Caucasian layer in Turkic Azeri is explained both by its contacts with the Caucasian languages and partially by the NC substrate in the northern Azeri dialects. The north-eastern part of Azerbaijan is still inhabited by the ethnically heterogeneous population, including, besides Azeri, also Iranian Tats and speakers of NEC languages, such as Lezgi, Avar, Tsakhur, Kryz, Budukh, Khinalug, Udi, etc. Such Daghestanian features as the presence of glottalized consonants, the lack of the possessive category, and the lack of vowel harmony are attested in the Kutkashen-Vartashen, Zakatala-Kakh, Kuba, Terekemy and Lower-Katrukhdialects of Azeri, conditioned by, respectively, Udi, Tsakhur and Lezgi substrates (cf. Džidaliev 1998: 125).



Similar non-Turkic features in phonetics, morphology and vocabulary are found also in some Kumyk dialects, specifically, in Kaytag, Lower Kumyk ("podgornyj") and partially in Buynaksk, where they are explained by the influence of Tabasaran, Lak and Avar substrates (Džidalaev 1998: 125; Mikailov 1954: 12–13). As to the Kaytag dialect, its non-Turkic phonological features include the presence of glottalized and geminate consonants, the affricates *ʒ* and *c*, the voiceless velar fricative *χ*, the pharyngealized *ā*, the lack of vowel harmony, etc. The possible substrate language in this case is supposed to have been the Khaydak dialect of Dargi. The other cases of the NEC substrate in Caucasian Turkic are: the Zakatala-Kakh dialects of Azeri, based on the Tsakhur and Avar substrates, the Kuba dialect of Azeri, based on Lezgi, the Lower-Katruk dialect of Azeri, based on Lak, the Tabasaran sub-dialects of Azeri, based on Tabasaran, and the Lower ("podgornyj") dialect of Kumyk, based on the Dargi and Avar substrates (Džidalaev 1987: 316).

In Armenia, which is currently nearly completely Armenian-speaking territory, the traces of ancient pre-Indo-European substrate languages (most importantly, Urartian) are found in toponyms and in the lexicon. Besides, some authors envisage also the presence of a Kartvelian substrate, which seems to be not East Kartvelian (i.e., Georgian), but West Kartvelian (i.e., Megrelo-Laz), which is indicated by the phonetic shape of the substrate words, which are closer to Megrelo-Laz (cf. Vogt 1988: 127–128). According to Pisowicz (1967), the substrate, which was supposedly close to the Caucasian languages, played a major role in consonant shift in Armenian.

Discussing the substrate phenomena in Armenian and Ossetic, Abaev (1995: 487–488, 498–499) thought that the absence of a morphologically marked accusative in both languages, the loss of grammatical gender and the development of postpositions are due to the common Caucasian influence or substrate. Cf. also K.-H. Schmidt (1972: 4) on the development of the common type of agglutinative declension in Armenian and Ossetic due to the areal pressure.

In general, though the study of substrate phenomena (including toponyms) in the Caucasus promises to bring about very interesting results, much is still to be done here, despite interesting work undertaken by some authors, cf. especially Abaev (1933; 1995), Megreliдзе (1938), Uturgaidze (1966), Fedorov (1983), Klimov (1986a; 1994), Vogt (1988), Fähnrich (1986; 1988), and Hewitt (1992).

#### 14. The position of Kartvelian among the Caucasian languages

Despite the fact that the Kartvelian languages manifest remarkable structural similarities with the North Caucasian languages on all linguistic levels, by some significant parameters they deviate from the canonical "Caucasian" type as attested in both North Caucasian groupings. Specialists even speak of a structural-typological split between NC and SC (cf. Klimov 1986a: 156). Indeed, when there are exceptions from the pan-Caucasian patterns, it is more often Kartvelian which deviates.



In phonology, though the SC languages, like NC, have glottalized consonants, they do not have labialized, palatalized, pharyngealized or strong consonants, which play important role in the North Caucasian languages and which are reconstructable (apart from palatalization) also for Proto-NC. The obstruent laterals, present in both NC groups and going back to Proto-NC, are not attested in modern SC languages, nor can they be reconstructed for Proto-Kartvelian. In general, the relative simplicity of the SC phonemic system renders it closer to such Indo-European languages of the area as Ossetic and Armenian, rather than to North Caucasian. As noted by G.V. Cereteli, "There is a full typological affinity between the phonological systems of Kartvelian, Armenian and Ossetic" (cf. Gamkrelidze and Mačavariani 1965: 46–47).

In morphology, there are no traces of nominal classes in SC, attested in both NC groups. The presence of relative pronouns in Kartvelian contrasts with their absence in NC. As to such proto-typical "Caucasian" morphosyntactic feature as ergativity, both NC groups are truly ergative, whereas SC, though displaying some ergative-like constructions is, according to Klimov and Alekseev (1980: 299), in essence nominative-active, rather than ergative (cf. also Harris 1990; Tuite 1999; but cf. Hewitt 1995). Passive diathesis is a normal phenomenon in Kartvelian, but not in NC.

In syntax, SC hypotactic structures are built as they are in IE, with finite verbs in both main and subordinate clauses, interconnected by complementizers. By contrast, in NC we normally have non-finite verbs in subordinate clauses, and no complementizers.

It has been noted (cf. Deeters 1957: 14–16; Charachidzé 1967: 62; Vogt 1988: 504) that in some important aspects Kartvelian shows a closer affinity with the IE structural type, than with the Caucasian one. Taking into account, in particular, the hypotactic constructions in Kartvelian, G. Schmidt (1952: 5) speaks about the "obvious Indoeuropeanization" of Kartvelian. Rogava (1977: 30) also thought that the development of hypotactic constructions in Kartvelian is due to the influence of IE languages. Gamkrelidze and Mačavariani (1965), on the basis of the analysis of Kartvelian resonants and ablaut phenomena, even proposed an Indo-European-Kartvelian Sprachbund. The same is suggested by Shimomiya (1978: 198) on the basis of IE-SC parallels in relative pronouns and in the structure of complex sentences, especially those containing relative clauses, which can be, according to him, ascribed to the influence of the IE syntax. Gamkrelidze and Ivanov (1984: 252–263, 871, 879) also suggested an ancient areal unity of these two families on the basis of the structural isomorphism of the Kartvelian and Indo-European consonant system, the presence of syllabic and non-syllabic resonants, the similar structure of root and affixal morphemes, the ablaut mechanisms, as well as the presence of the common cultural vocabulary.

Typologically, Kartvelian vacillates between NC and IE, sometimes sharing important features with the one, sometimes with the other. This particular intermediate position of Kartvelian may probably be regarded as an indication of the fact that it has evolved on a territory situated between the Proto-NC and Proto-IE habitats. This assumption, based on the structural traits of Kartvelian, can also be corroborated by the presence in Proto-Kartvelian of "cultural" vocabulary, connecting it, on the one hand,

with NC (cf. Nikolayev and Starostin 1984: 30)<sup>39</sup> and on the other hand, with IE (cf. Klimov 1986a: 171–202; 1994; Gamkrelidze and Ivanov 1984: 877–880). It has been assumed (cf. Vogt 1988: 505; Klimov 1986: 171–202) that before Kartvelian became a full-fledged member of the Caucasian “club” of languages, it underwent a substantial level of contact with Proto-Indo-European, or with some very old members of IE.<sup>40</sup> Deeters (1955: 26; 1957: 14, 16) regarded SC as a mixed type, standing between IE and NWC languages, a kind of an Indoeuropeanized West Caucasian, and was wondering whether this was the result of an IE superstrate in Kartvelian.

The complicated character of the evolution of Kartvelian can be demonstrated, for instance, by the analysis of its numerals. Thus, PK *\*sam-* ‘three’ (Geo *sam-*, Megr, Laz *sum-*, Svan *semi-*) might be brought into connection with Lak *šam-a*, Khin *pš<sup>o</sup>a* ‘three’, Tab *sum-* in *sum-č’ur*, Burshag Aghul *šin-* in *šin-c’ur* ‘thirty’ (‘three + ten’).<sup>41</sup> Another PK numeral, *\*xu(s<sub>1</sub>)t-* ‘five’, can be linked to such NC forms as Lak *x:u-*, Kubachi Dar *xu-*, Tab *xu-b*, Aghul *šafu-d*, Rut *xu-d*, Kryz *fī-d*, Archi *λ:ʷe-jt’u*, Khin *p<sup>h</sup>xu*, Bats *pxi*, Abkh *x<sup>o</sup>-ba*, Kab *t<sup>x</sup>ə* ‘five’.<sup>42</sup> The Kartvelian roots for ‘hundred’ (Geo *as-*, Megr-Laz *oš-*, Svan *ašir-*, *äšir-*) resemble such NC roots for ‘hundred’ as Lak *t:urš*, Dar *darš*, Lez *wiš*, Tab *warž*, Aghul *barš*, (Burshag) *warš*, Rut *wäš*, Tsakh *waš*, Archi *baš*, connected with the related forms in NWC (Abkh *s<sup>o</sup>ə*, Circ *ša*, Ub *s<sup>o</sup>a*).<sup>43</sup> On the other hand, the Kartvelian numerals *\*otxo-* ‘four’ and *\*eks<sub>1</sub>w-* ‘six’ are assumed to have been borrowed from IE *\*oktō(u)-* ‘eight’ (presumably, a dual from ‘four’) and *\*yek<sup>s</sup>-* ‘six’, respectively (cf. Klimov 1994: 59–62, 53). Finally, the two other Proto-Kartvelian numerals, *\*š(i)wid-* ‘seven’ and *\*arwa-* ‘eight’ are thought to have a Semitic origin (cf. Klimov 1994: 63). If we accept these comparisons, this would mean that Proto-Kartvelian was evolving in a habitat which had a close geographic contact and

39. It is possible that the NC and IE habitats were also in contact, as indicated by the NC “cultural” vocabulary of IE origin (cf. Nikolayev and Starostin 1984: 30–32). Cf. also the discussion of the structural-typological phonological parallels between NWC and IE in Colarusso (1981), who also suggests ancient areal contact, whereby the NWC habitat falls between PIE and PK (pp. 481, 496, 551). Kortlandt (1995: 94) speaks even of the possibility of IE-NWC Sprachbund, suggesting that the IE sound system originated as a result of strong NWC influence.

40. Against expectations, IE loans in Kartvelian in general do not show any specific Hittite (Anatolian) connections (cf. Klimov 1994: 15). According to Vogt (1988: 504), Kartvelian shows remarkable affinity with Greek and especially with Armenian.

41. Cf. Nikolayev and Starostin (1994: 978).

42. These connections were already proposed by Bork (1907: 23–25).

43. On the other hand, Klimov (1986: 191) notes the dependence of PK *\*as<sub>1</sub>ir-* ‘hundred’ on a Semitic source, cf. Akkadian *ešru* ‘ten’ (in Arabic *šašra*), although this is somewhat more remote semantically from both SC and NC. The comparison of the Kartvelian numeral with Semitic was proposed already by A. Trombetti (1902: 199). Blažek (1999: 85) suggests a possibility of Semitic origin of the NC root as well.

seemingly lively trade relations with the territories populated by the ancient speakers of NC and IE languages (and also probably of Semitic). Typologically, in the condition of a multilingual environment, numerals can often become the first victims of the influence of dominant, prestige, or economically important languages.<sup>44</sup>

Above we briefly discussed the possibility of a NC substrate in Kartvelian, more precisely, NEC in Georgian and Svan, and NWC in Megrelian and Laz. This would explain both remarkable similarities between the Kartvelian and North Caucasian languages, and the fundamental differences between them. Like in the case of Ossetic and Armenian, the process of “Caucasization” of Kartvelian could have been a result of combined effects of millennia-long contact and (partially) NC substrate. Even after Proto-Kartvelian split into three individual languages, the SC languages continued, as loanwords and toponymics of Georgia suggest, being in close contact with the separate NC languages. As to the structural and material parallels between SC and Proto-IE, these might have been the result of earlier contacts of Kartvelian, which preceded its contacts with NC. One can presume, that at some point the Proto-Kartvelian-IE ties were broken, due probably to the migration of Proto-Kartvelians towards the more northern Caucasian habitat, where the process of their “Caucasization” started, similar to what had happened with both Armenian and Ossetic, only much earlier and profoundly.

According to Gamkrelidze and Ivanov (1984: 880–881), the split of Common Kartvelian can be referred, based on a glottochronological analysis, to the beginning of the 2nd millennium B.C. The first wave of Kartvelian migrations from the original habitat in central and western parts of the Caucasus Minor to the west and north-west, towards the Kolkhida lowlands, resulted in the formation of Svan, which spread in western Transcaucasia and assimilated the local languages, probably of the West-Caucasian type, which partially could serve as a substrate for Svan. According to these authors, the next wave of the Kartvelian migration westwards, some nine centuries later, pushed the speakers of Svan to the northern mountains and these newly arrived tribes settled in the Kolkhida lowlands and the eastern Caucasian coast of the Black Sea, laying the basis of Megrelo-Laz. The Kartvelian dialects which remained within the confines of the western part of the Caucasus Minor were the ancestors of modern Georgian dialects, whose speakers moved in eastern and north-eastern direction and pushed the original languages of the East Caucasian type to the north. Already in historical times another wave of Georgian migration moved in the western direction, towards the Kolkhian lowlands, assimilating some Megrelo-Laz groups, splitting the Megrelo-Laz unity and causing the creation of the separate Megrelian and Laz languages.

---

44. As noted on this account by Gudjedjani and Palmaitis (1986: 23), “Numerals are weak in Kartvelian: nowadays the Svans often use Georgian numerals similarly as Georgians use Russian numerals”.

This scenario explains the spread of the Kartvelian dialects rather well from their original homeland in the Caucasus Minor, as well as the traces of the West Caucasian substrate in Megrelo-Laz and of the East Caucasian substrate in Georgian. As to Svan, containing the traces of both West Caucasian and East Caucasian influences, it seems that it has undergone a more complex evolution. The first, NWC, layer may go back chronologically to the period before the Svan migration to the north from the Kolkhian lowlands, when it was in contact with the local West Caucasian dialects, while the second, NEC, layer can be attributed to Svan's interaction with the speakers of NEC languages after migration to the northern mountains, in what is now Svanetia.

The combined effect of millennia-long contact and convergence with the geographically adjacent NC languages and partially the presence of the NC substrate in Kartvelian can thus be regarded as the most probable explanations for the structural similarities between the unrelated NC and SC languages, which form a genuine Sprachbund of linguistic families.<sup>45</sup>

## 15. Conclusions

The aim of the present paper was to summarize the evidence amassed in the specialized literature of numerous structural similarities between the two indigenous Caucasian families, North Caucasian and Kartvelian, as observed on various levels of their structure – phonological, morphological, partially syntactic, as well as in the semantic organization of their lexicon, in their cultural vocabulary and in phraseology. I argue that this evidence provides a solid basis for the assumption that the two unrelated Caucasian families have formed, over a period of several millennia of contact, a specific kind of union, which Trubetzkoy dubbed “the union of linguistic families”. Parallel to the pan-Caucasian Sprachbund, uniting the NC and NEC families, lower-level sub-regional “Sprachbünde” have also evolved, as well as local alliances between the individual languages in close contact.

Some of the features pertaining to the Caucasian language union diffused to the non-Caucasian languages of the area, in the first place, to Indo-European Ossetic and Armenian. These idioms, which have “Caucasian” traits in phonology, morphology, lexicon, phraseology and partially in syntax, can be regarded as peripheral members of the Caucasian Sprachbund. The Turkic languages of the Caucasus, i.e., Karachay-Balkar, Kumyk, Azeri, Nogay, which also share a number of Caucasian traits, are on the margin of the Caucasian linguistic area, as these traits are of a more superficial nature and involve mainly lexicon and phraseology. Additionally, some Caucasian features have “spilled over” to the contiguous languages of the area, such as Semitic Neo-Aramaic, Iranian Kurdish, Pontic Greek and those Anatolian Turkish dialects which have

---

45. Cf. a similar conclusion in Gabunia and Guzman Tirado (2002: 81–82).

strong Caucasian (Kartvelian) substrate. Finally, the most recent language spoken in the area, Russian, has remained outside of any areal alliance with the idioms constituting the Caucasian language union.

Though many of the pan-Caucasian features discussed here are found in other languages outside the Caucasus, the totality of both area-specific and the more general linguistic traits shared by the two Caucasian families form *a set which is unique to the Caucasus and is not found outside this particular area*. It is highly improbable that the two unrelated linguistic families, known to have existed in a close geographic proximity for millennia, could have developed so many common traits independently from each other.

Unlike some other cases of language unions, the main mechanism of the formation of the Caucasian Sprachbund was not the diffusion of certain linguistic traits from some centre to the other languages of the area, but convergence and substrate. More specifically, we are dealing here with the case of unilateral convergence: the Kartvelian languages developing in the direction of the NC structural type.

As to the structural similarities between the NEC and NWC languages, which form two discrete areas separated in the North by Indo-European Ossetic and in the south by Kartvelian, they can be explained as the manifestation of genetic inheritance and parallel development stimulated by tendencies inherent in their common ancestral idiom. There is only marginal interaction of one of the NWC languages, Kabardian, with the NEC Ingush language, which means that the structural similarities between the two NC groups cannot be explained by convergent development. Only a part of these similarities, namely, the cultural lexicon and phraseology, can be attributed to areal diffusion.

I would like to finish this paper by citing the statement made by Gerhard Deeters in his 1957 paper (p. 13): "All Caucasian languages display common features, so that in any case they, irrespective of their possible genetic relationship, can be regarded as forming a 'Sprachbund' ". Now we are in a better position, as I have tried to demonstrate in this paper, to substantiate this assertion by reliable data.

## References

- Abaev, V.I. 1933. Obščie èlementy v jazyke osetin, balkarcev i karačaevcev (Common elements in the languages of Ossetians, Balkars and Karachays). *Jazyk i myšlenie*, I. Leningrad.
- Abaev, V.I. 1958. *Istoriko-ètimologičeskij slovar' osetinskogo jazyka*. Tom I. A-K' (Historical-etymological dictionary of the Ossetian language. Vol. I. A-K'). Moskva-Leningrad: Akademija Nauk SSSR.
- Abaev, V.I. 1995. *Izbrannye trudy*. Tom II. *Obščee i sravnitel'noe jazykoznanie* (Selected works. Vol. II. General and comparative linguistics). Vladikavkaz: Ir.
- Abdokov, A.I. 1976. *K voprosu o genetičeskom rodstve abxazsko-adygskix i naxsko-dagestanskix jazykov* (On the question of genetic relationship of the Abkhazo-Adyghean and Nakh-Daghestanian languages). Nač'ik: El'brus.

- Abdokov, A.I. 1981. *Vvedenie v sravnitel'no-istoričeskiju morfologiju abxazsko-adygskix i naxsko-dagestanskix jazykov* (Introduction into comparative-historical morphology of the Abkhazo-Adyghean and North-Daghestanian languages). Načik: Kabardino-Balkarskij Gosudarstvennyj Universitet.
- Abdokov, A.I. 1983. *O zvukovyx i slovarnyx sootvetstvijax severokavkazskix jazykov* (On phonetic and lexical correspondences between the North Caucasian languages). Načik: El'brus.
- Abdullaev, I.X. 1991. Jazykovye kontakty i processy integracii v dagestanskix jazykax (Linguistic contacts and integration processes in the Daghestanian languages). *Problemy jazykovogo kontaktirovanija v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifičaja opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprelja. Tezisy dokladov*. Maxačkala, 40–41.
- Aikhenvald, A. 2003. Evidentiality in typological perspective. In Aikhenvald and Dixon 2003: 1–31.
- Aikhenvald, A. & Dixon, R.M.W. 2001. Introduction. In *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, A. Aikhenvald & R.M.W Dixon (eds.), 1–26. Oxford: OUP.
- Aikhenvald, A. & Dixon, R.M.W. (eds.) 2003. *Studies in evidentiality*. Amsterdam: John Benjamins.
- Alekseev, M.E. 1985. *Voprosy sravnitel'no-istoričeskoj grammatiki lezgijskix jazykov* (Problems of comparative-historical grammar of the Lezgian languages). Moskva: Nauka.
- Alekseev, M.E. 1991. Vnutrigruppovyje kontakty i sravnitel'no-istoričeskaja grammatika (Intra-group contacts and comparative-historical grammar). *Problemy jazykovogo kontaktirovanija v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifičaja opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprelja. Tezisy dokladov*. Maxačkala, 31.
- Baudouin de Courtenay, I.A. 1904. Jazykoznanie (Linguistics). *Enciklopedičeskij slovar'*, Vol. 31, F.A. Brokhaus i I.A. Efron (eds.).
- Baudouin de Courtenay, I.A. 1963. *Izbrannye raboty po obščemu jazykoznaniju* (Selected works on general linguistics). Moskva, t. II.
- Blažek, V. 1999. Numerals. *Comparative-etymological analysis and their implications*. (Saharan, Nubian, Egyptian, Berber, Kartvelian, Uralic, Altaic and Indo-European languages). Masarykova Univerzita v Brně.
- Bloomfield, L. 1933. *Language*. New York NY: Henry Holt.
- Bokarev, E.A. 1967. Xvaršinskij jazyk [govor s. Inxokari] (The Khvarshi language [The dialect of the Vil. Inkhokari]). *Jazyki narodov SSSR. Tom IV. Iberijsko-kavkazskie jazyki*, 421–435. Moskva: Nauka.
- Bork, F. 1907. *Beiträge zur kaukasischen Sprachwissenschaft*. Teil I. *Kaukasische Miscellen*. Königsberg.
- Catford, J.C. 1977. Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology* 6: 283–314.
- Cereteli, G.V. 1968. O jazykovom rodstve i jazykovyx sojuzax (On the linguistic kinship and language unions). *Voprosy Jazykoznanija*, No. 3.
- Cereteli, K.G. 1976. K voprosu o klassifikacii novoaramejskix dialektov (On the question of classification of Neo-Aramaic dialects). *Philologia Orientalis* IV. Tbilisi: Mecniereba, 224–232.
- Charachidzé, G. 1967. Position structurelle du géorgien parmi les langues caucasiqes. *Revue de l'École Nationale des Langues Orientales*. 4: 29–63. Paris: Presses Universitaires de France.
- Čikobava, A.S. 1942. Morfoložičeskie vstreči abxazskogo jazyka s kartvel'skimi jazykami (Morphological parallels between Abkhaz and the Kartvelian languages). *Izvestija Instituta jazyka, myšlenija i material'noj kul'tury*. Tbilisi.



- Čikobava, A.S. 1970. K voprosu ob otnošenii kartvel'skix jazykov k indoevropskim i sever-okavkazskim jazykam (On the question of relationship between the Kartvelian languages and Indo-European and North Caucasian languages). *Voprosy Jazykoznanija* 2.
- Čikobava, A.S. 1977. K voprosu o kontakte jazykov po materialam kartvel'skix bespismennyx jazykov (On the question of language contact on the material of unwritten Kartvelian languages). *Materialy pjatoj regional'noj naučnoj sessii po istoriko-sravnitel'nomu izučeniju iberijsko-kavkazskix jazykov*, 10–17. Ordžonikidze.
- Chirikba, V.A. 1996. *Common West Caucasian. The reconstruction of its phonological system and parts of its lexicon and morphology*. Leiden: Research School CNWS.
- Chirikba, V.A. 1998. K voprosu ob abxazskix zaimstvovanijax v megreľskom jazyke (On the question of Abkhaz loans in Megrel). *Voprosy Jazykoznanija* 4: 128–142.
- Colarusso, J. 1981. Typological parallels between Proto-Indo-European and North-West Caucasian languages. *Bono homini donum. Essays in historical linguistics. In memoriam of Alexander Kerns*, 475–557. Amsterdam.
- Deeters, G. 1931. Der abchasische Sprachbau. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Philol.-hist. Klasse*, 289–303. Berlin.
- Deeters, G. 1955. Gab es Nominalklassen in allen kaukasischen Sprachen? *Corolla Linguistica. Festschrift Ferdinand Sommer*, 26–33. Wiesbaden: Otto Harrassowitz.
- Deeters, G. 1957. Die Stellung der Kharthwelsprachen unter den kaukasischen Sprachen. *Bedi Kartlisa, Revue de Karthvélologie* 23: 12–16. Paris.
- Dibirov, I.A. 1991. Ob avarsko-caxurskix jazykovyx svjaziax [na materiale govorov] (On the Avar-Tsakhur linguistic contacts [On the material of dialects]). *Problemy jazykovogo kontaktirovanija v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifiķa opisaniķa kontaktnyx javlenij. Vsesojuznaja naučnaja konferenciķa. 10–13 apreķļa 1991 g. Tezisy dokladov*, 119. Maxačkala.
- Dibirov, I.A., Čerčiev M.Č. 1998. Process refonologizacii [Na materiale dagestanskix jazykov Alazanskoj doliny] (The process of re-phonologization [On the material of the Daghestanian languages of the Alazan valley]). *9<sup>th</sup> International Colloquium of Caucasian Studies. 15–19 June 1998. In Commemoration of Academician A.S. Chikobava Centenary. Abstracts of reports. Republic of Dagestan*, 175–176. Maxačkala.
- Dumézil, G. 1933. *Introduction à la grammaire comparée des langues caucasiennes du Nord*. Paris.
- Dumézil, G. 1937. Morphologie comparée et phonétique comparée. À propos des langues caucasiennes du nord. *Bulletin de la Société de Linguistique de Paris*, 38, fasc. 1, No. 112.
- Džavaxišvili, I. 1950. *Historical-ethnological problems of Georgia, the Caucasus and the Near East*. Tbilisi. (In Georgian).
- Džidalaev, N.S. 1987. K voprosu o dagestanskom substrate v kumyĥskom jazyķe [po dannym reči juŹnyx kumyĥov] (On the question of Daghestanian substrate in Kumyk [On the data of the speech of Southern Kumyks]). *Annual of Ibero-Caucasian Linguistics XIV*: 305–317. Tbilisi.
- Džidalaev, N.S. 1990. *Tjurkizmy v dagestanskix jazyķax. Opyt istoriko-ėtimologičeskogo analiza (Turkic loans in the Daghestanian languages. A sketch on the historical-etymological analysis)*. Moskva: Nauka.
- Džidalaev, N.S. 1998. Kavkazovedenie i problemy substrata (Caucasology and the problems of the substrate). *9<sup>th</sup> International Colloquium of Caucasian Studies. 15–19 June 1998. In commemoration of academician A.S. Chikobava centenary. Abstracts of reports*, 124–126. Republic of Dagestan, Maxačkala.

- Ėdel'man, Dž.I. 2007. Ešče raz o kavkazskom nazvanii pluga (Once more on the Caucasian designation of the plough). *Etimologija* 2003–2005, 327–333. Moskva: Nauka.
- Emeneau, M.B. 1980. *Language and linguistic area*. Selected and introduced by Anwar S. Dil. Stanford CA: Stanford University Press.
- Fähnrich, H. 1986. Gibt es ein nachisch-daghestanisches Substrat im Swanischen? *Georgica*, Heft 9, Jena-Tbilisi, 32.
- Fähnrich, H. 1988. Zu den nachisch-daghestanischen Lehnwörter im Swanischen. *Wiss. Ztschr. der Friedrich-Schiller-Universität Jena, Gesellschaftswiss. Reihe*, 37, H. 2, 117–121.
- Fedorov, J. 1983. *Istoričeskaja etnografija Severnogo Kavkaza (The historical ethnography of the Northern Caucasus)*. Moskva: Izdatel'stvo Moskovskogo Universiteta.
- Gabunia, Z., Guzman-Tirado, R. 2002. *Minoritarnye jazyki v sovremennom mire. Kavkazskie jazyki (The minority languages in the contemporary world. The Caucasian languages)*. Moskva: Rossijskaja Akademija Nauk.
- Gadžieva, N.Z. 1979. *Tjurkojazyčnye arealy Kavkaza*. Moskva: Nauka.
- Gamkrelidze, T.V. & Ivanov, V.V. 1984. *Indoevropskij jazyk i indoevropejcy. Rekonstrukcija i istoriko-tipologičeskij analiz prajazyka i protokul'tury (The Indo-European language and the Indo-Europeans. The reconstruction and historical-typological analysis of the proto-language and proto-culture)*. Tbilisi: Izdatel'stvo Tbilisskogo universiteta, Vol. I.
- Gamkrelidze, T. & Mačavariani, G. 1965. *Sistema sonantov i ablaut v kartvel'skix jazykax (The resonants system and ablaut in the Kartvelian languages)*. Tbilisi.
- Gamrekeli, V. 1958. Dvaletis t'op'onomik'is analizisatvis (On the analysis of place-names in Dvaleti). *Ist'oriis Inst'it'ut'is šromebi*, Vol. IV, Part I. Tbilisi.
- Greppin, J.A.C. 1981. Review of: Jazyki Azii i Afriki, III. Moscow, Nauka 1979. *Revue des Études Arméniennes* N.S. 15: 491–522.
- Grjunberg, A.L. 1997. Tatskij jazyk (The Tat language). *Jazyki mira. Iranske jazyki. I. Jugozapadnye iranske jazyki*. Moskva: Indrik, 141–154.
- Gudjedjani, C. & Palmaitis, M.L. 1986. Upper Svan: Grammar and Texts. – *Kalbotyra* XXXVII (4), Vilnius: Mokslas.
- Harris, A.C. 1990. Georgian: A language with active case marking. *Lingua* 80: 347–365.
- Hewitt, B.G. 1992. Languages in contact in N.W. Georgia: Fact or fiction? In *Caucasian perspectives*, George Hewitt (ed.), 244–258. Munich: Lincom.
- Hewitt, B.G. 1995. Georgian – Ergative, active, or what? In *Subject, voice and ergativity. Selected essays*, D.C. Bennett, T. Bynon & B.G. Hewitt (eds.), 202–217. London: University of London, School of Oriental and African Studies.
- Ibragimov, M.-R.A. 1991. Jazyki mežetničeskogo obščeniya narodov Dagestana v konce XIX-načale XX veka (The languages of inter-ethnic communication of the peoples of Daghestan at the end of the 19<sup>th</sup> – the beginning of the 20<sup>th</sup> century]. *Problemy jazykovogo kontaktirovanija v konkretnyx politničeskix regionax SSSR. Leksikografičeskaja specifika opisaniya kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 apreļa 1991 g. Tezisy dokladov*, 50–51. Maxačkala.
- Jakobson, R. 1985. O teorii fonologičeskix sojuzov meždu jazykami (On the theory of phonological unions between the languages). *Roman Jakobson. Izbrannye raboty*, 92–104. Moskva: Progress.
- Kakabadze, S.N. 1931. O megrel'skix nazvanijax v toponimike Imerii [Reziiume] (On Megrelian words in place-names of Imeri [A summary]). *Bjulleten' kavkazskogo istoriko-arxeologičeskogo Instituta* No. 8: 12.



- Keraševa, Z.I. 1977. Jazykovye kontakty i dialektnaja differenciacija (Language contacts and dialectal differentiation). *Materialy pjatoj regional'noj naučnoj sessii po istoriko-sravnitel'nomu izučeniju iberijsko-kavkazskix jazykov*, 32–39. Ordžonikidze.
- Kiazimov, I.B. & Musaev, M.M. 1991. Gruzinskie zaimstvovaniya v tureckom jazyke (The Georgian loanwords in Turkish). *Problemy jazykovogo kontaktirovaniya v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifika opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprlja 1991 g. Tezisy dokladov*, 112. Maxačkala.
- Klimov, G.A. 1964. *Ėtimologičeskij slovar' kartvel'skix jazykov* (The etymological dictionary of the Kartvelian languages). Moskva: Izdatel'stvo AN SSSR.
- Klimov, G.A. 1971. Review of: G. Dumézil. Documents Anatoliens sur les langues et les traditions du Caucase. IV. Récits Lazes (dialecte d'Archavi). Paris, 1967. *Ėtimologija* 1968: 257–259. Moskva.
- Klimov, G.A. 1978 (ed.). *Strukturnye obščnosti kavkazskix jazykov* (The structural affinities between the Caucasian languages). Moskva: Nauka.
- Klimov, G.A. 1986a. *Vvedenie v kavkazskoe jazykoznanie* (Introduction into Caucasian linguistics). Moskva: Nauka.
- Klimov, G.A. 1986b. Problems of Caucasian area linguistics. In *Aspects of language. Studies in honour of Mario Alinei*. Vol. I, *Geolinguistics*, N. Århammer et al. (eds.), 123–131. Amsterdam: Rodopi.
- Klimov, G.A. 1991. Nekotorye voprosy areal'nogo kavkazovedeniya (Some questions of the areal Caucasology). *Problemy jazykovogo kontaktirovaniya v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifika opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprlja 1991 g. Tezisy dokladov*, 8–9. Maxačkala.
- Klimov, G.A. 1994. *Drevnejšie indoevropeizmy kartvel'skix jazykov* (The most ancient Indo-European borrowings in the Kartvelian languages). Moskva: Nasledie.
- Klimov, G.A. & Alekseev, M.E. 1980. *Tipologija kavkazskix jazykov* (The typology of the Caucasian languages). Moskva: Nauka.
- Kortlandt, F. 1978. Notes on Armenian historical phonology II (The second consonant shift). *Studia Caucasica* 4: 9–16. Lisse: The Peter de Ridder Press.
- Kortlandt, F. 1995. General linguistics and Indo-European reconstruction. *RASK: Internationalt tidsskrift for sprog og kommunikation* 2: 91–109. Odense: Odense University Press.
- Kuipers, A.H. 1963. Caucasian. In *Current trends in linguistics*. Vol. 1, *Soviet and East European linguistics*, T.A. Sebeok (ed.), 315–344. The Hague: Mouton.
- Kumaxov, M.A. 1989. *Sravnitel'no-istoričeskaja grammatika adygskix [čerkeksskix] jazykov* (The comparative-historical grammar of the Adyghean [Circassian] Languages). Moskva: Nauka.
- Lewy, E. 1961. *Kleine Schriften*. Berlin: Akademie-Verlag.
- Mačavariani, G.I. 1966. K tipologičkoj xarakteristike obščekartvel'skogo jazyka-osnovy (Towards the typological characteristics of the common Kartvelian proto-language). *Voprosy Jazykoznanija* No. 1.
- Madieva, B. 1991. Regional'nye osobennosti dvujazyčija i mnogojazyčija v Dagestane [sociologičeskij aspekt] (The regional peculiarities of bilingualism and multi-lingualism in Daghestan [The sociological aspect]). *Problemy jazykovogo kontaktirovaniya v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifika opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprlja 1991 g. Tezisy dokladov*, 47–48. Maxačkala.
- Magometov, A.A. 1970. Ob odnom čislitel'nom v agul'skom jazyke (On one numeral in Agul). *Ibero-Caucasica* XVII: 167–175. Tbilisi.

- Magometov, A.A. 1977. Vlijanje jazykovyx kontaktov na tabasaranskij jazyk (The influence of language contacts on Tabasaran). *Materialy pjatoj regional'noj naučnoj sessii po istoriko-sravnitel'nomu izučeniju iberijsko-kavkazskix jazykov*, 40–46. Ordzhonikidze.
- Magometov, A.A. 1982. Vlijanje azerbajdžanskogo jazyka na vokalizm xinalugskogo jazyka (The Azeri influence on the vocalic system of Khinalug). *Annual of Ibero-Caucasian Linguistics* IX: 177–189. Tbilisi.
- Marr, N.Y. 1916. Kavkazovedenie i abxazskij jazyk (Caucasology and the Abkhazian language). *Žurnal ministerstva narodnogo prosvieščeniya*. LXIII, No. 5, otd. 4.
- Martinet, A. 1949. Phonology as functional phonetics. *Publications of the Philological Society* 15. London.
- Masica, C. 1976. *Defining a Linguistic Area. South Asia*. Chicago and London: The University of Chicago Press.
- Megrelidze, I.V. 1938. *Lazskij i megrel'skij sloi v gurijskom (Laz and Megrelian layers in Gurian)*. Moskva-Leningrad.
- Meillet, A. 1936. *Esquisse d'une grammaire comparée de l'arménien classique*. Vienne: Mékhitharistes.
- Melikišvili I.G. 1976. *Otnošenie markirovannosti v fonologii (On markedness in phonology)*. Tbilisi.
- Melikišvili I.G. 1983. K fonostatističeskoj karakteristike jazykov kavkazskogo areala, drevne-armjanskogo, obščekartvel'skogo i obščeeindoevropejskogo (Towards the phonostatistical characteristics of the languages of the Caucasus area, ancient Armenian, common Kartvelian and common Indo-European). *Genetičeskie, areal'nye i tipologičeskie svjazi jazykov Azii*, 218–229. Moskva: Nauka, Glavnaja redakcija vostočnoj literatury.
- Mikailov, Sh.I. 1954. Puti obrazovanija avarskogo literaturnogo jazyka (The paths of the formation of literary Avar). *Jazyki Dagestana*. Maxačkala.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago IL: The University of Chicago Press.
- Nikolayev, S.L. & Starostin, S.A. 1984. Severnokavkazskie jazyki i ix mesto sredi drugix jazykovyx semej Perednej Azii (The North Caucasian languages and their place among the linguistic families of the Middle East). *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka. Tezisy i doklady konferencii. Čast' 3. Jazykovaja situacija v Perednej Azii v X-IV tysjačelietii do n.e.*, 26–34. Moskva: Nauka, Glavnaja redakcija Vostočnoj Literatury.
- Nikolayev, S.L. & Starostin, S.A. 1994. *A North Caucasian etymological dictionary*. Moscow: Asterisk Publishers.
- Nurmagomedov, M.M. 1991. Ėlementy klassno-ličnogo spriaženija v avarskom jazyke kak rezul'tat vzaimodejstvija avarskogo i azerbajdžanskogo jazyka (The elements of class and personal conjugation in Avar as a result of interaction between Avar and Azeri). *Problemy jazykovogo kontaktirovanija v konkretnyx polietničeskix regionax SSSR. Leksikografičeskaja specifičeskaja opisanija kontaktnyx javlenij. Vsesojuznaja naučnaja konferencija. 10–13 aprelja 1991 g. Tezisy dokladov*, 76. Maxačkala.
- Pisani, V. 1956. *Ėtimologija: Istoriya-problemy-metod (Etymology: History-problems-method)*. Moskva.
- Pisani, V. 1959. *Saggi di linguistica storica*. Torino.
- Pisowicz, A. 1967. *Le développement du consonantisme arménien*. Wrocław.
- Pisowicz, A. 1997. Consonant shifts in Armenian dialects during the post-classical period revisited. *Armenian perspectives. 10th anniversary conference of the Association Internationale des Études Arméniennes*, N. Awde (ed.), 215–230. Richmond: Curzon.

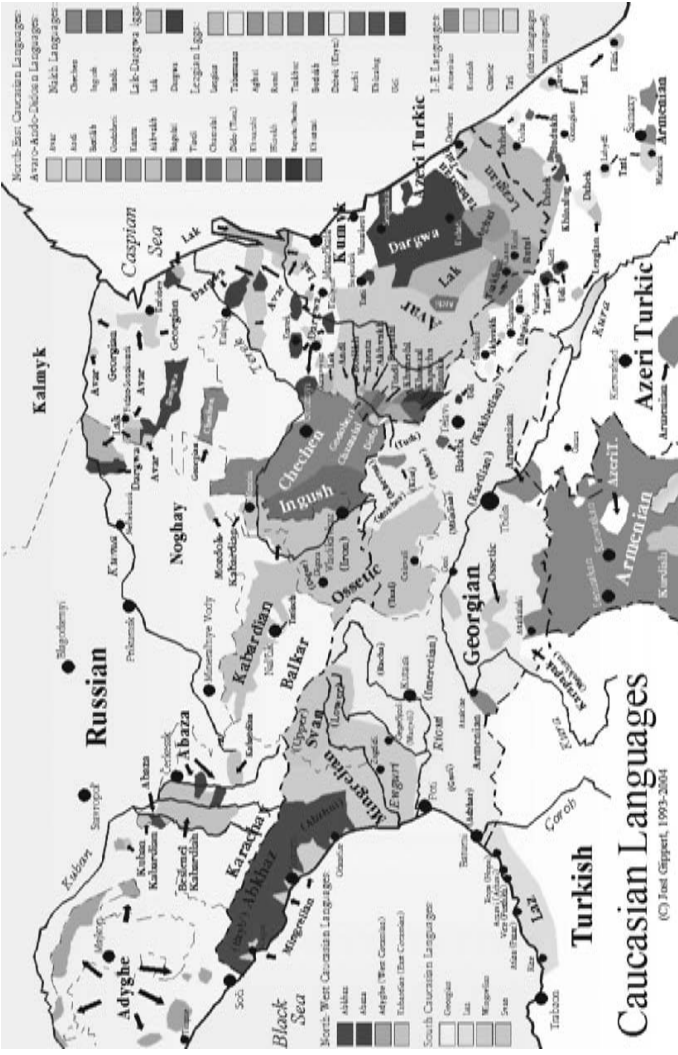
- Polák, V. 1950. L'état actuel des études linguistiques caucasiennes. *Archiv orientální* 18: 383–407.
- Provasi, E. 1981. Il superstrato lessicale Arabo nelle lingue caucasiche. In *Il Mondo Islamico tra Interazione e Acculturazione*, A. Bausani & B. Scarcia Amoretti (eds.), 53–66. Roma: Istituto di Studi Islamici. Università degli Studi di Roma.
- Rogava, G.V. 1977. K voprosu o vlijanii indoevropskix jazykov na grammatičeskiju strukturu kartvel'skix jazykov (On the question of the influence of Indo-European languages on the grammatical structure of the Kartvelian languages). *Materialy pjatoj regional'noj naučnoj sessii po istoriko-sravnitel'nomu izučeniju iberijsko-kavkazskix jazykov*, 26–31. Ordžonikidze.
- Rogava, G.V. 1987. K istorii atributivnoj sintagmy v iberijsko-kavkazskix jazykax (On the history of the attributive syntagma in the Ibero-Caucasian languages). *Annual of Ibero-Caucasian Linguistics* XIV: 9–16. Tbilisi.
- Schmidt, G. 1952. Über Aufgaben und Methoden der Kaukasologie. *Studia Orientalia, Edidit Societas Orientalis Fennica*. Bd. XVII, No. 4: 3–15. Helsinki.
- Schmidt, K.-H. 1972. Problemy genetičeskoj i tipologičeskoj rekonstruktsii kavkazskix jazykov (The problems of genetic and typological reconstruction of the Caucasian languages). *Voprosy Jazykoznanija* 4: 14–25.
- Shimomiya, T. 1978. *Zur Typologie des Georgischen [Verglichen mit dem Indogermanischen]. Mit einem Exkurs zur Sprachbundtheorie* [Gakushuin-Forschungsreihe 4]. Tokio: Gakushuin.
- Starostin, S.A. 1999. The problem of genetic relationship and classification of Caucasian languages: Basic vocabulary. In *Studies in Caucasian linguistics. Selected papers of the eight Caucasian colloquium*, H. van den Berg (ed.), 79–94. Leiden: Research School of Asian, African and Amerindian Studies (CNWS).
- Šaradzenidze, T. 1983. Izučenie konvergencii jazykov i jazykovye sojuzy (The study of the convergence of languages and the linguistic unions). *Genetičeskie, areal'nye i tipologičeskie svjazi jazykov Azii*, 69–79. Moskva: Nauka, Glavnaja redaksia vostočnoj literatury.
- Testelec, J.G. 1992. Ob odnom tipe reduplicirovannyx osnov v kartvel'skix jazykax (On one type of reduplicated stems in the Kartvelian languages). *Etimologija. 1990–1991*, 160–163. Moskva.
- Trombetti, A. 1902. Delle relazioni delle lingue caucasiche con le lingue camitosemitiche e con altri gruppi linguistici, I. *Giornale della Società Asiatica Italiana* 15: 177–201.
- Trubetzkoy, N.S. 1923. Vavilonskaja bašnja i smešenie jazykov (The tower of Babel and the merging of language). *Evrasijskij vremennik. Neperiodičeskoe izdanie pod red. P. Savickogo, P.P. Suvčinskogo i kn. Trubeckogo*. Kniga tretja, 107–124. Berlin. (Reprinted in Trubetzkoy 1995: 327–338).
- Trubetzkoy, N.S. 1987. Mysli ob indoevropskoj probleme (Thoughts on the Indo-European problem. N.S. Trubetzkoy. *Izbrannye trudy po filologii*, 44–59. Moskva: Progress.
- Trubetzkoy, N.S. 1995. *Istorija. Kul'tura. Jazyk (History. Culture. Language)*. Moskva: Progress, Univers.
- Trubetzkoy, N.S. 1922. Les consonnes latérales des langues Caucasiennes-Septentrionales. *Bulletin de la Société de Linguistique*, T. 23, No. 72 : 184–204.
- Trubetzkoy, N.S. 1928. Proposition 16. *Actes de 1er Congrès international de linguistes*, 17–18. Leiden: A.W. Sijthoff's Uitgevermaatschappij.
- Trubetzkoy, N.S. 1930. Nordkaukasische Wortgleichungen. *Wiener Zeitschrift für die Kunde des Morgenlandes* 37(1-2): 76–92. Wien.
- Trubetzkoy, N.S. 1975. *Letters and notes, prepared for publication by R. Jakobson*. The Hague: Mouton.

- Tuite, K. 1999. The myth of the Caucasian sprachbund: The case of ergativity. *Lingua* 108: 1–29.
- Turčaninov, G., Cagov, M. 1949. *Grammatika kabardinskogo jazyka. I* (*The grammar of the Kabardian language I*). Moskva-Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Uturgaidze, T. 1966. *Kartuli enis mtis k'ilota zogi tavisebureba* (Some characteristics of the mountain dialects of Georgian). Tbilisi.
- Vogt, H. 1942. La parenté des langues caucasiennes [Un aperçu général]. *Norsk Tidsskrift for Sprogvidenskap* 12. Oslo.
- Vogt, H. 1988. *Linguistique caucasienne et arménienne*. *Studia Caucasologica II* [Serie B: Skrifter 76], E. Hovdhaugen & F. Thordarson (eds.). Oslo: Norwegian University Press/The Institute for Comparative Research in Human Culture.
- Xajdakov, S.M. 1967. Arčinskij jazyk (The Archi language). *Jazyki narodov SSSR. Tom IV. Iberijsko-kavkazskie jazyki*, 608–626. Moskva: Nauka.

## Abbreviations

Abkh	Abkhaz	Lez	Lezgi
Ad	Adyghe	LOC	Locative
AOR	Aorist	Megr	Megrelian
Arm	Armenian	NEC	North-East Caucasian
ART	Article	NEG	Negative
Bagv	Bagvala	NWC	North-West Caucasian
Balk	Balkar	Oss	Ossetic
Bezh	Bezhta	PASS	Passive
Botl	Botlikh	PAvar	Proto-Avar
Bud	Budukh	PAvar-Andi	Proto-Avar-Andi
Cham	Chamala	PCirc	Proto-Circassian
Cech	Chechen	PDar	Proto-Dargi
Circ	Circassian (Adyghe and Kabardian)	PHunz-Bezh	Proto-Hunzib-Bezhta
Dar	Dargi	PIE	Proto-Indo-European
DAT	Dative	PK	Proto-Kartvelian
DYN	Dynamic	PL	plural
EC	East Caucasian	PLez	Proto-Lezgi
FIN	finite	PN	Proto-Nakh
FUT1	Future 1	POT	Potential
GEN	genitive	PRES	Present
Geo	Georgian	PTsez	Proto-Tsez
Hin	Hinukh	PTsez-Khvar	Proto-Tsez-Khvarshi
Hunz	Hunzib	Rut	Rutul
IE	Indo-European	SC	South Caucasian
Ing	Ingush	SG	singular
		SO	Klimov 1978

INSTR	Instrumental	Tab	Tabasaran
IOV	Indirect Object Version	TR	transitive
ITR	intransitive	Tsakh	Tsakhur
Kab	Kabardian	Tu	Turkish
Kar-Balk	Karachay-Balkar	Ub	Ubykh
Khin	Khinalug		
Khvar	Khvarshi		



Linguistic Map of the Caucasus (© J. Gippert).



# East Nusantara as a linguistic area<sup>1</sup>

Marian Klamer

Leiden University

Ger Reesink

Leiden University & Radboud University Nijmegen

Miriam van Staden

ACLC, University of Amsterdam

In this paper we consider how Eastern Indonesia may be treated as a linguistic area. We propose five defining linguistic features and we discuss their occurrence in some 40 Austronesian (AN) and non-Austronesian (NAN) languages of South Sulawesi, Flores, Sumba, Timor, Alor and Pantar, the Moluccas, Halmahera, the Bird's Head, and the Cenderawasih Bay. We propose that of these five areal features, three are Papuan features that have diffused into the Austronesian languages, while two Austronesian features have diffused into the Papuan languages. These Papuan features are: (1) possessor-possessum order in adnominal possession, (2) overt marking of the distinction alienable vs. inalienable possession, and (3) clause-final negation. While these features are not generally found in Austronesian, we will demonstrate that they occur in many Austronesian languages in East Nusantara and around the Bird's Head, as well as in the Papuan languages of this area. The Austronesian features are: (4) SVO as primary constituent order, and (5) an inclusive/exclusive opposition in the pronominal paradigm. These features are not found in Papuan languages in general, yet they are attested in both the Papuan and the Austronesian languages of East Nusantara, as we will demonstrate. Although the features do not all converge on the same isoglosses, together they define a linguistic area: East Nusantara. This area has Halmahera and the Bird's Head as its core, and radiates outwards to include the Moluccas and Alor/Pantar first, followed by the island Timor.

## 1. Introduction

Languages can be linguistic isolates, but they are seldom spoken in splendid isolation. Most groups of people have or have had extensive contact with speakers of different languages. Multilingualism is the norm rather than the exception and when groups migrate,

---

1. Research for this project was financed through two grants from the Netherlands Organisation for Scientific Research (NWO) and a grant from the European Science Foundation (ESFO). We wish to thank Pieter Muysken for his helpful comments and suggestions.



minge and split up, it is not surprising to find this reflected in the languages. Languages may change spontaneously or they may die entirely by themselves, but it is now a widely accepted view (Dixon, 1997) that the foremost source of language change and language death is through contact with other languages. In the description of spontaneous changes and inventions, genealogical relations between languages are of crucial importance but in contact-induced change we must also know how speakers of possibly unrelated languages interacted, and where they interacted. An area of interaction may be described as a linguistic area. Evidence of a shared history is used to delineate an area. Cultural commonalities between groups, state formation, or genealogical affiliation between languages are all evidence of historical links between groups, either directly or indirectly, for instance, through a common coloniser or in a chain relationship so that one group is in contact with two other groups that are not in contact with each other. Bio-genetic evidence linking groups may also be used. Groups of bio-genetically related people may be traced, and the area may be defined by migration patterns.<sup>2</sup>

But the notion of a linguistic area may also be approached from the ‘linguistics’ end as any area that is the focus of linguistic interest. The area may then be defined purely in topographical or geographical terms by stipulating a set of coordinates. Such a characterisation of an area may be relevant for, say, biologists studying linguistic diversity, or for linguists working on languages for which genealogical classification is highly problematic. In addition, an area can be defined on the basis of evidence from the languages themselves. For instance, a single area could be one in which a contact language is shared, in which a particular linguistic feature occurs, or in which the languages are typologically similar, even when they are genealogically unrelated. This is the more familiar approach in linguistics: “The term *linguistic area* refers to a geographical area in which, due to borrowing and language contact, languages of a region come to share certain structural features” (Campbell 1998: 299–300).

Typically, the various possible ways to delineate a linguistic area reinforce each other: linguists’ attention may be drawn to one particular feature, and this leads them to mark off a particular area based on geographical or topographical cues in which to look for other similarities. Generally speaking, any area defined on the basis of linguistic characteristics presupposes a shared history and contact between the speakers and the languages. In this paper we have defined the area East Nusantara, comprising of the easternmost part of insular South East Asia and Western New Guinea (cf. section 2.1 below), primarily

---

2. Genetic affiliation of peoples can, of course, by no means be conflated with genealogical affiliation of languages. In particular in the part of the world that we are interested in, it is very much an open question whether speakers of Austronesian or Papuan languages show differences in their genetic make-up. It is quite likely that intermarriage may have blurred the genetic differences between groups, or that language shift and language contact have led to a situation in which Austronesian languages are spoken by Papuan people. A simple example of this situation is of course the post-colonial world, in which Spanish is spoken by Argentinians of various descents, Dutch by African and Indian groups in Surinam, or English in India.

on the basis of a number of linguistic features that co-occur in genealogically distinct languages in this area but not generally outside it. We then examine possible foundations for contact induced change in a shared history, where evidence for genealogical relations between languages weak or absent.

Our hypothesis is that certain Austronesian languages in East Nusantara have absorbed Papuan features as the result of a shift process (Thomason 2001: 143). It is very likely that in various places the original 'Papuan speaking' populations were confronted with smaller but more powerful groups of 'Austronesian speaking' invaders. The indigenous peoples learned the Austronesian language imperfectly, keeping some of their 'Papuan routines.' Through intermarriage the two groups merged over time to become one homogeneous population, speaking the new variety of the Austronesian language including some Papuanisms. Such a scenario may account for the linguistic situation on those islands in the Moluccas where today only Austronesian languages are spoken, such as Buru and Banda.

At the same time, there are also instances where it is likely that Austronesian speakers incorporated Papuan features into their language as a result of contact with non-Austronesian speakers. An example is Alorese, the only indigenous Austronesian language spoken on Alor and Pantar, which has adopted some features from the various mutually related non-Austronesian languages that surround it Klammer (2007). Alor and Pantar are examples of regions outside Papua where non-Austronesian speaking populations persevered. This is also the case in Central Timor and North Halmahera where the people speaking non-Austronesian languages came to be surrounded by speakers of Austronesian languages. Some of the non-Austronesian languages adopted Austronesian features, such as the morphological distinction between inclusive and exclusive first person plural, and SVO constituent order.

These contact scenarios have interesting implications for our understanding of this linguistic area, because they may explain certain striking typological features found in the languages of Eastern Indonesia in particular. Himmelman (2005: 112ff) proposes two major typological groups on the basis of his typological research into the non-Oceanic Austronesian languages: 'symmetrical voice' languages and 'preposed possessor'<sup>3</sup> languages. Table 1 contrasts seven features on which these two language types differ. The 'symmetrical voice' languages include the Philippine-type languages and Western Indonesian-type languages, while the 'preposed possessor languages' include the Austronesian languages of Timor, the Moluccas and West Papua as well as the Malay varieties spoken in this area. In other words, the latter group coincides roughly with the languages of the area discussed in the present paper.

---

3. The preposed-possessor criterion refers to the most common or unmarked order found in possessive constructions. This means that it is not a requirement that all possessive constructions in a preposed possessor language show the order possessor-possessum, and conversely, in non-preposed possessor languages a possessor-possessum order may be allowed as well.

**Table 1.** Two major typological groups in the non-Oceanic Austronesian languages (from Himmelmann, 2005: 175)

Symmetrical voice languages	Preposed possessor languages
Symmetrical voice alternations	No or asymmetrical voice alternations
Postposed possessor in adnominal constructions	Preposed possessor in adnominal constructions
No morphosyntactic distinction between alienably/inalienably Possessed items	Morphosyntactic distinction between alienably/inalienably possessed items
Person marking only sporadically attested	Person marking prefixes or proclitics for S/A arguments
Numerals/quantifiers precede head	Numerals/quantifiers follow head
Negators in pre-predicate position	Clause-final negators
V-initial or SVX	V-second or –final

In section 4 below we argue that at least three of the characteristics of the ‘preposed possessor’ language type are, in fact, the result of diffusion from the non-Austronesian languages in the area. The features discussed are the preposed possessor pattern itself (sections 4.1.2 and 4.1.3), the morphosyntactic distinction between alienably/inalienably possessed items (section 4.1.1), and the presence of clause final, or post predicate, negators (section 4.2). Primary constituent order is discussed in 3.1 as a property of East Nusantara languages that has diffused from Austronesian into Papuan, just like the inclusive/exclusive distinction discussed in section 3.2. Of the other features, the absence of symmetrical voice alternations and the presence of person marking on the verb constitute an independent development, while noun phrase internal order is again an unrelated feature.<sup>4</sup> However, before we turn to a discussion of the linguistic features, we first present a geographical outline of the East Nusantara area, as well as a description of what is known about its history, followed by a sketch of the linguistic situation in the area.

2. East Nusantara

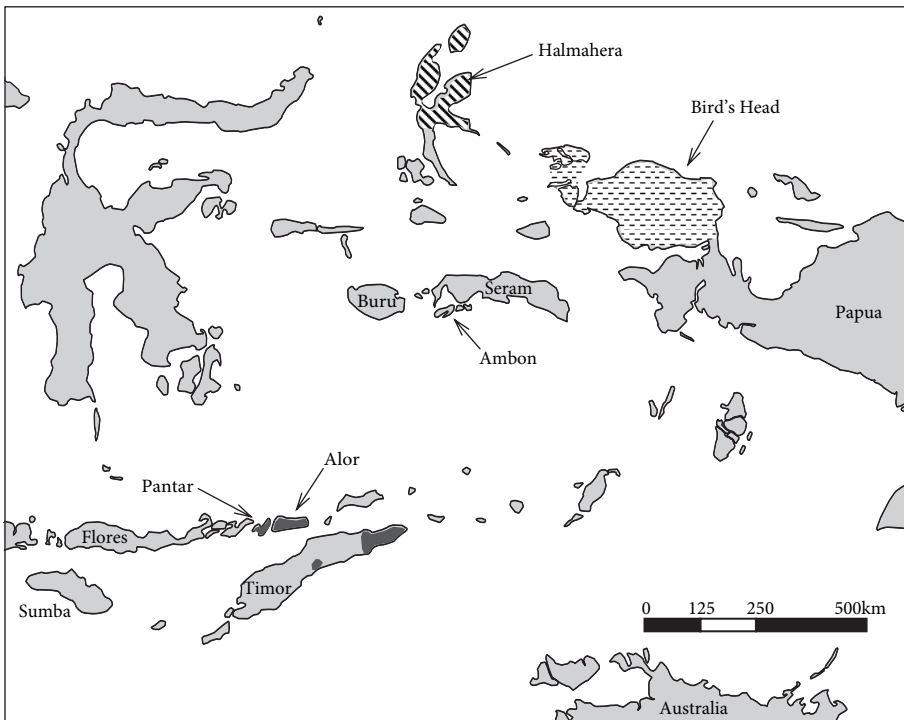
In this section, we first give a geographical outline of the area that we have labelled ‘East Nusantara’ (2.1). This is followed by a description of its common history and the origins of language contact (2.2). In section 2.3, we describe the overall linguistic

4. We have not investigated in detail these other features, but we know that asymmetrical voice alternations generally do not occur in the Papuan languages of the area. However, person marking on the verb and the placement of numerals/quantifiers after the noun are commonly found, so that diffusion from Papuan languages may account for their presence in the Austronesian languages.

situation of East Nusantara, and discuss the general features of the Austronesian and Papuan language families spoken in this area.

## 2.1 Geographical outline

The area of interest for the purposes of this paper we have labelled 'East Nusantara'. 'Nusantara' is Malay for 'islands in between', i.e., the Indo-Malaysian archipelago. East Nusantara comprises the islands of eastern Indonesia and East Timor: Halmahera, the Moluccas, Flores, Sumba, Sumbawa, Timor and Alor and Pantar. The Bird's Head of Papua belongs, strictly speaking, to mainland New Guinea and not to Nusantara (see Map 1). However, in this paper it is considered part of the linguistic area East Nusantara.



**Map 1.** The East Nusantara area. Papuan languages are spoken in Papua and the marked areas, Austronesian languages elsewhere.

At the outset we emphasise that the boundaries of the area are by no means clear-cut. There is clear evidence that the inhabitants of East Nusantara travelled to places outside the area, and there are genealogical relations between languages of this area and languages outside it. Especially parts of Sulawesi and New Guinea, not included at present, may have to be incorporated later. The geographical and historical centre of East Nusantara is the Moluccas, including Halmahera.

## 2.2 History

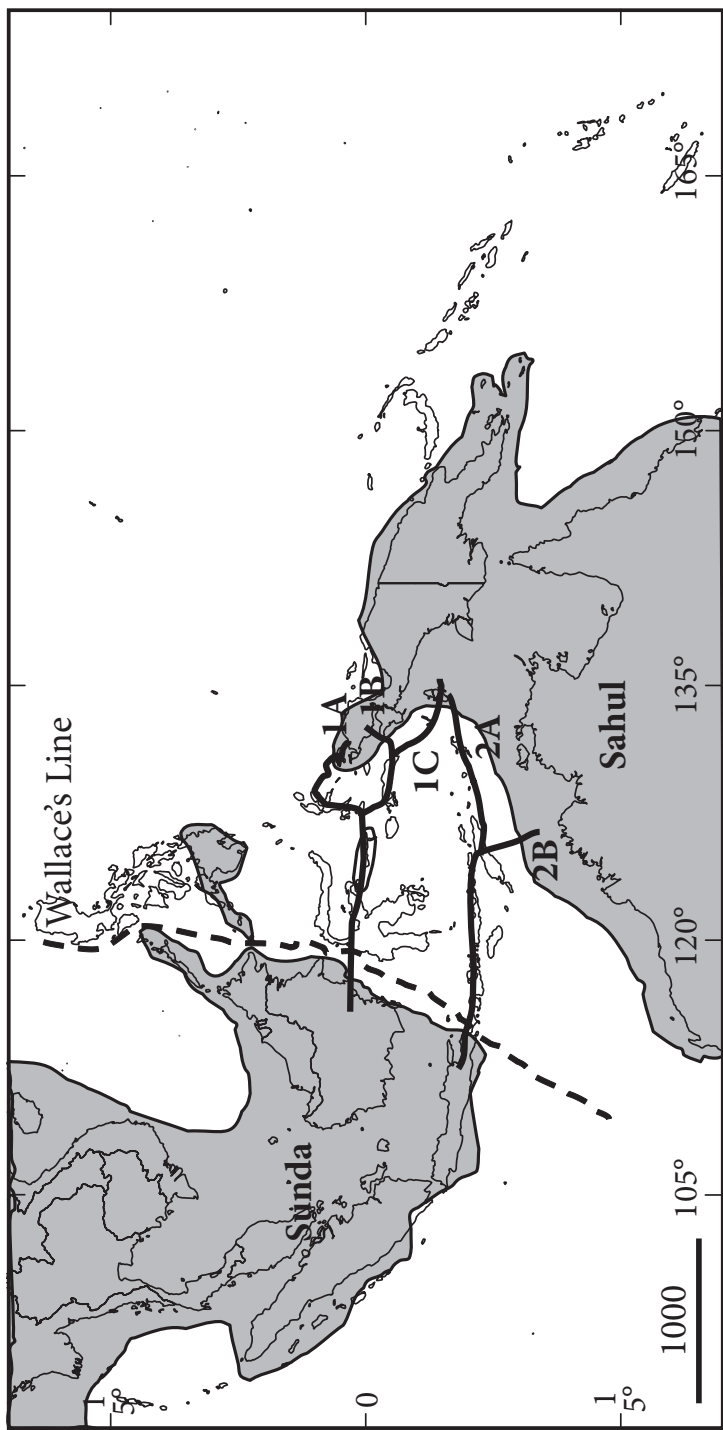
### 2.2.1 Early migrations

During the Pleistocene period, which lasted until approximately 9,000 BC, the land-masses of Australia and New Guinea were joined in a single continent called Sahul (see Map 2). Although the date for initial occupation of Sahul is still unresolved (Veth et al. 1998:162), it is generally agreed that the first human occupation was not later than 40,000 B.P., but possibly going back 50,000 years. These early colonists from Southeast Asia must have had the boat (or raft) technology that enabled them to cross the deep-water channels of the so-called Wallace line (and other channels) to reach the Moluccan islands and New Guinea, and eventually to travel as far as the Bismarck Archipelago and the Solomon Islands. The archaeological record contains dates of human settlement at various locations of more than 30,000 years ago from some Moluccan islands (i.e., Halmahera and Morotai; Bellwood 1998) and 26,000 years from the Bird's Head Peninsula (Pasveer 2003). It seems reasonable to assume, therefore, that ancestors of present-day speakers of Papuan languages had been present in the East Nusantara region for many millennia before the Austronesians arrived.

Birdsell (1977) hypothesizes that Sahul was populated by at least three groups of different people at times that the sea levels allowed relative easy crossing of the water divisions between Sunda and Sahul. He outlines two main routes from Sunda to Sahul, with different branches near the terminal points indicated on Map 2:

- (1) A northern route from Kalimantan through Sulawesi with three final alternatives:
  - (1a) From Sula via Obi to Halmahera and across to Waigeo (one of the Raja Ampat islands) with a landing on the Bird's Head;
  - (1b) From Sula via Buru and Seram to Misool (southern island of the Raja Ampat) as part of the Sahul shelf;
  - (1c) As (1b), but from Seram in south-eastern direction via smaller islands, such as Kai, with a landing point at the Aru islands (as part of Sahul).
- (2) A southern route from Sunda shelf through Bali and the Lesser Sunda islands to Timor with two final alternatives:
  - (2a) From Timor/Roti via Leti and smaller islands to Tanimbar with a landing on Aru;
  - (2b) From Timor/Roti directly south to the Sahul shelf.

Birdsell suggests three migration waves, approximately 50,000, 20,000 and 15,000 years ago, respectively, which could have followed some or all of the proposed routes, with (1b) and (2b) being the most attractive at 50,000 and 20,000 years ago, due to the low calibrated sea levels. The first group of immigrants, Birdsell suggests, would have been relatives of the negrito people of the Andaman islands, the *Semang*, also referred to as *orang asli*, of present inland Malaysia, and the *Agtas* of the Philippines. They are characterized as small in stature (pygmy-like) with tightly, spirally curled hair. The two later populations had a different phenotype, and are assumed to have



Map 2. Possible migration routes through Wallacea, after Birdsell, 1977

absorbed, extirpated or driven away the first settlers, some of which presumably ended up in the northern region of Sahul, present-day New Guinea. To what extent Birdsell's hypothesis can be proven for the present-day Australian continent is a matter of further research. The scenarios that make up his proposal all include the possibility of ancient populations in Wallacea (that includes our East Nusantara area) with connections to Papuan populations in New Guinea.

The little human genetic information that is available for this area shows that there is indeed an old connection between Timor and Halmahera regions and the Papuan mainland New Guinea. For example, Capelli et al. (2001) report a study which included a population sample from the Bird's Head. Its results identified a haplogroup of the Y chromosome that is mainly restricted to Melanesia. Outside Melanesia it has a high frequency in Alor, which Capelli et al. (2001) relate to the presence of Papuan languages in the region of Timor and the smaller islands of Alor and Pantar. The same study (Capelli et al. 2001:435) also reports deep splits between mainland Southeast Asian, insular Southeast Asian, and Melanesian Y chromosomes – with Polynesians closely associated with the Melanesian clusters, suggesting that this split may have happened at 12,000 B.P. or earlier. Because this haplogroup is also found in Australia, it suggests a common ancestry for Australia and Melanesia. Kayser et al. (2003) found four haplogroups on the Y-chromosome that most likely arose in Melanesia, before the Austronesian expansion. They have a distribution of high frequencies in the Highlands of New Guinea, and three of them are also found in Nusa Tenggara and the Moluccas, with higher frequencies in Papuan speaking populations than in Austronesian speaking groups.

In addition to the evidence from archaeology and human genetic studies, there are indications from linguistic studies that the greater New Guinea area was populated by different waves of migration. Nichols (1992; 1997; 1998) used statistically significant distributions of typological features to trace origin and dispersal of the world's languages. Since the traditional comparative method cannot reach further back in time than approximately 6,000 years, or ten millennia at the most (Nichols 2003; Rankin 2003), to determine genealogical ties between languages, relatively stable typological features can be used as 'historical markers' (shared by languages either because of genealogical descent or because of diffusion) to trace some common history.

For Sahul, Nichols (1997: 159–160) distinguishes two main strata of languages that are claimed to have a common geographical origin with similar strata in other parts of the world: The South-East-Interior (SEI) languages exhibit markers of the earliest colonisers of Sahul, the North-West-Coastal (NWC) languages have markers of the later colonisers, as given in Table 2.

In addition to the ten markers of Table 2, which as a cluster pattern in both New Guinea and Australia as they pattern on a global scale, differentiating the Old World (Africa and Eurasia) and the New World (the Americas), Nichols shows that the inclusive-exclusive opposition has a slightly different distribution in Sahul: it is almost universal in Australia, both in NWC and SEI languages, while in New Guinea

**Table 2.** Typical features of South-East-Interior and North-West-Coastal languages (after Nichols, 1997)

Earlier, SEI features	Later, NWC features
Dependent marking	Head marking
Ergative	Accusative; also stative-active, others
Lower passive/causative ratio	Higher (more passives)
Higher complexity	Lower complexity
No PPs	PPs
Fewer noun classes	More noun classes
No numeral classifiers	Numeral classifiers (minority feature even where relatively common)
More singular/plural neutralization	Less singular/plural neutralization (frequency is high even where less common)
No or few tones	Complex tone systems
One stop series	Two or more stop series

it is almost entirely confined to coastal and northern languages (1997: 150). Nichols concludes (1997: 160) that the eleven markers differ not only as to whether they polarize more strongly into north/south or coast/interior distributions, but also in the clarity of that division and the degree of parallelism between New Guinea and Australia. No two features have exactly the same distribution. This variety of distributions strongly suggests that there was more than one colonization per stratum.

This picture is somewhat refined in a following publication. Nichols (1998: 152–157) divides the language areas around the Pacific into three *provinces* on the basis of frequencies of a few other historical markers, in addition to some given in Table 2. The Pacific Interior province is dominated by descendants of a very early wave of colonization, whose languages are characterized by ergativity, rarity of head-marking, systematic marking of singular/plural (dual) oppositions on nouns,<sup>5</sup> minimal consonant systems (often limited to a single manner of articulation) and high frequency of derived intransitivity in the verbal lexicon. The Pacific Hinterland is a slightly expanded coastally oriented area, characterized by head marking, gender or other agreement classes in nouns, reduplicated plurals, extensive prefixation, causativisation as a regular derivational process in the verbal lexicon. This is a more recent stratum which has not penetrated deeply into the interior, present in both New Guinea and Australia. Finally, the Pacific Rim province, characterized mainly by identical SG and PL pronominal stems, *n* : *m* personal pronoun roots for first

5. Note the curious discrepancy with Table 2 after Nichols (1997), which suggests that singular/plural neutralization is predominant in South-East-Interior languages but much less so in the North-West-Coastal languages.



and second person, numeral classifiers, verb-initial word order, tones, and possessive classification, must have formed after New Guinea and Australia had been separated by rising sea-levels, since it is lacking in Australia but well represented in New Guinea. The latest wave of the Pacific Rim involved the Austronesians, who dispersed from Taiwan around 6,000 years ago, reaching the New Guinea area approximately 4,000 years ago. The west-east migration that Nichols assumes has been challenged by various linguists on the basis of linguistic evidence (Foley 2000; Ross 2005) or on the basis of historical documents (Voorhoeve 1989). In section 2.3.1 on the structure of Papuan languages we will elaborate on this issue.

Thus, it is plausible that these three provinces in the greater New Guinea area correlate to some degree with the multiple migrations proposed by Birdsell. We should emphasize the qualification 'to some degree', because we do not really know the time frames of each wave of colonization. Nichols (1998: 162) suggests that "more time probably elapsed between the Interior and Hinterland entries [possibly 50,000 to 20,000 years ago] than between the Pacific Hinterland and Pacific Rim strata", with the latter possibly starting 16 millennia ago and ending with the entry of the Austronesians.

#### 2.2.2 *The Austronesians*

Some three to four thousand years ago, the first Austronesians arrived through the Philippines in the Moluccan and New Guinea area (Bellwood 1997: 123). There is no evidence that all islands in the archipelago were inhabited when the Austronesians arrived, so that in some places they may be considered first settlers of the islands. However, in other places they will have encountered inhabited islands and various simplistic contact scenarios, violent and peaceful, are possible. In some places, they may have occupied whole islands. Nowadays, we find numerous islands, such as Seram, Buru, Biak, Manam, Manus, etc. that are completely 'Austronesian'. It is not clear whether size of island or population has any correlation with full or partial occupation by Austronesian speakers. The actual processes of linguistic replacement can no longer be determined. The Austronesians may have simply chased the ancestors of the current Papuan populations to other areas, or they may have conquered them, and through intermarriage and slavery obliterated the original languages while, perhaps, adopting some of their features. In other cases, they came to share certain islands, a situation that we still find today in many places in East Nusantara, for example, on Timor where the south-west is Austronesian and the north-east is Papuan; Halmahera where the south is Austronesian and the north is Papuan; Makian where on the east coast Austronesian Taba is spoken and on the west coast Papuan Moi (the endonym for West Makian); and Yapen where Papuan Yawa is spoken in central Yapen, and on either side we find Austronesian languages. It is important to realise at this point that we should not think of the dispersal of the Austronesians as a single event, just as the dispersal of the pre-Austronesians most likely was not. Periods of warfare and expansion were followed by more peaceful times in which trade relations between groups would be set up and allies would be found. In both situations language contact would occur: in times of war speakers of a language would have

been abducted and enslaved, in times of peace intermarriages would have introduced bilingual situations, as would trade.

We still know very little about the first 2–3,000 years of Austronesian presence in eastern Indonesia while Hindu influences found throughout Java, Bali and Lombok only reached the westernmost part of this area, there must have been contact between east and west. Clove trees (*Eugenia caryophyllata*) were indigenous only to the north Moluccan islands of Ternate, Tidore, Jailolo (Halmahera) and Bacan; nutmeg and mace were native to Banda (van Fraassen 1983: 3). Because these trees originated in the Moluccas, cloves and nutmeg serve as ‘tracers’ of contact between the Moluccas and the outside world. As long as 2,000 years ago cloves were transported to China and even to the Middle East. This means that there must have been trade relations far beyond the region even then, but we do not know who actually collected the spices in the Moluccas and there is no evidence for actual presence of, say, Indians, Arabs, Persians or Chinese in this period. It is likely that Austronesians from Java or Sulawesi traded cloves with the inhabitants of the islands and took them further West. These inhabitants would then be the predecessors of the current non-Austronesian speakers on e.g., Ternate, Tidore, Moti and Makian. But it is also possible that at the time some of the islands themselves were still uninhabited and that, in fact, it was the trade in cloves is what drove different groups of people, Austronesians and other, to establish settlements in the first place.

From the 12th century onward we can be a little more confident about historical developments. First, trade relations existed between the Moluccas and Alor/Pantar with groups from Java, Sulawesi, possibly China, and northern India. Islam was introduced to Ternate at around 1460 and to Banda around 1480. In Ternatan accounts of this event, no distinction is made between the arrival of the first Malay traders and the formal acceptance of Islam (Jacobs 1971: 104–105; Reid 1984: 24, cited in Dix Grimes 1991: 93). During the last decades of the 15th century both Ternate and Banda were incorporated into the greater Malayo-Muslim trading network of cities spread throughout Southeast Asia. By the 16th century, Malay had become a lingua franca over much of the archipelago (Bellwood 1997: 122).

Between the 13th and the 18th century the kingdoms in the North Moluccas increased their economic and political power in the region and Ternatans and Tidorese travelled south as far as Banda, north to Mindanao and east to the Bird’s Head of Papua. In the 17th century, Tidorese often led headhunting and raiding expeditions (*honggi*) to other islands. The traditional routes of these expeditions went southward to the Aru-Kei islands, Tanimbar, the Seram Laut Islands, Seram, Buru, Ambon, as well as northward to the Sulas, Banggai, and north Sulawesi (Andaya 1993: 192). Both Malay and the languages of Ternate and Tidore (at present still very similar) were used for intergroup communication. When the first Europeans arrived in the late 16th century, people from the Bird’s Head lived on Ternate and Tidore as slaves (*popuha* in these languages). In fact, in the entire area slave trade was very common, (cf. also Needham 1983) which must have implied the displacement of Austronesian speakers

to non-Austronesian speaking areas, and vice versa.<sup>6</sup> About two centuries later, the political and commercial relations between Tidore and the Moluccan islands towards the south, including Seram, Banda and Kei, appear to have remained just as tight. During the last quarter of the 18th century, the famous Tidore ruler Nuku, who rebelled against the Dutch East Indies Company, had to escape from Dutch expeditions directed against him and for several years travelled with a group of followers around the Moluccan archipelago (Andaya 1993: 219–232). The fact that this was possible for a Tidorese ruler suggests that the Moluccan islands were indeed considered an entity, and that this entity was connected with Tidore and Ternate (see Andaya 1993 for argumentation). Since Nuku is also reported to have traded Papuan slaves, sea cucumber, and tortoise-shell for gunpowder and ammunition from Banda, slave trade must still have been common practice at the time as well.

### 2.2.3 *European Colonisers and modern state formation*

Although the Europeans were mainly drawn to this part of the world for the spices, they soon embarked on missionary activities too. Some colonial powers were more adamant on spreading Christian faith than others, and of course the Portuguese and Spanish would propagate Catholicism while the Dutch advocated Protestantism. Not many Muslims were converted, but among those groups that had ‘animist’ traditions or were otherwise ‘non-religious’, Christianity was more successful. The reason that religions, both Islamic and Christian, are relevant for the determination of a linguistic area, is that along with religion new languages and genres were introduced. For example, Islam introduced Arabic orthography as well as Arabic as a language of religion, while the Dutch Protestant church on Ambon introduced a particular variety of Malay, called ‘High’ or ‘Church’ Malay. This literary Malay variety contrasted with ‘low’ Malay, the regional lingua franca and was introduced by the colonial government through a High Malay Bible translation in 1733 (Dix Grimes 1991: 98–99). This translation became a school text for education, and its language was the foundation for the church language in the Central Moluccas.

The different colonial powers also played an important role in determining how the different parts of the area developed into parts of nation states in the 20th century. After the Indonesian declaration of independence on August 17 1945, Indonesian was introduced as the national language. Through the educational system and the media, Indonesian as well as local Malay varieties have become increasingly dominant, and are steadily replacing the local languages of East Nusantara. Dutch New Guinea was included as Indonesia’s easternmost province Irian Jaya (now Papua) in 1962, and in

---

6. For example, in the 18th century we find references to old treaties which allowed Tidorese to buy slaves in the New Guinea area, in particular around what is now called Fak-Fak, on the southern shores of the MacCluer Gulf and the Bird’s Head in Papua. The Papuan slaves bought by the Tidorese from people in Fak-Fak had been recruited from the interior of Papua, probably from the Inanwatan area across the Gulf (Valentyn 1724, cited in Van Staden 2000: 8).

1974 the Indonesian army occupied the Portuguese colony of East Timor. Both the Papuans and the East Timorese peoples resisted incorporation into the Indonesian Republic, culminating in the Independent Republic of East Timor in 1999 and a name change and special status for Irian Jaya / Papua in the year 2000. In both areas, however, the influence of the Indonesian political and educational system and the official Indonesian language have been significant: on East Timor the older (educated) generations still have some command of Portuguese, but the people who had their education between 1974 and 2002, are fluent in Indonesian. In 2002, Portuguese has been re-introduced as the language of education in East Timor, besides Tetun as the national language. In Papua, many of the indigenous languages are being replaced by eastern Indonesian local variants of Malay.

### 2.3 The languages of East Nusantara

#### 2.3.1 *The Papuan languages*

The Papuan languages are both lexically and morphosyntactically a highly heterogeneous group, and it is often difficult to impossible to determine genealogical ties between the individual languages on the basis of the familiar methods of lexical comparison (for discussion and references, see for example Foley 1986, 2000). This in itself is not a surprise, since most successful reconstructions in other language families go back only as far as approximately 6,000 to maximally 10,000 years (Nichols 1998: 128), and have benefited from both archaeological and historical linguistic evidence. By contrast, the Papuan languages are the descendants of various waves of colonizers starting at least 40,000 years ago, but probably much earlier, as the linguistic age of New Guinea and Australia together is estimated at 60,000 years (Nichols 1998: 138), roughly correlating with the oldest archaeological evidence. This is far too long ago to apply the comparative method.

Wurm (1982) proposed five major phyla of 'Papuan' languages, as well as six minor ones and a number of isolates. More conservative estimates (e.g., Foley 1986) suggest that there are at least 60 different families (some of which consisting of only a few members or even isolates) for which genealogical ties cannot be established yet. The largest family for which there is general agreement is the Trans New Guinea (TNG) family, with close to 300 languages and some two million speakers. This family comprises about half the Papuan speaking population (Foley 2000: 363), but represents only a tiny fraction of the genealogical variation found in Papua. The label 'Papuan', then, does not refer to a superordinate category to which all the languages belong. Rather, the term is used for a negatively defined group of languages: the non-Austronesian languages spoken in New Guinea and archipelagos to the West and East. Nevertheless, there are a number of characteristics that, although too general to give true 'typological affinity' (Wurm 1982: 36), may point to a 'closer (though in some way secondary) affinity between these languages' (ibid.). This affinity could be genealogical in origin, but it could also be the result of language contact and language mixing, as both

Wurm (1982) and Foley (1986) stress. Foley (2000) is the most recent overview of general 'Papuan' characteristics, of which we list here only few of the more general ones, including those of particular relevance here.<sup>7</sup>

The phonemic inventories of languages in New Guinea tend to be simple. Generally, the number of segmental phonemes is approximately two dozen, although an extreme exception is Yele, or Yéli Dnye, spoken on Rossel island, which has a total of 94 contrasting sounds (Henderson 1995: 11). The great majority of Papuan languages have only a single liquid phoneme. In the Austronesian languages, by contrast, a phonemic distinction between /r/ and /l/ is virtually universal. Tone systems are found in a number of Papuan families, including various sub-families of TNG, Skou, Lakes Plain, and as we will see in a number of Papuan languages of East Nusantara, but often not in all their member languages.

Syntactically, Papuan languages are overwhelmingly head-final, with SOV constituent order. Typical of the TNG family, but not restricted to it, is clause chaining, often with some concomitant switch reference system. Such a system basically encodes whether the following clause in the chain has the 'same' or a 'different' subject, although in many languages it is more aptly described as marking changes in topic or setting.

Morphologically, verbs are the most complex word class in many Papuan languages, such as the major groupings of TNG, Sepik, and Trans-Fly languages. The majority of Papuan languages are head marking. The canonical verbal structure for the TNG languages is a bound pronominal agreement prefix for object, and a suffix for subject, often as portmanteau with TAM distinctions. Verbal prefixing for subject is found in various language groups along the north coast, which generally have few other affixal categories: most of the Papuan languages of the East Nusantara region, the Skou family, Torricelli, Lower Sepik and some East Papuan languages. A number of languages in the north coastal region also exhibit a greater degree of morphological complexity in nouns than is found in most TNG languages. Numeral classifiers are widespread in the Papuan languages of East Nusantara; noun class systems are found in isolates in northern Papua as well as in the North Halmahera family, in members of the Torricelli and Lower Sepik-Ramu families and various East Papuan languages. Roughly coinciding with these groups, although not in the East Nusantara region, are languages with nouns inflected for number in an extremely irregular fashion. Reduced nominal classification of gender in pronouns (often just for 3sg) is a typical Papuan feature. It is found in most non-TNG languages along the north coast, and in some TNG languages along the Indonesian – Papua New Guinea border. An inclusive/exclusive contrast in the first person plural pronouns – a universal feature of Austronesian languages – is found in many Papuan languages neighbouring Austronesian languages, but typically absent in others.

---

7. For the examination of 'Papuan' characteristics it is important to realize that they do not define some Papuan essence of all these languages. Some features may only typically be found in certain subgroupings.

There are some indications that the Papuan languages of the East Nusantara area reflect traces of at least two original strata. The marking of gender, which represents a reduced system of nominal classification, is a feature that appears to be stable through time only when reinforced by gender systems in neighbouring languages (Nichols 2003: 303). Gender and extensive noun class systems are widespread in the Papuan language families along the northern rim of where Papuan languages are spoken: from North Halmahera all the way to the Solomon Islands in the Pacific, coinciding with Nichols' Pacific Rim or North-West-Coastal populations. The distribution of tone systems in Papuan languages of the Bird's Head may be another marker of this stratum, although tone systems are also rather widely attested in various sub-families of TNG of the interior and not available in most of the West Papuan languages. In other words, at least in New Guinea, tone is not a very distinctive areal or genealogical feature.

The Papuan languages of the Timor-Alor-Pantar and South Bird's Head families are claimed to be members of the large Trans New Guinea family. According to Foley (2000: 395), the languages of this family closely fit Nichols' South-East Interior profile (e.g., ergative, dependent marking, fewer noun classes, no or few tones; see 2.2.1), but they "believe the migration pattern expected by Nichols' summary". Foley suggests, as does Ross (2005), that the homeland of this family is located somewhere in the Eastern highlands of New Guinea and that the languages spread (as a result of language shift? or by means of peoples' migrations?) from east to west, all the way to the Timor area. Voorhoeve (1989: 82) likewise addresses this question. He suggests that the migration may have been east-west, on the basis of a tradition of speakers of the non-Austronesian language Fataluku in East Timor, according to which they originally came from the Kei islands in the East (Capell 1972). And in an unpublished grammar sketch of the Iha language, a non-Austronesian language spoken in south West Papua, the Dutch Roman Catholic missionary Coenen mentions that in pre-contact days the Iha speakers went on slave expeditions all the way to the Kei and Tanimbar islands. This suggests at least the existence of east-west maritime contacts between the two ends of the chain Papua-Timor, and a point in between, Kei (Voorhoeve 1989: 82).

Yet, it seems more parsimonious to assume that these patches of Papuan languages are remnants of ancient continuous populations than to assume that New Guinea highlanders migrated back over water to small islands such as Alor and Pantar. The Timor-Alor-Pantar languages and their putative relatives of the South Bird's Head and Bomberai peninsula may well be part of the early South-East-Interior populations, while the northern groups (present-day North Halmahera and most of the Bird's Head and Yawa) appear to belong to the North-West-Coastal (= Pacific Rim/Hinterland) migration(s). In other words, whether the Papuan languages presently spoken in North Halmahera and in Timor-Alor-Pantar are the result of east-west migrations, but still predating the arrival of the Austronesian speaking populations, or whether they are remnants of more widespread Papuan populations throughout the archipelago, it is clear that there is an old connection between Timor and Halmahera regions and the Papuan mainland New Guinea, as the human genetic studies indicate (see section 2.2.1).

Lexically, the Papuan languages of East Nusantara have little in common. The Papuan languages of Timor-Alor-Pantar are related, and in North Halmahera, too, we find a clear set of related languages. But a conservative estimate gives nine distinct families of Papuan languages in East Nusantara, (see also Map 3):

#### Cenderawasih Bay

- (1) Yawa (isolate) (Jones 1986; Reesink 2005)

#### The Bird's Head, with three families and three isolates.

- (2) East Bird's Head family (Voorhoeve 1975; Reesink 2002a): Meyah; Sougb
- (3) West Bird's Head family (Voorhoeve 1987): Moi; Tehit; Moraid; Seget
- (4) Hatam and (extinct) Mansim (Reesink 2002a)
- (5) Mpur
- (6) Maybrat
- (7) Abun

#### North Moluccas

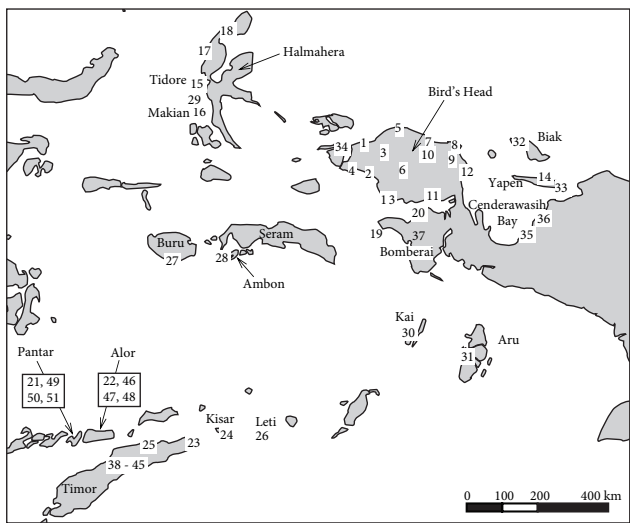
- (8) The North Halmahera family with four subgroups or languages (Voorhoeve 1987, 1989): the dialect chain: Galela, Tobelo, Pagu; Sahu; Tidore-Ternate; West Makian

#### Southern Bird's Head and Timor area

- (9) The Trans New Guinea family with four subgroups in East Nusantara:
  - South Bird's Head, with Inanwatan (Voorhoeve 1975; Wurm 1982; Berry and Berry 1987; De Vries 2004)
  - West Bomberai: Iha, Baham
  - West Timor-Alor-Pantar: Bunak, Abui, Adang, Klon, Kafoa, Blagar, Nedebang, Teiwa, Lamma
  - East Timor: Oirata, Makasai (Ross 2005)

There are some indications in the lexicon and the bound morphology, in particular the subject cross-referencing on the verb, that suggest a very distant common origin for Yawa, the Northern Bird's Head languages and the North Halmahera family, (see Reesink 1996, 1998, 2005; Ross 2005, to appear, for discussion and references). Evidence for assigning Inanwatan and the Timor-Alor-Pantar languages to the TNG family is extremely slender (Pawley 1998: 683), but Ross (to appear) presents several pronominal forms in Proto-West Bomberai-TAP that reflect forms in Proto Trans New Guinea. Within these two groups, the East Timor family occupies a position midway between the West Bomberai and West Timor-Alor-Pantar languages, sharing different pronominal innovations with each, presenting evidences of an erstwhile dialect chain. Inanwatan is possibly related to the Marind family, but its position in the TNG family is also highly uncertain (De Vries 1998, 2001, 2004).





**Papuan languages:**

**Bird's Head:**

- 1      Moi
  - 2      Tehit
  - 3      Moraid
  - 4      Seget
  - 5      Abun
  - 6      Maybrat
  - 7      Mpur
  - 8      Mansim
  - 9      Hatam
  - 10     Meyah
  - 11     Moskona
  - 12     Sougb
  - 13     Inanwatan
- Yapen island**
- 14     Yawa
- Tidore island**
- 15     Tidore
  - 16     West Makian
- North Halmahera**
- 17     Sahu
  - 18     Galela
- Bomberai peninsula**
- 19     Iha
  - 20     Baham
- Pantar island**
- 21     Blagar
  - 49     Nedebang
  - 50     Teiwa
  - 51     Lamma
- Alor island**
- 22     Kabola/Adang
  - 46     Abui
  - 47     Klon
  - 48     Kafoa
- Timor island**
- 23     Makasai
  - 45     Bunak
- Kisar island**
- 24     Oirata

**Austronesian languages:**

- Timor island**
- 25     Tetun
  - 38     Idate
  - 39     Isri
  - 40     Kemak
  - 41     Lakalei
  - 42     Lolein
  - 43     Mambai
  - 44     Tokodede
- Leti island**
- 26     Leti
- Buru**
- 27     Buru
- Ambon**
- 28     Asilulu
- Makian island**
- 29     Taba
- Kai island**
- 30     Keiese
- Aru island**
- 31     West-Tarangan
- Biak**
- 32     Biak
- Yapen island**
- 33     Ambai
- Raja Ampat islands**
- 34     Ma'ya
- Bomberai**
- 37     Irarutu
- Cenderawasih coast**
- 35     Waropen
  - 36     Mor

Map 3. Location of a selection of languages of East Nusantara



2.3.2 *The Austronesian languages*

The classification of a language as ‘Austronesian’ is far less problematic than for the Papuan languages and does imply a clear genealogical relationship. However, within the various sub-groupings of Austronesian, there is still much ongoing research to determine the precise classification of particular languages. The reasons for these problems are probably rapid migration as well as prolonged and complex contact situations between languages, which have led to diffusion of features and borrowing that may now obscure original genealogical relations between languages. The vast majority of Austronesian languages of East Nusantara belong to the large group of the Central Eastern Malayo Polynesian (CEMP), which has approximately 600 members and comprises the languages of eastern Indonesia and almost all the Pacific languages (Blust 1993). In East Nusantara we find two major sub branches of CEMP: the Central Malayo-Polynesian (CMP) languages and the South Halmahera/ West New Guinea (SHNWG) languages. The latter constitute again a subgroup of the larger Eastern Malayo-Polynesian (EMP) family, which includes the Oceanic languages (Blust 1993: 274). Figure 1 gives the three major branches and a necessarily non-exhaustive list of member languages. For an indication of their geographical location, see the Appendix.

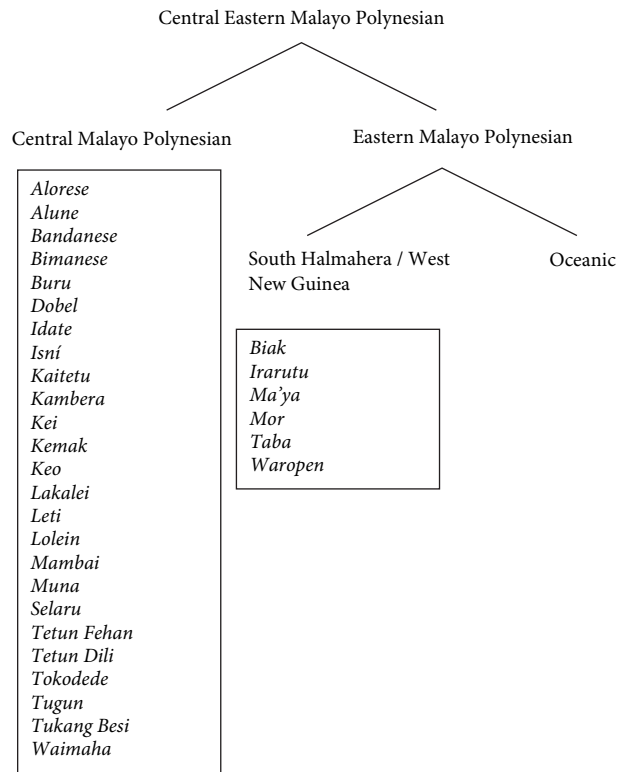


Figure 1. Some Central Eastern Malayo Polynesian languages discussed in this paper, per subgroup, in alphabetical order.

The classification of a language as either CMP or SWHNG is difficult (Blust 1993: 271ff.; Ross 1995; Grimes 2000). A few characteristics of SWHNG are: loss of vowel between nasal and following stop, shift of *\*e* to *\*o* in penultimate position, and the replacement of *\*anak* with *\*natu* for ‘child’. Diagnostic features of CMP include: glide truncation in diphthongs, postnasal voicing, loss of prepenultimate initial vowels, and the replacement of *\*qasu* by *\*masu* for ‘smoke’. On the whole it appears that the SWHNG languages are less conservative in their basic vocabularies than most CMP languages (Blust 1993: 245), which may be due to more extensive contact (substrate?) with Papuan languages.

Some typological characteristics of Austronesian languages in which they contrast with Papuan languages in general are (i) a phonemic distinction between /r/ and /l/, (ii) a predominance of bisyllabic lexical morphemes (CVCV), (iii) if possessors are affixed, they are suffixed rather than prefixed (Klamer 2002), (iv) common occurrence of reduplication, and (v) a distinction between the 1st plural inclusive and exclusive. Syntactically, the Austronesian languages are typically head-initial, i.e., they are verb initial or verb second, and their negator precedes the predicate.

### 3. Austronesian features in Papuan languages

In this section we discuss two features that appear to have diffused from Austronesian to Papuan languages. First, we consider constituent order, arguing that the SVO structure found in some Papuan languages in our survey is a contact phenomenon. Next, we discuss the distinction between 1PL exclusive (‘we without you’) and 1PL inclusive (‘we including you’). When this distinction is marked in the pronominal paradigm of a Papuan language in our sample, we assume it is the result of diffusion.

#### 3.1 Primary constituent order

All the Austronesian languages in East Nusantara have SVO constituent order, correlating with the typical head-initial phrase structure found in Austronesian languages (Clark 1990; Tryon 1995; Foley 1998; Klamer 2002). Apart from VO constituent order, such languages typically have prepositions rather than postpositions, clause-initial/preverbal/pre-predicate complementisers and negators, and possessed nominals preceding the possessor. In East Nusantara, the Austronesian languages virtually all have prepositions, with the exception of Alorese which has postpositions, and all, including Alorese, have clause initial complementisers. We return to the issue of word order in the possessive construction in 4.1 and the placement of the negator in 4.2, where the Austronesian languages give a more diverse picture.

By contrast, Papuan languages are generally head-final, with OV constituent order, postpositions, final complementisers, possessor-possessum order, and clause final negators (Foley 1986, 2000). However, in East Nusantara we find both SOV and SVO constituent

order in Papuan languages. In Alor/Pantar, all the Papuan languages have SOV as the basic constituent order (cf. Steinhauer 1995; Nitbani et al. 2001; Kratochvíl, 2007, Klamer, *forthc.*, Baird, *in press*), and this is also the case in Timor (Makasai: Brotherson 2003: 78, 80; Bunak: Friedberg 1978).<sup>8</sup> In Halmahera, however, as Voorhoeve (1987, 1994) argued, all the Papuan languages originally had SOV order, but a few (Sahu, Ternate-Tidore and West Makian) have now shifted to SVO constituent order (see also Reesink 1998: 633; Foley 2000: 393). There are occasional examples with VO order in descriptions of North Halmaheran languages, as for example in Pagu (Wimbish 1991: 103):

- (1) *Yo-uit-isa ya-siguti ma naok.*  
 3PL-descend-land 3PL-unload ART fish  
 ‘When they got out, they unloaded the fish.’ Pagu

The South Bird’s Head languages (de Vries 1996, 2001) also have SOV order, and so does isolate Yawa spoken in the Cenderawasih Bay (Jones 1986; 1991). However, most of the Papuan languages spoken in the Bird’s Head have SVO constituent order: the West Bird’s Head languages, Moi (Menick 1996) and Tehit (Flassy and Stokhof 1979), the isolates Abun (Berry and Berry 1999), Maybrat (Dol 1999), Mpur (Odé 2002a), and the small families in the eastern Bird’s Head, Hatam-Mansim (Reesink 1999) and Meyah-Sougb (Gravelle 2002; Reesink 2002a). Some of these Papuan SVO languages show evidence of head-final phrase order in other areas. Tidore, for instance, has clause final complementisers, although it also has prepositions, and all of the SVO languages have post-predicate negation. Outside East Nusantara, SVO word order is rare among Papuan languages, found only in a number of languages along the north coast of New Guinea in areas where contact with Austronesian may be assumed. Although spontaneous shift from SOV to SVO is possible, it is reasonable to assume that in East Nusantara the shift is the result from contact with Austronesian.

### 3.2 Inclusive/Exclusive opposition

It is a general feature of Austronesian languages, reconstructed even for Proto-Austronesian, to have an opposition inclusive-exclusive for the first person plural. The Austronesian languages of East Nusantara follow this pattern, with the exception of local varieties of Malay (Van Minde 1997; Baird, Klamer and Kratochvíl 2004). The inclusive/exclusive distinction is not generally found in Papuan languages, spoken in the interior of New Guinea. Yet, many of the Papuan languages in East Nusantara have

---

8. SVO is attested as minor constituent order pattern in some of the languages of Alor/Pantar (Klon, Teiwa) and Timor (Bunak). What exactly determines this minor order is (yet) unclear; it could be the sign of an ongoing language shift, but may also be determined by pragmatics or discourse considerations. Therefore we consider the major pattern only, and classify these languages as SOV here.

the distinction. In this section, we show that this Austronesian feature has diffused into Papuan languages.

All the Austronesian languages in our sample have the inclusive/exclusive opposition for the first person plural. Although the inclusive/exclusive opposition for first person plural is not generally 'Papuan', in East Nusantara we find that the majority of the Papuan languages do have this opposition. It is found in all the Papuan languages of Alor/Pantar surveyed by Stokhof (1975: 16–17), with the exception of Kolana in East Alor. In Timor, both Bunak (Friedberg 1978: 25) and Makasai (Brotherson 2003: 28) make the distinction as well as all North Halmahera languages (Voorhoeve 1987). In the Bird's Head, the extent to which the distinction is encoded differs. The EBH family, Meyah and Sougb, have a robust distinction in non-singular first person pronominal forms (Reesink 2002a), as do the WBH family (Reesink 1996) and Inanwatan of the South Bird's Head (De Vries 1996). Hatam, however, marks the distinction not in the free pronouns but only in the verbal prefixes, where the inclusive form is identical to 3PL and the exclusive is the same as 3SG POS. Other languages in the Bird's Head, such as the isolates Maybrat, Abun and Mpur, located more centrally in the peninsula, do not have the distinction. Outside East Nusantara, the distinction is found in some Papuan languages mainly along the north coast of New Guinea, but also in a few languages spoken in the interior.

Interestingly, none of the local Malay varieties spoken in Papua, the Moluccas and Alor/Pantar have the distinction. We have no explanation for the loss of this distinction in the contact language when both the indigenous languages and the lexifier of the contact language do. Perhaps it is the result of European and other foreign traders learning the contact language imperfectly, but this is mere speculation. What we can say is that it is highly unlikely that the distinction entered the Papuan language through these Malay varieties, but that it is a much older feature of the Papuan languages of East Nusantara. There is some disagreement still on whether the forms are borrowed from Austronesian languages. Voorhoeve (1994: 661) suggests that they are, but Ross (2005) is not convinced and believes that it is possible that the presence of the distinction predates even the arrival of the Austronesians, because it is also found in Senagi and Border languages, spoken in the interior of New Guinea, for which an Austronesian contact scenario is unlikely. Yet in East Nusantara, it appears that the inclusive/exclusive distinction for the first person plural, a typically Austronesian feature, occurs just in those Papuan languages that have had a long history of contact with surrounding Austronesian languages.

#### 4. Shared Papuan features

In this section we review a number Papuan features found in both Papuan and Austronesian languages in the area under discussion. Three of these have to do with the categorisation and expression of possessive relations: alienability (4.1.1), the order of the possessor and possessum in adnominal possession (4.1.2), and the morphological marking in the possessive construction (4.1.3). The fourth Papuan feature concerns the

occurrence of final negators (4.1.3). Tone, finally, is found in a number of Austronesian subgroups, but typically not in the CEMP languages. At the same time it is weakly linked to Papuan languages. It is remarkable then, that in East Nusantara we find a small set of Papuan languages with tone, but moreover, two neighbouring Austronesian languages that also exhibit tone. It appears that this feature has also diffused from Papuan to Austronesian (4.3).

#### 4.1 Possessive constructions

##### 4.1.1 *Alienability*

For the Austronesian languages, the alienability distinction has been claimed to be an innovation of the CEMP subgroup (Blust 1993: 258), which includes the majority of the Austronesian languages of East Nusantara. It does not occur in the Western Austronesian languages. This innovation must have occurred prior to the population of Oceania, as Ross (2001: 138) hypothesizes that “it is also probable that the formal distinction between alienable and inalienable possession entered Proto-Oceanic or an immediate precursor through Papuan contact”. In view of the data at hand, this hypothesis appears correct. Virtually all the Papuan languages of East Nusantara do have this distinction, and wherever it occurs in the Austronesian languages of East Nusantara these languages are spoken in areas with Papuan contact. Furthermore, as we demonstrate in sections 4.1.2 and 4.1.3, the structure of the possessive constructions in these languages also warrants a contact scenario with inalienable possession marked more conservatively than alienable possession virtually everywhere.

Where languages mark the difference between alienable and inalienable possession, the latter group typically contains terms designating a “close biological or social bond between two people” (Heine 1997: 11) body parts and other part-whole relations, spatial relations, and objects “essential for one’s livelihood or survival” (*ibid.*; see also Chappell and McGregor 1996; Nichols 1992). In the languages of East Nusantara it is typically kinship and body part terms that are included in this category of inalienables. The inclusion of spatial relations and some artefacts is reported for only a few. In one exceptional language, Austronesian Taba of the North Moluccas, only part-whole relations are marked differentially, while body parts and kinship terms are treated as ‘alienable’ (Bowden 2001:233–34). For Taba this means that inalienables have obligatory expression of the possessive relationship, as in (2a), while for alienables, the possessor may be omitted (2b):

- |     |    |                        |    |                    |      |
|-----|----|------------------------|----|--------------------|------|
| (2) | a. | <i>meja ni wwe</i>     | b. | <i>Wwe mhonas.</i> |      |
|     |    | table 3SG.POS leg      |    | leg be.sick        |      |
|     |    | ‘the leg of the table’ |    | ‘My leg is sore.’  | Taba |

It has been questioned whether the alienability distinction is similar to ‘gender’ in the sense that it categorises the lexicon, or whether it should be treated rather as a semantic relation between the possessor and the possessum (Heine 1997: 17, cf. also Lynch 1978). Grimes (1991: 287), in his treatment of Buru possession, argues in favour

of the latter when he discusses the different uses of the word *olo* ‘head’, used inalienably in (3a) and alienably in (3b):

- (3) a. *Da iko tu olo-n.*  
 3SG go with head-3SG.GEN  
 ‘He went with (accompanied possession) its (pig’s) head.’  
 b. *Da iko tu nak olo.*  
 3SG go with 3SG.POS head  
 ‘He went with (comitative) his (social/political) head.’ Buru

Most treatments of the distinction in East Nusantara do not report on this issue and for our discussion we will include all descriptions of alienability as one phenomenon.

Typically, the difference between the two categories is marked morphologically. Svorou (1993: 198ff) observes that inalienables tend to be zero marked while alienables have some morphological marking. We find this in only one language in East Nusantara. Papuan Abun (Berry & Berry 1999: 79) expresses alienable possession with a ligature *bi* between possessor and possessum as in (4a), while inalienable possession is expressed by the simple juxtaposition of possessor and possessum as in (4b):

- (4) a. *an bi nji bi nggon bi nu*  
 3SG POS brother POS wife POS house  
 ‘his brother’s wife’s house’  
 b. *Wo Kwai tik Sepenyel gwes.*  
 fish Kwai pull Sepenyel leg  
 ‘The kwai fish pulled Sepenyel’s leg.’ Abun

Inanwatan (De Vries 1996: 104–106), however, arguably has the reverse, with inalienable nouns marked by a person prefix on the possessum, as in (5a), and constructions with alienable nouns with a gender marked possessive pronoun that precedes the possessum without any further marking:

- (5) a. *Ná-wir-i me-tutú-rita-bi.*  
 1SG-belly-M 3SG-hurt-DUR-M  
 ‘I (male) have pain in my belly (lit. my [male] belly hurts).’  
 b. *Owó-i nárido-wo méqaro-wo.*  
 that.F-SG 1SG.POS-F house-F  
 ‘That is my house.’ Inanwatan

Klon of Alor island (Baird, in press) and Bunak in central Timor (Friedberg 1978: 28–30) have a similar pattern, in which inalienably possessed items are inflected for the person and number of their possessor, as in (6a), whereas the possessor of alienably possessed items is expressed by a separate pronoun, as in (6b):

- (6) a. *g-agar*  
 3-mouth  
 ‘His/her/their-mouth.’

- b. *gie deu*  
3 house  
‘His/her/their house’
- Bunak

More commonly the distinction is marked through different morphological marking, possibly combined with different word order. In this context, Lichtenberk (1985) distinguishes two types of possessive constructions: ‘direct possession’ which involves a construction in which the possessor is directly cross-referenced on the possessum and ‘indirect possession’, which has a ligature or ‘possessive classifier’ of some kind. In his overview of Oceanic languages, Lichtenberk (1985: 103) found that the distinction between ‘direct’ and ‘indirect’ possession may be considered the ‘hallmark of the Oceanic subgroup’ (1985: 95–96), whereby the direct construction is typically used to express inalienable possession, and the indirect one is used for alienable possession.

The northern Papuan languages of East Nusantara, tend to conform to this pattern also. The inalienables typically take a prefix that derives from a paradigm (nearly) identical to the subject or object prefixes found on verbs indicating person and number of the possessor, while alienable possession is expressed with the possessive prefix attached to a possessive ligature that is often of likely verbal origin. For instance, throughout the eastern Bird’s Head we find inalienable possession expressed by a subject prefix on the possessum, as in Mpur (Odé 2002a: 62) and Hatam (Reesink 1999: 49), which both include the words for ‘name’ in the category of inalienables:

- (7) a. *An-muk* b. *an-tar jan*  
2SG-name 2SG-POS house  
‘your name’ ‘your house’
- Mpur
- (8) a. *A-nyeng tou i?* b. *a-de singau*  
2SG-name who QUEST 2SG-POS knife  
‘What is your name?’ ‘your knife’
- Hatam

In Yawa (Cenderawasih Bay) inalienable nouns (9a) have a prefix identical to the undergoer prefixes used on transitive verbs and uncontrolled intransitive verbs, as in (9b) (Jones 1986: 44–49). The expression of alienable possession involves a ligature and a different set of person markers as in (10) (Linda Jones, unpublished texts):<sup>9</sup>

- (9) a. *in-aneme* b. *In-awabea*  
1SG-hand 1SG-yawn  
‘my hand’ ‘I yawn.’
- Yawa
- (10) *Weti sy-a ana-syora yamo, syopi no naije.*  
so 1SG-POS NOM-speech TOP arrive LOC there  
‘So as for my speech, it’s finished.’
- Yawa

9. Sougb has a different set of affixes for both the inalienables, when they are directly prefixed to the noun, and the alienables when they are prefixed to a possessive ligature. (Reesink 2002a: 218).

There are some languages, however, in which, as in Maybrat, additional differences are found, or in which the distinction is marked in a different manner, as for instance in Tidore. In Maybrat (Dol 1999: 149) the inalienables are inflected like verbs with a subject prefix, but in addition, there is a different word order for alienable and inalienable possession. In inalienable possession the order is possessor-possessum (11a), but alienably possessed nouns are followed by a relator *ro* and the possessor, as seen in (11b). In this example, the possessor is itself an inalienably possessed noun *t-atia* ‘my-father’ with a subject prefix:<sup>10</sup>

- (11) a. *fnia m-ao*      b. *amah ro t-atia*  
          woman 3U-foot      house pos 1SG-father  
          ‘the woman’s foot’      ‘my father’s house’  
          Maybrat

In Tidore inalienably possessed nouns, including also words for names, boats and houses, require a possessive prefix or an invariant marker *ma-*, which with alienably possessed nouns can only be used for third person neuter possessors (Van Staden 2000: 125–126; 253). This means that with alienable nouns agreement between free pronoun and possessive pronoun is obligatory, while with inalienable nouns there is the choice between person/number marking and invariant *ma*. In addition, alienable nouns frequently occur without a possessor whereas this is very unusual for inalienable nouns:

- (12) a. *Mina ma-ronga nage?*  
          3SG.F INAL-name who  
          ‘What’s her name?’  
       b. *Ma-fola nde jang bahaya!*  
          INAL-house 3NH.here beautiful very  
          ‘This house is really beautiful!’  
          Tidore

In the Papuan languages of Timor, Alor and Pantar we find that the ‘direct’ versus ‘indirect possession’ opposition works less well. As indicated above, Klon and Bunak are like Inanwatan with a ‘direct’ construction for inalienable possession, but no ligature with alienables. Teiwa, like Tidore, marks alienable and inalienable possession with the same set of forms, but for alienably possessed items the possessor prefix is optional, while for inalienably possessed items the prefix is obligatory. Other languages of this area have different person markers for alienable and inalienable possession. Prefix or bound pronoun is typically used to mark objects, free form for subjects, as in Blagar (Steinhauer 1993: 150–151):

- (13) a. *n-amal*      b. *ne quu*  
          1SG-voice      1SG tuber  
          ‘my voice’      ‘my tuber’  
          Blagar

10. U = unmarked, used for both 3SG.F and 3PL



Adang, finally, has an obligatory possessor noun for alienables and while the genitive particle is optional (14a.), for inalienables this is reversed with an obligatory genitive prefix and an optional possessor noun (14b.) (Haan 2001, section 5.3):

- |      |    |                     |         |           |    |               |                 |       |
|------|----|---------------------|---------|-----------|----|---------------|-----------------|-------|
| (14) | a. | <i>hiu</i>          | (?e)    | <i>dɔ</i> | b. | (John)        | <i>ʔa-taang</i> |       |
|      |    | chicken             | 3SG.GEN | egg       |    | John          | 3SG.GEN-hand    |       |
|      |    | 'the chicken's egg' |         |           |    | 'John's hand' |                 | Adang |

On the whole, however, it appears that in the Papuan languages of East Nusantara the alienability distinction is a unified phenomenon. It features in virtually all the Papuan languages of East Nusantara, with the exception of some of the North Halmahera languages (e.g., Tobelo, Holton, p.c. see also Van Staden *forthc.*). Although it is not a universal feature in the Papuan languages, the distinction between alienable and inalienable possession is found in a number of different Papuan families (recall section 2.3.1) and can be seen as a 'Papuan trait'.

In the Austronesian languages, the distinction gives a more diverse picture, both in terms of its occurrence and in terms of the construction types that express it. It occurs in the majority of Austronesian languages of East Nusantara, but within this area clear borders can be discerned. It does not generally occur west of Alor/Pantar and Timor. Keo, Bimanese and Kambera, for instance, do not make the distinction. Also languages to the north of Alor/Pantar, such as Muna (Van den Berg 1989) and Tukang Besi (Donohue 1999: 346), lack the distinction. On Timor only a subset of Austronesian languages (e.g., Waimaha, Lakalei, Isní, Lolein, and Kemak on East Timor, cf. Hull 2001a: 123–125) have the distinction. Yet, east of Timor it is a common feature. As such it crosscuts genealogical boundaries since all these languages belong to either the CEMP or the SHNWG. The marking of the distinction is again typically a combination of word order and morphological marking but among the Austronesian languages there are few that have identical systems. We find languages with prefixes or suffixes, and languages without any affixation on the possessum, languages in which the possessor precedes or follows the possessum, languages with and without possessive ligatures, that in turn may but need not be inflected, etc. It is not always easy to characterise these languages in terms of Lichtenberk's (1985) 'direct' and 'indirect' possessive constructions. Nevertheless, two areas may be discerned: west and east, with a boundary just west of the Bird's Head.

In the western part of East Nusantara, the inalienables often have a possessor suffix and the alienables the word order possessor-possessum. For example, in Kaitetu (Seram) the inalienably possessed noun as well as the possessor pronoun take the same possessor suffixes, while the inalienables have just the free pronouns preceding the possessum (Collins 1983: 28):

- |      |    |                   |    |              |             |         |
|------|----|-------------------|----|--------------|-------------|---------|
| (15) | a. | <i>ale mata-m</i> | b. | <i>ale-m</i> | <i>luma</i> |         |
|      |    | 2SG eye-2SG.POS   |    | 2SG-2SG.POS  | house       |         |
|      |    | 'your eye'        |    | 'your house' |             | Kaitetu |

The former construction may be characterised as a 'direct' possessive construction, but the latter is obviously not an 'indirect' one. In various other Austronesian languages in

the Moluccas and on Timor, e.g., Selaru (Tanimbar Archipelago), Kei, Buru, (Grimes 1991: 283, 331) and Kemak (East Timor), the same or a similar situation is observed.

Yet frequently different combinations are found, as for instance in Lakalei (Hull 2001a: 123–125) where the inalienables take a possessor suffix as well as a preposed possessor pronoun, and the alienables take a postposed possessive pronoun:

- |      |    |                |    |                  |         |
|------|----|----------------|----|------------------|---------|
| (16) | a. | <i>au amak</i> | b. | <i>arbau auk</i> |         |
|      |    | 1sg father:1sg |    | buffalo 1sg      |         |
|      |    | 'my father     |    | 'my buffalo'     | Lakalei |

An entirely different way to make the distinction is found in Alorese, the indigenous Austronesian language of Alor/Pantar, which marks the third person singular differently depending on whether the possessive relation is alienable (*ni*) or inalienable (*no*) (Klamer, 2007):

- |      |    |                                               |         |
|------|----|-----------------------------------------------|---------|
| (17) | a. | <i>Sa no maring ni ahho: “..”</i>             |         |
|      |    | CONJ 3SG say 3SG.POS dog                      |         |
|      |    | 'Then he told his dog: “..”                   |         |
|      | b. | <i>Pada hal, kujo ha gaki no leing terus.</i> |         |
|      |    | in fact crab this bite 3SG.POS leg then.      |         |
|      |    | 'In fact, the crab did bite his leg.'         | Alorese |

Recall also Taba in example (2) that does not mark alienable possession morphologically and has a possessive prefix only for inalienables. However, word order in both constructions is the same: possessor-possessum (Bowden 2001: 233–234).

By contrast, in the Austronesian languages to the east, in particular in the Cenderawasih Bay, the possessor of inalienable nouns is prefixed to the possessum, as for instance in Biak and Ambai, spoken on Yapen island, and also in Waropen. The distinction between 'direct' possession for inalienables and 'indirect' possession for alienables in these languages often does hold. In Biak (18) (Van Hasselt 1905: 37) and Ambai (19) (Silzer 1983: 89), non-singular possessors of inalienable objects are expressed by prefixes on the possessum while singular possessors give suffixes, as in (18a) and (19a) respectively.<sup>11</sup> The corresponding (b) examples give the alienable constructions that in both languages involve possessive ligatures, again prefixed when the possessor is plural. In Biak this ligature is of obvious verbal origin:

- |      |    |                  |                 |                   |
|------|----|------------------|-----------------|-------------------|
| (18) | a. | <i>Sno-ri</i>    | <i>sno-m-ri</i> | <i>ko-sno-sna</i> |
|      |    | name-1,3SG       | name-2SG-SG     | 1INC-name-PL      |
|      |    | my, his/her name | your name       | our names         |

11. In Biak, first and third person singular are identical and the second person singular is distinguished only by the addition of the Austronesian cognate form *-m*. Number of the possessum is marked by a suffix.

- b. *Awa ko-be-na na mbra.*  
 mango 1INC-possess-3PL.INAN 3PL.INAN ripe  
 Our mango's are ripe. Biak
- (19) a. *awe-ku awe-mu ta-nu-mi*  
 foot-1SG foot-2SG 1INC-head-PL  
 'my foot' 'your foot' 'our heads'
- b. *ne-ku wa ne-mu fian ta-ne romi*  
 POS-1SG canoe POS-2SG food 1INC-POS garden  
 'my canoe' 'your food' 'our garden' Ambai

Waropen, a third Cenderawasih Bay language, has a direct possession construction with prefixes for inalienables, as in (20a) (Held 1942: 48). The same forms are also used with a possessive morpheme in indirect possession constructions for alienable nouns, as in (20b) (Held 1942: 45):

- (20) a. *ra-worai* b. *ra-i ruma*  
 1sg-leg 1SG-POS house  
 'my leg' 'my house' Waropen

This arrangement is similar to various languages of the eastern Bird's Head, Hatam, Meyah and Sougb (Reesink 1999, 2002a: 217; Gravelle 2004).

Summing up, we find that the Papuan languages virtually all have a distinction between alienable and inalienable possession and that, furthermore, like the Oceanic languages the former tend to be expressed in 'direct' possessive constructions, and the latter – to a somewhat lesser degree – in 'indirect' possessive constructions. For the Austronesian languages of East Nusantara, the easternmost languages conform to this Papuan pattern, while the languages furthest west, such as Keo, Bimanese and Kambera, and north, such as Tukang Besi and Muna, are like the western Austronesian languages lacking this distinction altogether. In between, i.e., the area between Timor and the Bird's Head, we find languages that do distinguish alienable and inalienable possession but show variability in the expression of this distinction. Blust (1978, 1993) and Lichtenberk (1985) have argued that the 'direct' possessive construction for inalienable possession is an innovation in the CEMP group of Austronesian languages. On the basis of the data reviewed in this section, however, we find that rather as an innovation in the CEMP group as a whole, it appears that it is only a subset of the CEMP languages in East Nusantara that adopted the distinction, as a result of contact with Papuan languages. For the Austronesian languages on islands west and north west of Timor/Alor/Pantar there is no evidence of contact with Papuan populations. Thus, the data presented in this section support the hypothesis that the alienable-inalienable distinction (with the concomitant 'indirect' and 'direct' constructions) is an areal feature of East Nusantara and the Bird's Head that diffused from the Papuan languages to the Austronesian ones. It was in this area that the distinction entered the precursor language(s) of Proto-Oceanic (Ross 2001:138), making it a distinctive feature of the Oceanic languages today.

#### 4.1.2 *The order of possessor and possessum*

An old diagnostic characterizing the non-Austronesian languages of the Moluccas and the Bird's Head is the possessor (ligature) possessum order in a possessive noun phrase. Cowan (1953: 10) mentions compounds like 'chin-hair' for 'beard' as evidence for what he calls the 'Papuan genitive construction' in various languages of the Bird's Head, and Van der Veen (1915: 92–95) gives a Galela example (21) as illustrative for the North Halmahera languages:

- (21) *O baba awi tahu*  
 ART father 3SG.M.POS house  
 'father's house' Galela

This order contrasts with what is found in the Western Austronesian languages, as for example in Standard Indonesian *rumah saya* '(lit. house I) my house' where the possessor follows the possessum. This is what is found also in the Austronesian languages of the western part of East Nusantara, such as Kambara, whether the possessor is a noun (22) or a pronoun (23):

- (22) *Uma tau*  
 house person  
 'someone's house' Kambara
- (23) *Uma-nggu (nyungga)*  
 house-1SG.POS I  
 'my house' Kambara

In the eastern part of East Nusantara, however, we find the Papuan order also in the Austronesian languages. This so-called 'reversed Genitive' has been a long-standing topic in comparative work of the Moluccan languages. A central question has been whether it should be taken as a diagnostic for genealogical subgrouping of Austronesian languages (e.g., Brandes 1884), or whether it is a clear non-Austronesian feature (e.g., Van der Veen 1915). Grimes (1991: 287, 495–506) suggests that the reversed Genitive order is due to contact with non-Austronesian languages of the area, and Himmelmann (2005), likewise, uses the 'preposed possessor' order as a typological, rather than a genealogical, feature of some Austronesian languages in Eastern Indonesia (recall Table 1). Recently, several authors have argued that the 'reversed Genitive' should be seen in connection with the semantics and expression forms of the possessor. For example, Collins (1983: 27–29) argues that it is important to make a distinction between alienable possession, which typically has Possessor-Possessum order, and inalienable possession with Possessum-Possessor order. Similarly, the order may depend on the expression of the possessor as a pronoun or a full noun phrase, and in some cases on yet other factors.

Before we link the word order in the possessive construction to contact, genealogy, or general typology, let us first examine the languages in some detail. In the Papuan languages of East Nusantara, regardless of basic constituent word order in the clause, the possessor occurs before the possessed noun, at least whenever the possessor

is expressed by a full noun (phrase) and the possessive relation is ‘inalienable’. For example, SOV languages like Teiwa (24) on Pantar (Klammer, forthcoming), and Adang on Alor (Haan 2001: 163; Stokhof 1975: 20), Galela in the North Moluccas (cf. example (21) above), and Inanwatan in the Southern Bird’s Head (De Vries 2004: 64), and Yawa in the Cenderawasih Bay (Jones 1986: 47) all have this order:

- (24) *Iman ga-yivar ga’an un tei luxun goxu’ pati.*  
 3PL 3-dog that.one ASP tree high bark PROG  
 ‘Their dog is barking up a tree.’ Teiwa
- (25) *Jadi suda órewo agá aiba-séro iko-we-ge-i.*  
 so allright woman POS voice-word follow-3SU-do-PAST.M  
 ‘So, allright, he followed the instructions of the woman.’ Inanwatan
- (26) *Natanyer apa-jaya Ø-awabe-to.*  
 Nathaniel 3SG.M.POS-father 3SG.M-yawn-PERF  
 ‘Nathaniel’s father is yawning.’ Yawa

But also the Papuan SVO languages, such as Tidore in the North Moluccas (Van Staden 2000: 250), and most languages of the Bird’s Head, illustrated here by Abun (Berry & Berry 1999: 82), have possessor-possessum order:

- (27) *Cole ma-giba yo-foluji.*  
 bra 3NH.POS-strap 3NH-come.loose  
 ‘The bra strap has come loose (by itself).’ Tidore
- (28) *an bi nji bi nggon bi nu*  
 3SG POS brother POS wife POS house  
 ‘his brother’s wife’s house’<sup>12</sup> Abun

In Maybrat, in the centre of the Bird’s Head, only the inalienable possessive constructions conform to the possessor-possessum order. In the case of alienable possession, the possessor is in post-nominal position linked to the possessum by an invariant relator *ro*, which is also used to introduce relative clauses (Dol 1999: 149):

- (29) a. *fnia m-ao* b. *amah ro t-atia*  
 woman 3U-foot house POS 1SG-father  
 ‘the woman’s foot’ ‘my father’s house’ Maybrat

This order need not have the same origin in all Papuan languages. First of all, it appears that there is a universal preference for the ‘possessor-possessum’ order, despite further typological characteristics of languages. It is more common to find the possessor preceding the possessum in SOV languages, than it is to find the possessum before the possessor in SVO languages. This preference has been

12. Abun *bi* is most likely of Biak origin: *ve* ‘possess’ has infix *y* for 3SG: *vye*; compare Abun loan verbalizer *bi* ~ and Biak verbalizer *ve*.

related to two principles governing word order in many languages: animate before inanimate and definite before indefinite (Heine 1997: 135, cf. also Clark 1978). The tendency for possessor-possessum order in the Papuan languages may simply be in line with this. However, some languages, like Tidore, show indications of erstwhile SOV order and the placement of the possessor is thus in line with other placement patterns that accompany this dominant constituent order (cf. section 3.1 above). Furthermore, in languages like Hatam it may be the verbal origin of the possessive ligature that has led to the possessor-possessum order. In this language, the ligature is cross-referenced with a 'subject' marker co-referential with the possessor. The word order in the possessive construction then mirrors the order of subject, verb and object in the clause.

As indicated above, if the possessor is pronominal or if the possessed is inalienable, there may be deviations from this possessor-possessum order. Hatam and Meyah, for example, allow the alternative order possessum-possessor if the possessum is alienable, as in (30), as an alternative to the far more frequent pre-nominal position if the possessum is alienable, in (31) (Reesink 1998: 623, 1999: 81; Gravelle 2004: 278):

- (30) *Munggwom ji-de=nya i-pim mindei i?*  
 child 2PL-POS=PL 3PL-cry what Q  
 'Why are your children crying?'  
 (31) *Ji-de munggwom=nya i-pim mindei i?*  
 2PL-POS child=PL 3PL-cry what Q  
 'Why are your children crying?' Hatam  
 (32) *Eita meiteb (ongga) efen.*  
 give machete (REL) 3SG.POS  
 'He/she gave (him) the machete (that) he/she owned' Meyah

The motivation for the word order differences in Hatam and Meyah is still unclear, possibly related to a relative clause construction, as argued by Gravelle (2004: 278), but in some other languages emphasis may play a role. For instance, Tehit and Moi (Menick, n.d., Reesink 1998: 622–624) allow for a pronominal possessor after the possessum, as in the following example, for reasons of emphasis:

- (33) *Na-saalek n-e-kuwok n-a ø-o-w se!*  
 2SG-carry 2SG-POS-stringbag 2SG-POS NEAR.ADDR-SG-3SGM/NH PERF  
 'Carry your (own) stringbag!' Moi

In the Austronesian languages of East Nusantara possessor-possessum order is also frequently attested, although there are also languages with possessum-possessor as the only option. Languages of the latter type include the Austronesian languages in the western part of East Nusantara, such as Kambera (recall examples (22) and (23) above), but also, rather exceptionally, Biak spoken East of the Bird's Head. Possessive phrases clearly have a verbal origin in this language, consisting of a verb with a prefix

cross-referencing the possessor and a suffix marking number and gender of the possessum (Van Hasselt 1905: 37).<sup>13</sup>

- (34) *Awa ko-be-na na mbra.*  
 mango 1INC-possess-3PL.INAN 3PL.INAN ripe  
 'Our mangos are ripe.' Biak

When the possessor is a noun phrase, either possessor (35) or possessum (36) may come first, although the possessive pronoun (or cross-referenced ligature) is always in a phrase-final position (Steinhauer 2005; Van den Heuvel 2006):

- (35) *Ya-mam snon=ya rum v(y)e=di.*  
 1SG-see man-3SG.DEF house <3SG>-possess=3SG.DEF  
 'I see the house of the man.'
- (36) *Ya-mam rum snon=ya v(y)e=di.*  
 1SG-see house man-3SG.DEF <3SG>-possess=3SG.DEF  
 'I see the house of the man.' Biak

Other languages that have alternative orders are Tetun Fehan (Van Klinken 1999: 142–143) and Tetun Dili (Williams-van Klinken et al. 2002: 33–35) on Timor, the latter only for alienables. However, the reversed genitive represents over 80% of the textual examples in the corpus of Tetun Fehan, and pronominal possessors also virtually always precede the possessum in everyday Tetun Dili:

- (37) *tumukun nia=kan<sup>14</sup> fé=n*  
 village.head 3SG=POS wife=GEN  
 'village head's wife' Tetun Fehan
- (38) *asu ó=k*  
 dog 2SG-POS  
 'your dog' Tetun Fehan
- (39) a. *João nia liman \*liman João nian*  
 John POS hand hand John POS  
 'John's hand'
- b. *João nia uma uma João nian*  
 John POS house house John POS  
 'John's house' 'John's house' Tetun Dili

In Tetun Fehan, the structure of the two possessive constructions is rather different. In the case of a preposed possessor, the clitic =*kan* is virtually obligatory but it does not

13. When treated as verbal constructions, they display the also unusual SOV or even OVS and OSV constituent orders in this otherwise SVO language. The word order in the possessive construction is therefore remarkable in more than one way.

14. The possessive clitic =*kan* is optional but strongly preferred for pre-nominal possessors

occur with the postposed possessor, which requires  $=k$ , as in (38). Furthermore, inalienably possessed items (kinship terms, part-whole relations, and terms such as ‘house’) take a genitive clitic  $=n$  on the possessum, when the possessor precedes the possessum, as in (37), but not when it follows. In Tetun Dili, the postposed possessor usually expresses a more general relationship of association between the possessor and possessum.

The reversed genitive is found as the only option in Alorese and Alor Malay, the Austronesian languages of Pantar and Alor, as well as in the languages of central Timor, such as Idate and Mambai (Klamer, field notes 2003). In the Moluccas, Leti (Van Engelenhoven 1995: 170) has this order (43) as well as Buru (Grimes 1991: 282), Dobel (Hughes 2000: 146), Tugun on Wetar (Hinton 2000: 116), Bandanese (Collins & Kaartinen (1998: 536)), Kei (Geurtjens 1921: 19) and Taba (Bowden 2001: 230). Further to the east, it is found in Wandamen (Cowan 1955: 47), Ambai (Silzer 1983: 124), and Waropen (Held 1942: 44–49). As such it is clearly the dominant pattern in the Austronesian languages of East Nusantara:

- (40) *au ama*  
1SG father  
‘my father’ Mambai
- (41) *Ni ning laffo juang.*  
3SG POS village far  
‘His village is far (from here).’ Alorese
- (42) *ni mahina-n*  
3SG wife-POS  
‘his wife.’ Idate
- (43) *puat-e lavar-ne*  
woman-DEI sarong-POS  
‘the woman’s sarong’ Leti

Both typologically and genealogically the possessor-possessum order in the Austronesian languages of East Nusantara is unexpected since SVO languages tend to have the order possessum-possessor, as indeed the Western Austronesian languages do. As with the Papuan languages, a general preference for this order typologically or the verbal nature of the ligature are possible explanations. For the latter, consider Buru (Grimes 1991) in which the possessive construction is ambiguous between a predicative and attributive reading. Grimes (1991: 287) gives the following schema with just one translation, but an alternative reading is ‘my house’:

SUBJECT <sub>Actor</sub>	VERB	OBJECT <sub>Undergoer</sub>
Possessor	Inflected possessive word	Possessed
<i>Yako</i> 1SG ‘I have/own a house’	<i>nango</i> 1SG.POS	<i>huma saa</i> house one



Yet, if we consider the geographical distribution of the languages with this unexpected order, we find that these languages are all spoken in close proximity to Papuan languages, or in places where historically Papuan influences may have been strong: Timor-Alor-Pantar in the west, North Halmahera in the north, and the Bird's Head in the east. Kambera, as a language that has the 'Western Austronesian' possessum-possessor order, significantly does not border a Papuan language. The only exception is Biak, which prefers the 'Western Austronesian' order, despite its geographical location and intense Papuan contact evidenced in other parts of the grammar. The languages with this order also all have the distinction between alienable and inalienable possession, which may also be traced to Papuan contact. This distribution confirms Grimes' (1991: 292) hypothesis that the "Austronesian languages [calqued] on the order of the genitive construction of languages in the area prior to the arrival of the Austronesians." The evidence for a 'contact' rather than 'innovation' account is strengthened further if we also take into consideration the placement of possessive affixes and ligatures. It may be assumed that these are more resilient to change than the possessor noun (phrase) and may reveal 'older' stages of the languages.

#### 4.1.3 Possessive ligatures and affixes

So far we have seen that both the alienability distinction and the order of the possessor and possessum in the possessive phrase are striking features in the Austronesian languages of East Nusantara that are not easily accounted for by genealogy or typology. In this section we examine the position of possessive ligatures and affixes in particular in the Austronesian languages to demonstrate that these reveal remnants of an erstwhile possessum-possessor order, strengthening our claim that the current structure of the possessive constructions in the Austronesian languages of East Nusantara is due to contact with neighbouring Papuan languages.

Not all languages in East Nusantara cross-reference the possessor on the possessum, but for the languages that do we may find prefixes or suffixes. In addition, languages may (and indeed do) mark alienable and inalienable possession differently. An important difference between the Austronesian and the Papuan languages of East Nusantara is the position of the possessive affix or ligature. In the Papuan languages, this affix or ligature invariably occurs before the possessum, but in the Austronesian languages both orders are frequent. Recall examples (40) to (43) above, in which Leti and Idate have suffixes and Alorese has a ligature preceding the possessum. A complication is found in some of the Moluccan Austronesian languages, such as Buru (Grimes 1991: 279), Kei (Geurtjens 1921: 19) and Taba (Bowden 2001: 230), which have a ligature between the possessor and the possessum, like many of the Papuan languages, but with a suffix marking the possessor rather than a prefix:

- |      |    |                       |    |                    |      |
|------|----|-----------------------|----|--------------------|------|
| (44) | a. | <i>yak ni-k mapin</i> | b. | <i>ni-m capeyo</i> |      |
|      |    | 1SG POS-1SG wife      |    | POS-2SG hat        |      |
|      |    | 'my wife'             |    | 'your hat'         | Taba |

Frequently, the position of the affix or ligature is different for alienables and inalienables. Bandanese (Collins and Kaartinen 1998: 536), for instance, has a single affix that suffixes to inalienably possessed nouns and prefixes to alienably possessed nouns:

- |      |    |                     |    |                   |           |
|------|----|---------------------|----|-------------------|-----------|
| (45) | a. | <i>ak mata-n(u)</i> | b. | <i>ak nu-rumo</i> |           |
|      |    | 1SG eye-1SG.POS     |    | 1SG 1SG.POS-house |           |
|      |    | 'my eye'            |    | 'my house'        | Bandanese |

East of the Bird's Head, Ambai (Silzer 1983: 124) has a similar phenomenon with ligature, *ne*, but now the possessor is suffixed when it is singular (46) and prefixed when it is plural (47):

- |      |    |                   |    |                   |       |
|------|----|-------------------|----|-------------------|-------|
| (46) | a. | <i>ne-ku wa</i>   | b. | <i>ne-mu fian</i> |       |
|      |    | POS-1SG canoe     |    | POS-2SG food      |       |
|      |    | 'my canoe'        |    | 'your food'       | Ambai |
| (47) | a. | <i>ta-ne romi</i> | b. | <i>e-ne munu</i>  |       |
|      |    | 1INC-POS garden   |    | 3PL-POS house     |       |
|      |    | 'our garden'      |    | 'their house'     | Ambai |

The morphological markers suggest that many of the Austronesian languages of East Nusantara have retained the suffixing nature of the possessor marking on the possessum or on the ligature while adjusting the word order in the construction to that of surrounding Papuan languages. Some of the Austronesian languages of the Cenderawasih Bay show a partial shift to prefixes, as for instance Biak and Ambai, which have suffixes for singular and prefixes for plural possessors; in others, like Waropen, the shift is complete. Because all the Oceanic languages have retained a cognate set of suffixes,<sup>15</sup> we know that the language(s) ancestral to Proto-Oceanic must have left the Cenderawasih Bay (Lynch, Ross & Crowley 2002: 57) before the shift in Waropen, Ambai and Biak took place. Considering both the placement of the possessor noun (phrase) and the ligature or affix, we can set up a scale from 'typically' (western) Austronesian, with Kambara as a clear representative, to 'typically' Papuan as in Alorese or in Bandanese alienable constructions. Languages that have preposed ligatures with person marking suffixes like Taba are also toward the Papuan end of the scale, while languages that have postposed ligatures or suffixes on the possessum are toward the Austronesian end.

Summing up, we find that although it is not striking to find the possessor-possessum order in SVO languages it is remarkable that this order is found i. often in combination with postposed ligatures or possessor suffixes on the possessum noun; ii. in alienable constructions rather than inalienable constructions; and iii. precisely in those Austronesian

15. Lichtenberk (1985: 112–113) notes that "there is overwhelming evidence that POC had possessive suffixes and not prefixes." Although he reports a "veritable embarrassment of riches" in Oceanic possessive suffixes, the reconstructed POC set is quite like that of Proto-Austronesian, with reflexes easily identified in Kaitetu of the Moluccas and various Oceanic languages of New Britain, the homeland of POC.

languages that have been in contact with Papuan languages. Significantly, it is not found in the languages that lack the alienable/inalienable distinction, which is another feature that may be traced to Papuan contact. Since the adaptations to the Papuan structure are clearest in areas with the strongest Papuan influences, and absent in parts where Papuan contact was limited or absent, we see the emergence of the typical East Nusantara possessive construction as an areal feature that is the result of contact rather than as a spontaneous innovation in a reconstructed ancestor of a genealogically defined group.

#### 4.2 Post-predicate negation

As mentioned in section 3.1, verb final or clause final negation is typologically associated with SOV constituent order. It is therefore an unexpected feature in Austronesian languages, which are predominantly SVO, while in Papuan languages, which are typically, though not exclusively, SOV it is expected to occur widely. Indeed, in the majority of Papuan languages negation is either expressed by a pre-verbal adverb or particle, or through some morphological modification of the final verb (Reesink 1998, 2002b). Clause-final negation is found in Papuan families belonging to the TNG, such as the Dani languages and, tentatively, the South Bird's Head languages such as Inanwatan, and the Papuan languages of the Timor area. It is also found further along the north coast of New Guinea (Sentani), in some of the Torricelli Phylum languages, both those with SVO and those with SOV, as well as in East Papuan languages. In Austronesian languages, where we almost universally find pre-verbal/pre-predicate or initial negators, final negation is not found, except in East Nusantara and in some Oceanic languages.

In East Nusantara, the typical TNG negation that involves a suffix on the verb is found only in Inanwatan. As the example shows, negative verb forms may optionally be preceded by an additional negative adverb *náwo* (De Vries 1996: 111):

- (48) (*Náwo*) *né-se-sa-aigo*  
 (not) 1SG.S-walk-FUT-NEG  
 'I am not going to walk.' Inanwatan

In most other Papuan languages, such as Yawa (Jones 1991: 102), of Yapen island in the Cenderawasih Bay, the negator appears to be a free form, always in sentence-final position:

- (49) *Yancea beare ruwijirati bauname joen?*  
 Yance be why marry NEG  
 'Why hasn't Yance married yet?' Yawa

This goes for both the SVO languages of the Bird's Head, such as Maybrat (50) (Dol 1999: 127) and Hatam, and for the SOV languages of Timor such as Bunak (51) (Friedberg 1978: 36, 57), Teiwa and Lamma (52) of Pantar (Nitbani et al. 2001: 90, cf. 93, 134, 135):

- (50) *Ana m-amo Kumurkek fe.*  
 they 3PL-go Kumurkek NEG  
 'They do not go to Kumurkek.' Maybrat

- (51) *Halali a niq, halali a loi niq.*  
 3DU eat NEG 3DU eat good NEG  
 'They two don't eat, they two can't eat.' Bunak
- (52) *Nang sakka kauwa.*  
 I strong NEG  
 'I am not strong.' Lamma

There is one language family that shows some deviation from this pattern: the North Halmahera family. Although these languages are generally described as having clause final negation (Van der Veen 1915: 98), the data available for some of these languages suggest that the negative element is primarily attached to the verb, with some material possibly following, as in Pagu (Wimbish 1991: 57):

- (53) *Muna ma-oko-wa-si ma gasi-ko.*  
 3SG.F 3SG.F-seawards-NEG-still ART sea-seawards  
 'She hasn't yet gone seawards to the beach.' Pagu

It appears that this language initially had the same structure as Inanwatan with a verbal suffix that in a verb final language will always occur in clause final position. When a language becomes less rigidly verb final, there are two possibilities: the suffix retains its position on the verb, as appears to be the case in Pagu, or the suffix retains its position in the clause detaching itself from the verb. Indeed, in related Tidore, which has completely shifted to SVO, the negator is a separate morpheme that occurs rigidly clause-finally. (Van Staden 2000: 232).

Like Inanwatan, this last language also has double negation. The preverbal element *kama* places focus on the negated verb and is often used to disambiguate the scope of negation in complex clauses (Van Staden 2000: 235–240):

- (54) *Fangato kama hoda mina mo-oro nyao ua.*  
 1SG.M NEG see 3SG.F 3SG.F.A-take fish NEG  
 'I did not see her steal the fish.'
- (55) *Fangato hoda mina kama mo-oro nyao ua.*  
 1SG.M see 3SG.F NEG 3SG.F.A-take fish NEG  
 'I saw she did not steal the fish.' Tidore

Such optional preverbal adverbs are also found in Abun of the Bird's Head (Berry and Berry 1999: 135; Reesink 2002b: 255) and Adang of Timor. The final negator may occur alone, as in (56a), or with a preverbal negator that may be omitted, as in (56b). Since the postverbal negator is always obligatory, we assume that this is the basic one (Haan 2001: 75, 76):

- (56) a. *Ince (è) sam don nene.*  
 Ince NEG go shopping NEG  
 'Ince did not go shopping.'
- b. *\*Ince è sam don*  
 Ince NEG go shopping Adang

Most of the Austronesian languages of East Nusantara have the typical Austronesian pre-verbal/pre-predicate/clause-initial position for their negators. For example on Timor, Tetun Fehan (Van Klinken 1999: 228), Tetun Dili (Williams-van Klinken et al. 2002: 86), Mambai (Hull 2001b: 10), Tokodede, Kemak, and Lakalei (Hull 2001a: 171–173) all have preverbal negators, as do Leti (Van Engelenhoven 1995: 213) in the south west Moluccas and Dobel (Hughes 2000: 162) in the southeast Moluccas. Further west, Kambara on Sumba (Klamer 1998: 77, 107–108, 142), Bimanese on Bima (Owens 2000: 127–137) and Keo on Flores (Baird 2001: 339) all have similar negation strategies:

- (57) *Ami la h' osan.*  
 1PL.EXC NEG have money  
 'We don't have money.' Tetun Fehan
- (58) *Muani ta=na-natu surt-e.*  
 man:INDEX NEG=3SG-send:INDEX letter-INDEX  
 'The man did not send the letter.' Leti
- (59) *Nda ku-hili beli-ma-nya-pa*  
 NEG 1SG.N-again return-EMPH-3SG.D-IMPF  
 'I am not going back to him again.' Kambara

However, East Nusantara also hosts a significant number of Austronesian languages with clause-final negators. Geographically, most of these languages are spoken in the area close to the Bird's Head or otherwise in close proximity to Papuan languages. For example, the Austronesian languages on Timor do not have final negation, but Alorese, the only indigenous Austronesian language in the Alor/Pantar area does (Klamer, 2007):

- (60) *Fe guo guo, matto oro uttang unung tapi fe dapa lahhe.*  
 they call call frog at forest inside but they find NEG  
 'They called and called the frog in the woods but they didn't find it.' Alorese

And in the Cenderawasih Bay and in the 'neck' of the Bird's Head, as well as in the Bomberai Peninsula, south of the Bird's Head, most Austronesian languages have final negation. Examples are Biak (see Reesink 2002b: 249), Mor (61) (Laycock 1978: 300), Ambai (62) (Silzer 1983: 215), Waropen, Wandamen-Windesi (Cowan 1955: 58) and Irarutu (Matsumura and Matsumura 1990: 9):

- (61) *Igwa-n rueta va.*  
 1SG-eat something NEG  
 'I do not eat something.' Mor
- (62) *Y-okon dian we Yani kaka.*  
 1SG-give fish to Yani NEG  
 'I didn't give any fish to Yani.' Ambai

In the Moluccas the picture is mixed with many languages with preverbal negation, but on the island Makian, west of Halmahera, Taba has clause final negation (Bowden

2001: 335), as do the Central Moluccan languages Buru (Grimes 1991: 232) and Alune (Florey 2001: 100), and South Moluccan Kei (Geurtjens 1921):

- (63) *Nik calana kuda-k asfal te.*  
 1SG.POS trousers be.black-APPL bitumen NEG  
 ‘My trousers are not blackened with bitumen.’ Taba
- (64) *A=a-somi be dana mo.*  
 2SG=CAU-embarrass COMP take NEG  
 ‘Don’t be embarrassed to take it.’ Alune

In the majority of cases, the forms of these negators appear to come from two different origins: from *\*ba ~βa ~(u)wa*, found in both Austronesian and Papuan languages and from Proto-Austronesian (PAN) *\*ta*. Which negator a language has is not necessarily predictable on the basis of its genealogy. For instance Waropen and its dialect Waropen-Napan have forms that can be traced to different origins: Waropen has *afa ...womo*, possibly cognate with the former (compare also Buru *moo* and Alune *mo* below), and Waropen Napan *ambe...te*, clearly cognate with the latter (Held 1942: 80–81):<sup>16</sup>

- (65) *Yenggea ambe i-totonako te.*  
 my.leg NEG it-hurt NEG  
 ‘My leg doesn’t hurt.’ Waropen-Napan

Similarly, in the south Moluccas, one Kei dialect has *dem* ‘not’ while another has *waeid* (Geurtjens 1921: 38). The former is again a reflex of PAN *\*ta*; the latter is a cognate form of *ba ~βa ~(u)wa*:

- (66) *U-ba nangan waeid*  
 1SG-go forest NEG  
 ‘I don’t go to the forest.’ Kei

In other words, this typically Papuan feature occurs in a number of Austronesian languages in New Guinea, the Moluccas, including Halmahera, and Alor/Pantar. Reesink (2002b), therefore, hypothesizes that clause-final negation originates from the verb-final Papuan languages. It has been retained in Papuan languages that changed to SVO word order, and has diffused into Austronesian languages in East Nusantara. We have both linguistic and historical evidence to believe that this is indeed the case. The historical evidence is that speakers of Austronesian and Papuan languages in the Moluccas and Papua have been in contact for a long time, as discussed in sections 2.2–2.2.3. As additional linguistic evidence for diffusion there is the sound correspondence *ba ~βa ~(u)wa* which, as Reesink (2002b) points out, links the negative markers *bar* and *big* in non-Austronesian Mansim and Hatam (Bird’s Head) and *(u)wa* in the non-Austronesian

16. Waropen and its dialect Waropen-Napan are the only Austronesian languages with double negation.

languages of North Halmahera to  $\beta a$  in Austronesian Biak, Mor and Wandamen in the Cenderawasih Bay region, and to various other languages in the Moluccas, such as Kei, and possibly even Buru and Alune. At the same time, a number of Austronesian languages in East Nusantara, and one Papuan language, Abun in the northwest of the Bird's Head, have reflexes of PAN  $*ta$ . It is plausible that one of the negative morphemes reconstructed for Proto-Oceanic,  $*b^w\text{ali}$ , with a truncated  $*b^w\text{a}$  (Lynch, Ross & Crowley 2002: 88), has its origin in the contact area of West Papuan and Austronesian in Maluku and the Bird's Head area, while POc  $*ta$  is the canonical Austronesian negative morpheme.

### 4.3 Tone

Austronesian has a number of branches with tonal languages, most notably the Formosan languages, but it is a phenomenon not at all associated with the CEMP languages. Lexical tone is weakly linked to Papuan languages to the extent that tonal languages are found in various Papuan families. In East Nusantara we find two Papuan languages that clearly have tone: isolates Mpur and Abun. Mpur has four lexical tones, a fifth complex contour tone is a phonetic compound of two lexical ones (see Odé 2002b). Abun is claimed to have three lexical tones, but only two of these (low and high) still provide minimal pairs, e.g., high marks plural, as in *an* '3sg' and *án* '3pl'. Their functional load is low and it is clear that they are in the process of disappearing (Berry & Berry 1999: 20). Meyah and Sougb of the East Bird's Head family have a few contrasts in tone. On some monosyllabic words two contrastive tones can be identified, while on polysyllabic words, only the stressed syllable receives high tone. These languages could be analysed as either tone languages, since there is a clear contrast in tone (cf. Van Zanten and Dol, to appear), or as pitch-accent languages, since the function of tone is restricted to signalling stress in polysyllabic words. This is the position taken in Reesink (2002a) for Sougb, Gravelle (2002) for Meyah, and Donohue (1997: 366) in his overview of tone systems in Papuan languages.

No tone or pitch distinctions are found in the other Papuan languages of the Bird's Head,<sup>17</sup> or indeed in any of the other Papuan languages of East Nusantara. Yet, just west of the Bird's Head on the Raja Ampat islands, there are two Austronesian languages, Ma'ya and Matbat, with tonal systems as in Mpur. For Matbat (2001: 102), Remijsen analyses five lexical tones, and for Ma'ya (2001: 119), two contour tones and one level tone. Both languages also have some Papuan vocabulary, although Matbat more so than Ma'ya. The non-Austronesian lexical items in Matbat and Ma'ya do not all have clear correspondences in the Papuan languages of the Bird's Head,

---

17. Dol (1999) and Odé (2002b) report that going by native speaker claims there may be some remnants of tone in Maybrat, but experiments and measurements did not reveal any tonal distinctions.

which is not so strange, given the lexical diversity among the languages of the peninsula. Nevertheless, there are a few items in the 100-item Swadesh lists, provided in Remijsen (2001: 140–153), that are suspiciously similar to the equivalents in precisely the two tonal languages Abun and Mpur, which are, furthermore, not related. For instance, the word for ‘sago’ is *bi*<sup>[high]</sup> in Ma’ya, *bei* in Abun and *bi*<sup>[low]</sup> in Mpur. Remijsen suggests that the tone systems are a remnant of a presently extinct Papuan tone language spoken in the Raja Ampat archipelago at the time of the earliest Austronesian arrival. Another possible Austronesian tone language in East Nusantara is reported in Laycock (1978: 290, cf. also Odé 2002b: 8). This language, Mor, is spoken in the Cenderawasih Bay and supposedly has two tones, but the language is almost extinct and the status of these tones at present or in the past is unclear. A tonal area may then be established stretching from the Raja Ampat islands with Matbat and Ma’ya, the northern coastal languages of the Bird’s Head Abun and Mpur, and perhaps including the East Bird’s Head family in the southeast and even possibly Mor in the Cenderawasih Bay.

## 5. Discussion and conclusion

In East Nusantara we are dealing with a heterogeneous group of Papuan languages (2.3.1) that have a number of features in common with the Austronesian languages that neighbour them: (1) the possessor-possessum order in adnominal possession, (2) the overt marking of the distinction alienable vs. inalienable possession, and (3) clause-final negation. While these features are not generally found in Western Austronesian (see Himmelman, as given in Table 1), they do occur in many Austronesian languages in East Nusantara (see Table 3). Tone, finally, has diffused to just a few Austronesian languages (Ma’ya and Matbat on the Raja Ampat islands).

At the same time, we identified a number of Austronesian areal features as typical for Austronesian languages, but uncommon in Papuan languages: (1) SVO as primary constituent order, and (2) an inclusive/exclusive opposition. In East Nusantara they are both attested in a number of Papuan languages (see Table 4):

Clearly, the features in Table 3 and Table 4 do not all converge on the same isoglosses. However, all five features overlap in Halmahera and the Bird’s Head; four (3 non-Austronesian, 1 Austronesian) overlap in Alor/Pantar, the Moluccas, Halmahera, and the Bird’s Head & surroundings; while three (2 non-Austronesian, 1 Austronesian) overlap in Timor, Alor/Pantar, the Moluccas, Halmahera, and the Bird’s Head & surroundings. Together, the features define a linguistic area that has Halmahera and the Bird’s Head as its core, and radiates outwards to first include the Moluccas and Alor/Pantar, and then Timor. The fact that the Austronesian languages of East Nusantara agree in a number of features with the diverse Papuan languages in this area could be due to various contact scenario’s, since we find Austronesian languages with Papuanisms and Papuan languages with Austronesian traits.



**Table 3.** Non-Austronesian features in Austronesian languages of East Nusantara and the Bird’s Head

non-Austronesian Feature	Austronesian Languages with this feature according to region
Possessor-Possessum	Alor/Pantar (Alorese) Timor (e.g., Tetun Fehan, Tetun Dili, Idate, Mambai) Central and South Moluccas (e.g., Leti, Buru, Dobel, Wetar, Bandanese, Kei) Halmahera (e.g., Taba) Cenderawasih Bay (e.g., Wandamen, Ambai, Waropen)
Clause-final negation	Alor/Pantar (Alorese) Central and South Moluccas: all (e.g., Buru, Alune, Kei) Halmahera (e.g., Taba) Cenderawasih Bay (e.g., Biak, Irarutu, Ambai, Mor, Waropen). Not found in Timor languages
Alienable/inalienable	Alor/Pantar (Alorese) Timor (e.g., Tetun Fehan, Tetun Dili, Lakalei, Isní, Lolein, Kemak, Waimaha) Central and South Moluccas (e.g., Kaitetu, Selaru, Kei, Buru) Halmahera (e.g., Taba) Cenderawasih Bay (e.g., Biak, Ambai, Waropen)

**Table 4.** Austronesian features in non-Austronesian languages of East Nusantara and the Bird’s Head

Austronesian Feature	non-Austronesian languages with this feature according to region
SVO constituent order	Some Halmahera languages (e.g., Sahu, Ternate, Tidore, West Makian) but not all Bird’s Head, except the SBH languages (e.g., Inanwatan) Not in Alor/Pantar, Timor
Incl/excl distinction	Alor/Pantar (e.g., Teiwa, Lamma, Blagar, Adang, Abui, Kui, Klon, Kafoa, Hamap, etc.) Timor (Bunak, Makasai) Halmahera (e.g., Tidore) Bird’s Head, except for three isolates in the center: Maybrat, Abun, Mpur

Considering these data, the first observation to make is that contact is not a one-way process. The Austronesian and Papuan languages have influenced each other. At the same time it is also clearly not the case that the typology of these languages is the result of one ‘clash of civilisations’ between the newly arrived Austronesians and the original Papuan communities. East Nusantara has been, for centuries, a highly dynamic area with a long history of migration, intense (slave) trade, and many shifts in power and dominance. This is reflected in the complexity of its linguistic situation.

It is difficult to date the various changes in the individual languages, but there is one point of departure that we can use, which is the origin of Proto Oceanic. East Nusantara is the link between Proto Malayo-Polynesian (PMP) and Proto Oceanic. As we described in section 2.3.2, the Oceanic languages are a direct descendant of the Eastern Malayo Polynesian languages, just as the SHWNG languages like Biak, Taba and Waropen. But although the ancestors of the Proto Oceanic speakers probably lived in the Cenderawasih Bay area (Blust 1978; Lynch et al. 2002: 57), too little was known about the Austronesian languages of East Nusantara, i.e., the Central Eastern Malayo Polynesian languages, to use as a basis for reconstructing Proto Oceanic. For this reason, traditionally PMP is used to examine innovations in Proto Oceanic. Yet, on the basis of the present study we can now examine which of the typical characteristics of the East Nusantara languages that are not found in Western Austronesian languages do again occur in the Oceanic languages, or indeed in reconstructed POc. If a change is present in POc, we have a fair indication that the change occurred, at least in some languages, prior to the peopling of Oceania.

The first of these characteristics is the possessive construction. We noted two important aspects of this construction in the East Nusantara languages. The first is the introduction of the alienable-inalienable distinction, and a concomitant distinction between direct and indirect possession, which has occurred in the vast majority of the Austronesian languages of East Nusantara. The second is the change in order from possessed-possessor to possessor-possession. Here we find variation although there is a clear trend towards possessor-possession. In particular in those languages that are spoken in areas with many Papuan languages, e.g., North Halmahera and Papua, the order is typically possessor-possession. When we then consider the Oceanic languages and reconstructed POc, it appears that the alienability distinction and direct-indirect possession is present in virtually all languages and it is almost certainly a characteristic of POc (Lynch et al. 2002: 69). However, when we look at the actual order of the possessor and the possessum, the Oceanic languages, despite some variation (e.g., in Fijian and in some Western Melanesian languages), favour possessor suffixes in direct possession and possessive classifiers with suffixes following the possessed noun in indirect possession. Possessor nouns, similarly, follow the possessed (Lynch et al 2002: 40). It appears then that the changes in order have occurred over a long period of time, and are in fact still ongoing, while the introduction of the alienability distinction was a very early influence from the Papuan languages on Austronesian.

The second was the position of the negator clause finally, or at least: after the predicate. Western Austronesian languages typically have pre-predicate negation, but in East Nusantara we find post-predicate negation in many languages. In section 4.2 above, we also showed that the forms of the negators in the Austronesian languages reflected two sources. One, *\*ta*, is clearly Austronesian and typically occurs in pre-predicate position, both in the languages of East Nusantara and in the Oceanic languages; the other *\*ba ~ʃa ~(u)wa* is found also in the Papuan languages of North Halmahera and Papua (not in those of Alor/Pantar and Timor) and typically occurs clause finally. In their

reconstruction of POc, Lynch et al. (2002: 88) give *\*b<sup>w</sup>ali* or *\*b<sup>w</sup>a* as a POc negative verb, and we suggest that these have the same origin. Unlike *\*ta*, this negative verb typically occurs in post-clausal clausal position, although some pre-clausal negators are also found (Lynch et al. 2002: 91). Yet, while the post-predicate position of this negator in the Papuan languages makes perfect sense, since they are – or were – SOV, for the verb initial Oceanic languages, the typical post-predicate position of this particular negator is odd. Lynch et al. (2002: 88) explain the position of this negative verb as a later development that occurred when individual Oceanic languages changed from verb initial to verb medial or even final, but on the basis of our findings we propose that an alternative account could be that the POc copied the form as well as its position in the clause from a Papuan source. A shift to a clause initial position in some Oceanic languages is the result of a later reanalysis of this verb. Further evidence enhancing this view is that in the Oceanic languages, the negative verb virtually never has person marking.

In brief, we suggest that two of the three Papuan features of East Nusantara occurred already in at least some of the Austronesian languages before Proto Oceanic emerged and split up, about 3,500 years ago. The shift in order in the possessive construction may have occurred early in some Austronesian languages, but at least in the ancestor language(s) of POc it did not, and the variation that we find today in the Austronesian languages also suggest that it was probably a later and more gradual development.

Dating the Austronesian influences on the Papuan languages is even more difficult. It appears that the change from SOV to SVO order is relatively recent. First, many of the Papuan languages of East Nusantara still have SOV order, or show remnants of erstwhile SOV order. In the languages of North Halmahera, for example, Ternate-Tidore, West Makian (Voorhoeve 1982) and Sahu (Visser and Voorhoeve 1987) have SVO order, while the other North Halmaheran languages, such as Galela (van Baarda 1908) and Pagu (Wimbish 1991) still have SOV (see Voorhoeve 1987, 1994). And even Ternate-Tidore have a number of features that typically come with SOV order, e.g., clause final conjunctions (Van Staden 2000: 38). Second, if POc was indeed verb initial, as also proposed for Western Austronesian languages (Himmelman 2005, see Table 1), then it follows that the ancestral Austronesian languages of the East Nusantara area were likewise verb-initial. If the change in constituent order in the Papuan languages is indeed as we suggest the result of contact with Austronesian languages, then this cannot have occurred until the Austronesian languages themselves had changed to SVO.

The introduction of an inclusive/exclusive distinction in the pronominal paradigm again suggests an earlier development, found in all the Papuan languages regardless of genealogical affiliation. As Nichols (2003: 304) observes, the borrowing of this feature involves ‘the opposition in the abstract’, while the formal expression comes from native resources.

These findings support the possible historical settlement of East Nusantara and the Bird’s Head that we gave in 2.2.1. and provide additional evidence for the hypotheses put forth by Grimes (1991), Voorhoeve (1994), and Ross (2001) (cf 2.2.1 and 2.3.1),

which argue that the innovations in the Austronesian languages typologically defined as *preposed possessor languages* (Himmelman 2005) are due to substratal Papuan influence. There were a number of Papuan speaking populations before Austronesian-speaking peoples entered this region. It is likely that the Moluccan Austronesian languages are the result of language shift. We assume that many Moluccan islands were once occupied by speakers of Papuan languages who were confronted by an influx of Austronesian speakers several millennia ago, see section 2.2. The two populations must have mixed, adopting the Austronesian languages of the more powerful invaders, but with some Papuan features.

Apparently, not all Papuan populations shifted completely to Austronesian languages. Scattered throughout the East Nusantara area a number of Papuan pockets remained: (1) on the small islands Alor and Pantar, (2) in the northeastern part of the island Timor, (2) in North-Halmahera and nearby small islands Tidore, Ternate and Makian, (3) on the tip of the Bomberai peninsula, (4) on Yapen island in the Cenderawasih Bay, and (5) in virtually all of the Bird's Head. For whatever reasons, these populations maintained their indigenous languages, but they did have prolonged contact with the various 'Papuanized' Austronesian speakers. The effects of this contact vary considerably, but the most noticeable pattern taken from the Austronesian languages is the introduction of the inclusive/exclusive distinction in the pronominal paradigm. In most of these areas Papuan languages maintained their SOV order, only some of the North-Halmahera and almost all of the BH languages rearranged the constituent order in the clause to V-medial.

Our conclusion is that these regions together constitute a linguistic contact area. The data also indicate that this area was not defined by a single wave of diffusion, but rather that several waves, taking place at different points in time (and perhaps going in various directions), have shaped to its present form.

## References

- Adelaar, K.A. & Himmelman, N.P. (eds). 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge/Curzon.
- Andaya, L.Y. 1993. *The world of Maluku: Eastern Indonesia in the early Modern Period*. Honolulu HI: University of Hawaii Press.
- Baarda, M.J. van. 1908. *Leiddraad bij het bestuderen van 't Galéla'sch Dialekt, op het eiland Halmahera*. Den Haag: Martinus Nijhoff.
- Baird, L. 2002. A grammar of Kéo: An Austronesian language of East Nusantara. PhD dissertation. Australian National University.
- Baird, L. In press. A *grammar of Klon*. Canberra: Pacific Linguistics.
- Baird, L., Klammer, M. & Kratochvíl, F. 2004. Alor Malay as a distinct variety of Malay. Paper presented on the 8th International Symposium on Malay/Indonesian linguistics (ISMIL 8), Universiti Sains Malaysia, Penang, Malaysia, 31 July–2 August 2004. [www.let.leidenuniv.nl/aapp/](http://www.let.leidenuniv.nl/aapp/)
- Bellwood, P. 1997. *The prehistory of the Indo-Pacific Archipelago*. 2<sup>nd</sup> edn. Honolulu HI: University of Hawaii Press.

- Bellwood, P. 1998. From Bird's Head to bird's eye view: Long term structures and trends in Indo-Pacific prehistory. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, J. Miedema, C. Odé & R.A.C. Dam (eds.), 951–975. Amsterdam: Rodopi.
- Berg, R. van den. 1989. *A grammar of the Muna language*. Leiden: KITLV Press.
- Berry, K. & Berry, C. 1987. A survey of the South Bird's head stock. *Workpapers in Indonesian languages and cultures* 4: 81–117. Abepura: SIL/Department of Education and Culture, Percetakan Universitas Cenderawasih.
- Berry, K. & Berry, C. 1999. *A description of Abun, a West Papuan language of Irian Jaya*. Canberra: Pacific Linguistics.
- Birdsell, J.B. 1977. The recalibration of a paradigm for the first popling of greater Australia. In *Sunda and Sahul: Prehistoric studies in Southeast Asia, Melanesia and Australia*, J. Allen, J. Golson & Jones (eds.), 113–167. New York NY: Academic Press.
- Blust, R.A. 1978. Eastern Malayo-Polynesian: A subgrouping argument. *Pacific Linguistics* C-61: 181–234.
- Blust, R.A. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32: 241–293.
- Bowden, J. 2001. *Taba, description of a South Halmahera language*. Canberra: Pacific Linguistics.
- Brandes, J.L.A. 1884. Bijdragen tot de vergelijkende klankleer der Westerse afdeeling van de Maleisch-Polynesische taalfamilie. PhD dissertation. University of Utrecht.
- Brotherson, A. 2003. A spatial odyssey: Referring to space in Makasai. Honours Thesis, Australian National University.
- Campbell, L. 1998. *Historical linguistics: An introduction*. Edinburgh: EUP.
- Capell, A. 1972. Portuguese Timor: Two more non-Austronesian languages. *Oceania Linguistic Monographs* 15.
- Capelli, C., Wilson, J.F., Richards, M., Stumpf, M.P.H., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V.L., Tsang-Ming Ko & Goldstein, D.B. 2001. A predominantly indigenous paternal heritage for the Austronesian-Speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics* 68: 432–443.
- Chappell, H. & McGregor, W. (eds.). 1996. *The Grammar of inalienability: A typological perspective on body part terms and the part-whole relation*. Berlin: Mouton de Gruyter.
- Clark, E.V. 1978. Locational: Existential, locative, and possessive constructions. In *Universals of human language*, Vol. 1: *Method and theory*, Vol. 2: *Phonology*. Vol. 3: *Word structure*, Vol. 4: *Syntax*, J.H. Greenberg, C.A. Ferguson & E.A. Moravcsik (eds.), 85–126. Stanford CA: Stanford University Press.
- Clark, R. 1990. The Austronesian languages. In *The major languages of East and South East Asia*, B. Comrie (ed.), 173–184. London: Routledge.
- Collins, J.T. 1983. *The historical relationship of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.
- Collins, J.T. & Kaartinen, T. 1998. Preliminary notes on Bandanese: Language maintenance and change in Kei. *Bijdragen tot de Taal-, Land- en Volkenkunde* 154–4: 521–570.
- Cowan, H.J.K. 1953. *Voorlopige resultaten van een ambtelijk taalonderzoek in Nieuw-Guinea*. 's-Gravenhage: Martinus Nijhoff.
- Cowan, H.J.K. 1955. Notes on Windesi grammar. *Oceania* 26: 42–48.
- Dix Grimes, B. 1991. The development and use of Ambonese Malay. In *Papers in Austronesian linguistics* 1, H. Steinhauer (ed.) 83–123. Canberra: Pacific Linguistics.
- Dixon, R.M.W. 1997. *The rise and fall of languages*. Cambridge: CUP.
- Dol, P. 1999. A grammar of Maybrat. PhD dissertation, Leiden University.
- Donohue, M. 1997. Tone systems in New Guinea. *Linguistic Typology* 1: 347–386.

- Donohue, M., 1999. *A grammar of Tukang Besi*. Berlin: Mouton De Gruyter.
- Engelenhoven, Aone van. 1995. A description of the Leti language. PhD dissertation. Leiden University.
- Flassy, D.A.L. & Stokhof, W.A.L. 1979. A note on Tehit (Bird's Head – Irian Jaya). *NUSA* 7: 35–83.
- Florey, M. 2001. Verb and valence in Alune. In *Explorations in valency in Austronesian languages* [La Trobe Papers in Linguistics 11], P. Austin, B. Blake & M. Florey (eds.), 73–120.
- Foley, W.A. 1986. *The Papuan languages of New Guinea*. Cambridge: CUP.
- Foley, W.A. 1998. Towards understanding Papuan languages. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, J. Miedema, C. Odé & R.A.C. Dam (eds.), 503–518. Amsterdam: Rodopi.
- Foley, W.A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29: 357–404.
- Fraassen, C.F. van. 1983. Ternate, de Molukken en de Indonesische archipel: Van soa-organisatie en vierdeling: Een studie van traditionele samenleving en cultuur in Indonesië (2 volumes). PhD dissertation, Leiden University.
- Friedberg, C. 1978. *Comment fut tranchée la liane celeste et autre textes de littérature orale bunaq (Timor, Indonésie)*. Paris: Société d'études linguistiques et anthropologiques de France Sela.
- Geurtjens, H. 1921. *Spraakleer der Keieesche Taal* [Verhandelingen van het Bataviaasch genootschap van Kunsten en Wetenschappen, Deel LXIII]. Den Haag: Martinus Nijhoff.
- Gravelle, G. 2002. Morphosyntactic properties of Meyah word classes. In *Languages of the Eastern Bird's Head*, G.P. Reesink (ed.), 109–180. Canberra: Pacific Linguistics.
- Gravelle, G. 2004. Meyah: An East Bird's Head language of Papua, Indonesia. PhD dissertation, Vrije Universiteit, Amsterdam.
- Grimes, C.E. 1991. The Buru language of Eastern Indonesia. PhD dissertation, Australian National University.
- Grimes, C.E. (ed.). 2000. *Spices from the East: Papers in languages of eastern Indonesia*. Canberra: Pacific Linguistics.
- Haan, Johnson W. 2001. The grammar of Adang: A Papuan language spoken on the island of Alor - East Nusa Tenggara - Indonesia. PhD dissertation, University of Sydney.
- Hasselt, F.J.F. van. 1905. *Spraakkunst der Numfoorsche Taal*. Den Haag: Martinus Nijhoff.
- Heine, B. 1997. *Possession: Cognitive sources, forces and grammaticalization*. Cambridge: CUP.
- Held, G.J. 1942. *Grammatica van het Waropensch* [Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen, deel LXXVII]. Den Haag: Martinus Nijhoff.
- Henderson, J.E. 1995. *Phonology and grammar of Yele, Papua New Guinea*. Canberra: Pacific Linguistics.
- Heuvel, W. van den. 2006. *Biak: Description of an Austronesian language of Papua*. Utrecht: LOT.
- Himmelman, N.P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In *The Austronesian languages of Asia and Madagascar*, K.A. Adelaar & N.P. Himmelman (eds.), 110–181. London: Routledge/Curzon.
- Hinton, B. 2000. The languages of Wetar: Recent survey results and word lists, with notes on Tugun grammar. In *Spices from the East: Papers in languages of eastern Indonesia*, C.E. Grimes (ed.), 105–129. Canberra: Pacific Linguistics.
- Hughes, J. 2000. The morphology of Dobel, Aru, with special reference to reduplication. In *Spices from the East: Papers in languages of eastern Indonesia*, C.E. Grimes (ed.), 131–180. Canberra: Pacific Linguistics.
- Hull, G. 2001a. A morphological overview of the Timoric Sprachbund. *Studies in languages and cultures of East Timor* 4: 98–205.
- Hull, G. 2001b. *Mambai language manual (Ainara dialect)*. Winston Hills: Sebastião A. da Silva Project.



- Jacobs, H.T.T. (ed.). 1971. *A treatise on the Moluccas (c.1544). Probably the preliminary version of António Galvão's Histiória das Molucas*. Rome: Loyola Press.
- Jones, L.K. 1986. The question of ergativity in Yawa, a Papuan language. *Australian Journal of Linguistics* 6: 37–55.
- Jones, L.K. 1991. Question words in Yawa. Canberra: *Pacific Linguistics* A-73: 97–105.
- Kayser M., Brauer, S. Weiss, G., Schiefenhövel, W., Underhill, P., Peidong Shen, Oefner, P., Tommaseo-Ponzetta, M. & Stoneking, M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *American Journal of Human Genetics* 72: 281–302.
- Kratochvíl, F. 2007. *A grammar of Abui: A Papuan language of Alor*. Utrecht: LOT.
- Klammer, M. 1998. *A grammar of Kambara*. Berlin: Mouton de Gruyter.
- Klammer, M. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41: 363–383.
- Kalmer, M. 2007. Alorese: An atypical Austronesian language. Paper presented at the Fifth International East Nusantara Conference (ENUS 5), 1–3 August 2007, Nusa Cendana University, Kupang, Indonesia, and the Third Conference on Austronesian Languages and Linguistics (ALL3), 21–22 September 2007, SOAS, London.
- Klammer, M. To appear. One item, many faces: 'Come' in Teiwa and Kaera. To appear in *Typological and areal analysis: Contributions from East Nusantara*, M. Ewing & M. Klammer (eds.) Leiden: KITLV Press.
- Klammer, M. Forthcoming. *A grammar of Teiwa*.
- Klinken, C. van. 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics.
- Laycock, D. 1978. A little Mor. In *Second International Conference on Austronesian Linguistics: Proceedings, Fascicle I. Pacific Linguistics*, S.A. Wurm & L. Carrington (eds.), 285–316. Canberra: Pacific Linguistics.
- Lichtenberk, F. 1985. Possessive constructions in Oceanic languages and in Proto-Oceanic. In *Austronesian Linguistics at the 15th Pacific Science Congress*, A. Pawley & L. Carrington (eds.), 93–140. Canberra: Pacific Linguistics.
- Lynch, J. 1978. Verbal aspects of possession in Melanesian languages. *University of Hawaii working papers in linguistics* 5(9): 1–22.
- Lynch, J., Ross, M. & Crowley, T. (eds.) 2002. *The oceanic languages*. Richmond: Curzon Press.
- Matsumura, T. & Mastumura, M. 1990. Irarutu grammar notes. Unpublished manuscript. Jayapura: Cenderawasih University and Summer Institute of Linguistics.
- Menick, R.H. 1996. Verb sequences in Moi. In *Studies in Irian languages*, G.P. Reesink (ed.), NUSA 40: 41–60.
- Minde, D. van. 1997. Malayu Ambong. PhD dissertation. Leiden University.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago IL: University of Chicago Press.
- Needham, R. 1983. *Sumba and the slave trade*. [Working Paper no. 31]. Clayton, Australia: Monash University, Centre of SE Asian Studies.
- Nichols, J. 1997. Sprung from two common sources: Sahul as a linguistic area. In *Archaeology and linguistics*, P. McConvell & N. Evans (eds.), 135–168. Melbourne: OUP.
- Nichols, J. 1998. The origin and dispersal of languages: Linguistic evidence. In *The Origin and Diversification of Language*, N.G. Jablonski & L.C. Aiello (eds.), 127–170. San Francisco CA: Memoirs of the California Academy of Sciences, Number 24.
- Nichols, J. 2003. Diversity and stability in language. In *The handbook of historical linguistics*, B.D. Joseph & R.D. Janda (eds.), 283–310. Oxford: Blackwell.

- Nitbani, S.H., Kosmas, J., Wona, S. & Naley, H. 2001. *Struktur Bahasa Lamma*. Jakarta: Pusat Bahasa, Departemen Pendidikan Nasional.
- Odé, C. 2002a. A sketch of Mpur. In *Languages of the Eastern Bird's Head*, G.P. Reesink (ed.), 45–107. Canberra: Pacific Linguistics.
- Odé, C. 2002b. *Mpur prosody: An experimental-Phonetic analysis with examples from two versions of the fentora myth* [ELPR Publication series A1-003]. Osaka: Endangered Languages of the Pacific Rim.
- Owens, M. 2000. Agreement in Bimanese. MA Thesis. University of Canterbury, New Zealand.
- Pasveer, J. 2003. The Djief hunters, 26,000 years of lowland rainforest exploitation on the Bird's Head of Papua, Indonesia. PhD dissertation, Rijksuniversiteit Groningen.
- Pawley, A.K. 1998. The Trans New Guinea Phylum hypothesis: A reassessment. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, J. Miedema, C. Odé & R.A.C. Dam (eds.), 655–690. Amsterdam: Rodopi.
- Rankin, R.L. 2003. The comparative method. In *The Handbook of Historical Linguistics*, B.D. Joseph & R.D. Janda (eds.), 183–212. Oxford: Blackwell Publishing.
- Reesink, G.P. 1996. Morpho-syntactic features of the Bird's Head languages. In *Studies in Irian languages*, G.P. Reesink (ed.), NUSA 40: 1–20.
- Reesink, G.P. 1998. The Bird's Head as Sprachbund. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, J. Miedema, C. Odé & R.A.C. Dam (eds.), 603–642. Amsterdam: Rodopi.
- Reesink, G.P. 1999. *A grammar of Hatam*. Canberra: Pacific Linguistics C-146.
- Reesink, G.P. (ed.) 2002a. *Languages of the Eastern Bird's Head*. Canberra: Pacific Linguistics.
- Reesink, G.P. 2002b. Clause-final negation: Structure and interpretation. *Functions of Language* 9: 239–268.
- Reesink, G.P. 2005. West Papuan languages: Roots and development. In *Papuan Pasts, Studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*, A. Pawley, R. Attenborough, J. Golson & R. Hide (eds.), 185–218. Canberra: Pacific Linguistics.
- Reid, A. 1984. The islamization of Southeast Asia. In *Historia: Essays in commemoration of the 25th anniversary of the Department of History, University of Malaya*, M. Abu Bakar, A. Kaur and A.Z. Ghazali. (eds.), 13–33. Kuala Lumpur: Malaysian Historical Society.
- Remijsen, B.C. 2001. *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT.
- Ross, M.D. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian Dictionary*, D.T. Tryon (ed.), 1: 45–120. Berlin: Mouton de Gruyter.
- Ross, M.D. 2001. Contact-induced change in oceanic languages in North-West Melanesia. In A.Y. Aikhenvald & R. M.W. Dixon (eds.) *Areal diffusion and genetic inheritance*, 134–166. Oxford: OUP.
- Ross, M.D. 2005. *Pronouns as a preliminary diagnostic for grouping Papuan languages*. In *Papuan Pasts, Studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*, Andrew Pawley, R. Attenborough, Jack Golson, and Robin Hide (eds.) 15–65. Canberra: Pacific Linguistics.
- Ross, M.D. To appear. Defining the Trans New Guinea family: preliminary evidence from pronouns. In *Papuan languages and the Trans New Guinea phylum*. A. Pawley, M. Ross and M. Osmond (eds.), Canberra: Pacific Linguistics.
- Silzer, P.J. 1983. Ambai: An Austronesian Language of Irian Jaya, Indonesia. PhD dissertation. ANU, Canberra.
- Staden, M. van. 2000. Tidore. A linguistic dscription of a language of the North Moluccas. PhD dissertation, Leiden University.



- Staden, M. van. Forthcoming. Possessive clauses in East Nusantara, the case of Tidore. In *The Expression of Possession*, Mc Gregor (ed.), Berlin: Mouton de Gruyter.
- Steinhauer, H. 1985. Number in Biak: Counterevidence to two alleged language universals. *Bijdragen tot de Taal-, Land- en Volkenkunde* 141: 462–485.
- Steinhauer, H. 1993. Sisters and potential wives: Where linguists and anthropologists meet. Some notes on kinship in Blagar (Alor Archipelago). In *Vrienden en verwanten: Liber amicorum Alex van der Leeden*, P. Haenen & A. Trouwborst (eds.), 147–168. Leiden/Jakarta: DSALCUL/IRIS.
- Steinhauer, H. 1995. Two varieties of the Blagar language (Alor, Indonesia). In *Tales from a concave world: Liber amicorum Bert Voorhoeve*, C. Baak, M. Bakker & D. van der Meij (eds.), 265–296. Leiden University: Department of Languages and Cultures of South East Asia and Oceania.
- Steinhauer, H. 2005. Biak. In *The Austronesian languages of Asia and Madagascar*, K.A. Adelaar and N.P. Himmelman (eds.), 793–823. London: Routledge/Curzon.
- Stokhof, W.A.L. 1975. *Preliminary notes on the Alor and Pantar languages (East Indonesia)*. Canberra: Pacific Linguistics.
- Svorou, S. 1993. *The grammar of space*. Amsterdam: John Benjamins.
- Thomason, S.G. 2001. *Language contact. An introduction*. Edinburgh: EUP.
- Tryon, D.T. 1995. The Austronesian languages. In *Comparative Austronesian dictionary: An Introduction to Austronesian studies*. Part 1: *Fascicle*, D.T. Tryon (ed.), 5–44. Berlin: Mouton de Gruyter.
- Veen, H. van der. 1915. *De Noord-Halmahera'se taalgroep tegenover de Austronesiese talen*. Leiden: Van Nifterik.
- Veth, P., Spriggs, M., Jatmiko, A. & O'Connor, S. 1998. Bridging Sunda and Sahul: The archaeological significance of the Aru Islands, southern Moluccas. In *Bird's Head approaches* [Modern Quarternary Research in Southeast Asia 15.], G.-J. Bartstra (ed.), 157–177. Rotterdam: Balkema.
- Visser, L.E. & Voorhoeve, C.L. 1987. *Sahu-Indonesian-English dictionary and Sahu grammar sketch* [Verhandelingen van het KITLV 126]. Dordrecht: Foris.
- Voorhoeve, C.L. 1975. *Languages of Irian Jaya: Checklist. Preliminary classification, language maps, wordlists*. Canberra: Pacific Linguistics.
- Voorhoeve, C.L. 1982. *The Makian languages and their neighbours*. Canberra: Pacific Linguistics.
- Voorhoeve, C.L. 1987. The non-Austronesian languages in the North Moluccas. In *Halmahera dan Raja Empat sebagai Kesatuan majemuk*, [Buletin LEKNAS, Vol. II, No. 2, 1983; published 1987], E.K.M. Masinambow (ed.), 13–39. Jakarta: Lembaga Ekonomi dan Kemasyarakatan Nasional, Lembaga Ilmu Pengetahuan Indonesia.
- Voorhoeve, C.L. 1989. The masked bird: Linguistic relations in the Bird's Head area. In *Peoples on the move*, P. Haenen & J. Pouwer (eds.), 78–101. Nijmegen: Centre for Australian and Oceanic Studies.
- Voorhoeve, C.L. 1994. Contact-induced change in the non-Austronesian languages in the north Moluccas, Indonesia. In *Language Contact and Change in the Austronesian world*, T. Dutton and D.T. Tryon (eds.), 649–674. Berlin: Mouton de Gruyter.
- Vries, L. de. 1996. Notes on the morphology of the Inanwatan language. In *Studies in Irian languages*, G.P. Reesink (ed.). NUSA 40: 97–127.
- Vries, L. de. 1998. Some remarks on the linguistic position of the Inanwatan language. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, J. Miedema, C. Odé & R.A.C. Dam (eds.), 643–653. Amsterdam: Rodopi.

- Vries, L. de. 2001. An Introduction to the Inanwatan language of Irian Jaya. In *Between worlds: Linguistic papers in memory of David J. Prentice*, K.A. Adelaar & R. Blust (eds.), 77–93. Canberra: Pacific Linguistics.
- Vries, L. de. 2004. *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua, Indonesia*. Canberra: Pacific Linguistics.
- Williams-van Klinken, C., Hajek, J. & Nordlinger, R. 2002. *Tetun Dili: A grammar of an East Timorese language*. Canberra: Pacific Linguistics.
- Wimbish, S.G. 1991. An introduction to Pagu through the analysis of narrative discourse. MA thesis, The University of Texas at Arlington.
- Wurm, S.A. 1982. *The Papuan languages of Oceania*. [Ars Linguistica 7]. Tübingen: Gunter Narr.
- Zanten, E. van & Dol, P. To appear. Word stress and pitch accent in Papuan languages. In *Stress patterns of the world II: The data*, R. Goedemans & H.G. van der Hulst (eds.). Amsterdam: John Benjamins.

## Abbreviations

1	first person	N	nominative
2	second person	NEG	negator
3	third person	NH	non-human
A	actor	OBJ	object
ADDR	addressee	PERF	perfect
APPL	applicative	PL	plural
ART	article	POS	possessive
ASP	aspectual marker	QUEST/Q	question marker
CAU	causative	S	subject
COMP	complementiser	SG	singular
D	dative	SU	subject
DEF	definite	TOP	topic
DEI	deictic	U	undergoer
DU	dual		
DUR	durative		
EMPH	emphasis		
EXC	exclusive		
F	feminine		
FUT	future		
GEN	genitive		
IMPF	imperfective		
INAL	inalienable		
INAN	inanimate		
INC	inclusive		
INDEX	indexer		
M	masculine		

# Appendix

The following table gives an alphabetical list of languages mentioned in the paper, with general classification and (where applicable) subgroup classification, geographical location and references.

Table 5. Languages cited

Language	Genetic affiliation	Subgroup (if applicable)	Geographical area	References
Abui	PAPUAN	Timor-Alor-Pantar of TNG	Alor island	Kratochvíl 2007
Abun	PAPUAN	Family-level isolate of WBH Stock of WPP	NW Bird's Head	Voorhoeve 1987; Also known as <i>Karon Pantai</i> , a family-level isolate of WPP
Adang	PAPUAN	Timor-Alor-Pantar of TNG	Alor island	Wurm 1982
Alorese	AN	CMP	Alor & Pantar	fieldnotes 2003
Alune	AN	CMP	C Moluccas	Floreay 2001
Ambai	AN	SHWNG	Yapen island, Cenderawasih Bay, E of Bird's Head	Silzer 1983
Ambonese	AN	WMP	C Moluccas	Van Minde 1997
Malay				
Bandanese	AN	CMP	C/S Moluccas	Van Fraassen 1983, Collins & Kaartinen 1998
Biak (+ Numfor dialect)	AN	SHWNG	Biak island, Cenderawasih Bay, E of Bird's Head	Steinhauer 1985, 2005; Van den Heuvel 2006
Bimanese	AN	CMP	Bima	Owens 2000
Blagar	PAPUAN	Timor-Alor-Pantar of TNG	Pura and Tereweng island, W of Alor	Steinhauer 1993, 1995
Bunak	PAPUAN	Timor-Alor-Pantar of TNG	C/E Timor	Friedberg 1978, Klammer fieldnotes 2002
Buru	AN	CMP	C Moluccas	Grimes 1991

(Continued)

Table 5. Continued

Language	Genetic affiliation	Subgroup (if applicable)	Geographical area	References
Dobel	AN	CMP	SE Moluccas	Hughes 2000
Galela	PAPUAN	N Halmahera family of N Moluccan Stock of WPP	N Halmahera	Voorhoeve 1987; Voorhoeve 1994
Hatam	PAPUAN	Hatam-Mansim family	E Bird's Head	Voorhoeve 1987; Reesink 1999
Idate	AN	CMP	E Timor	Hull 2001a, Klammer fieldnotes 2002
Inanwatan	PAPUAN	Inanwatan Family of SBH Stock	S Bird's Head	Voorhoeve 1975; De Vries 1996, 1998, 2001
Irarutu	AN	CMP/SHWNG?	Bomberai Peninsula, S of Bird's Head	Matsumura and Matsumura 1990; Blust 1993
Isní	AN	CMP	E Timor	Hull 2001a
Kaitetu	AN	CMP	Seram	Collins 1983
Kambera	AN	CMP	E Sumba	Klammer 1998
Kei	AN	CMP	Kai islands, SE Moluccas	Geurtjens 1921; Blust 1993
Kemak	AN	CMP	E Timor	Hull 2001a, Klammer field notes 2003
Keo	AN	CMP	E Flores	Baird 2002
Klon	PAPUAN	Timor-Alor-Pantar of TNG	Alor island	Baird In press
Lakalei	AN	CMP	E Timor	Hull 2001a, Klammer field notes 2003
Lamma	PAPUAN	Timor-Alor-Pantar of TNG	Pantar	Nitbani et. al. 2001
Leti	AN	CMP	Leti, S Moluccas	Blust 1993; Van Engelenhoven 1995
Lolein	AN	CMP	E Timor	Hull 2001a
Makasai	PAPUAN	Timor-Alor-Pantar of TNG	E Timor	Brotherson 2003
Ma'ya	AN	SHWNG	Raja Ampat islands, W of Bird's Head	Blust 1978; Remijnsen 2001
Mambai	AN	CMP	E Timor	Hull 2001b, Klammer field notes 2002
Mansim	PAPUAN	Hatam-Mansim	E Bird's Head	Voorhoeve 1975; Reesink 2002a. Earlier known as <i>Borai</i> of

(Continued)

Table 5. Continued

Language	Genetic affiliation	Subgroup (if applicable)	Geographical area	References
Marind	PAPUAN	Marind family of TNG	SE Papua	Borai-Hattam Stock-level Family belonging to WPP Foley 1986
Matbat	AN	SHWNG	Raja Ampat islands, W of Bird's Head	Remijnsen 2001
Maybrat (Brat)	PAPUAN	Family-level isolate of WBH Stock of WPP	C Bird's Head	Voorhoeve 1987; Dol 1999
Meyah (Meax)	PAPUAN	East Bird's Head family	E Bird's Head	Voorhoeve 1987; Gravelle 2002
Moi	PAPUAN	WBH family of WBH Stock of WPP	W Bird's Head	Voorhoeve 1989
Mor	AN	SHWNG	Mor island, Cenderawasih Bay, E of Bird's Head	Laycock 1978
Moraid	PAPUAN	WBH family of WBH Stock of WPP	W Bird's Head	Voorhoeve 1987
Mpur	PAPUAN	Stock-level isolate of WPP	northeast Bird's Head	Odé 2002a, 2002b. Earlier known as <i>Amberbaken</i> stock-level isolate of WPP (Voorhoeve 1975).
Muna	AN	CMP	S Sulawesi	Van den Berg 1989
Pagu	PAPUAN	North Halmahera family of North Moluccan Stock of WPP	North Halmahera	Voorhoeve 1987; Voorhoeve 1994: 650; Wimbish 1991
Sahu	PAPUAN	North Halmahera family	North Halmahera	Voorhoeve 1994
Seget	PAPUAN	WBH family of WBH Stock of WPP	southWest Bird's Head	Voorhoeve 1987
Selaru	AN	CMP	Tanimbar Archipelago, S of Seram	Blust 1993
Sentani	PAPUAN	Sentani Family of Sentani Stock TNGP	northeast Papua	Voorhoeve 1975: 41

(Continued)

Table 5. Continued

Language	Genetic affiliation	Subgroup (if applicable)	Geographical area	References
Sougb (known as <i>Manikion</i> )	PAPUAN	Member of the EBH Stock-level Family of WPP	E Bird's Head	Voorhoeve 1987; Reesink 2002a
Taba	AN	SHWNG	Makian island, Halmahera	Blust 1978; Bowden 2001
Tehit	PAPUAN	WBH family of WBH Stock of WPP	SW Bird's Head	Voorhoeve 1987
Tetun Dili	AN	CMP	East Timor	Williams-van Klinken et al. 2002
Tetun Fehan	AN	CMP	Timor	Van Klinken 1999
Tehit	PAPUAN	WBH	SouthWest Bird's Head	Flassy and Stokhof 1979
Teiwa	PAPUAN	Timor-Alor-Pantar of TNG	Pantar, E of Alor	Klamer, forthcoming
Tidore	PAPUAN	N Halmahera family of the N Moluccan Stock of WPP	Tidore island, Halmahera	Van Staden 2000; Voorhoeve 1987
Tokodede	AN	CMP	E Timor	Hull 2001a, Klamer field notes 2002
Tugun	AN	CMP	Wetar, SW Moluccas	Hinton 2000
Tukang Besi	AN	CMP	S Sulawesi	Donohue 1999
Waimaha	AN	CMP	E Timor	Hull 2001
Wanda- men	AN	SHWNG	'neck' of the Bird's Head	Cowan 1955 (Windesi-Wandamen); Reesink 1996
Waropen	AN	SHWNG	coast of Cenderawasih Bay, E of Bird's Head	Held 1942; Blust 1993
West- Makian	PAPUAN	Family-level isolate of N Moluccan Stock of WPP	Makian island, Halmahera	Voorhoeve 1987; Voorhoeve 1994
Yawa	PAPUAN	Yawa Stock-level isolate of Geelvink Bay Phylum	Yapen island	Voorhoeve 1975; Jones 1986; Reesink 2005



# The Guaporé-Mamoré region as a linguistic area

Mily Crevels and Hein van der Voort

The Guaporé-Mamoré region is one of the world's linguistically most diverse regions, with over 50 languages representing eight different stocks and 11 genetic isolates. In spite of the fact that these languages diverge enormously at the lexical level, they do seem to share a considerable number of structural features. Most of these languages are seriously endangered, but fortunately recent research initiatives have contributed to their documentation and led to new insights and data. This paper reviews the present research situation with regard to these languages and examines possible areal linguistic relationships in the Guaporé-Mamoré region.

## 1. Introduction

The Guaporé and Mamoré are two great rivers of the Southwestern Amazon region. The Guaporé River forms the border between the Brazilian federal state of Rondonia and the Bolivian departments of Santa Cruz and Beni, where the river is called Iténez. It winds slowly through endless forested and often flooded lowland in a northwesterly direction. At Puerto Avaroa on the Bolivian side and Surpresa on the Brazilian side, it turns north and flows into the Mamoré River that descends from the Bolivian Andean highlands in the south. Shortly after Guayaramerín/Guajará-Mirim, the Mamoré joins with the Beni River to form the Madeira River, which continues for roughly 1000 kilometers before it flows into the Amazon River close to Manaus.

The Guaporé and Mamoré rivers together drain a part of the tropical lowlands where traditionally over 50 different indigenous languages are spoken. Since these languages represent numerous language families and unclassified languages that may be isolates, it is one of the linguistically most diverse regions of South America. This linguistic diversity, which includes, besides languages from the Arawakan, Chapacuran, Jabutian, Nambikwaran, Panoan, Tacanan, and Tupian families no less than 11 unclassified languages, is highly threatened with extinction. The speakers' numbers have dwindled under the ecological, physical, social and Western cultural pressures of the national societies. Nowadays more than half of the languages have less than 50 speakers and one third has less than 10 speakers.



The existing classifications of the Guaporé-Mamoré languages are often based on thin evidence, since the languages in question lacked thorough professional documentation and description. Fortunately, the past decennium has seen a considerable increase in the number of linguists conducting fieldwork in the region, producing grammars and dictionaries of the different languages. Several initiatives (e.g., those of Muysken and of Wetzels in The Netherlands, and Moore in Brazil) have led to long-term descriptive projects on Guaporé-Mamoré languages, and from comparisons of the preliminary results it now becomes clear that our ideas about the relationships between the languages of the region have to be revised. It appears that, apart from new insights concerning certain genetic relationships, the accumulated data also point to relationships of linguistic diffusion. In the present chapter we will tentatively claim that the Guaporé-Mamoré region is a linguistic area in the traditional sense, possibly with several sub-areas.

The languages that will be mentioned in the following are numbered and listed in Table 2 below. The numbers in the list correspond to the numbers on Map 1, which provides the geographic location of those languages.

### 1.1 Traditional cultures and habitat

The indigenous cultures on the right and left side of the Guaporé are markedly different. The Guaporé or Iténez River forms a border between traditional indigenous cultural areas. The culture area on the Bolivian side is known as the Moxos or Moxo-Chiquito cultural complex (Lévi-Strauss 1948, Denevan 1966, Crevels 2002). The culture area on the Brazilian side is often referred to as the Guaporé cultural complex (Lévi-Strauss 1948, Galvão 1960).

On the basis of material cultural traits, Lévi-Strauss (1948) divides the Guaporé cultural complex into two parts: the Chapacura area to the west of the Branco River and the Tupí area to its East. In her comparative ethnohistorical work Maldi (1991: 225) suggests that certain Chapacura-speaking groups in Brazil originated from the disintegrated Jesuit missions on the Bolivian side. Furthermore, she defines a specific section of Lévi-Strauss' Tupí area as the Marico cultural complex of Rondonia. The Marico cultural complex includes not only cultures of Tupí-speaking peoples, but also of speakers of Jabutian languages and of isolated languages, and is characterized by a combination of the following traits:

- seminomadic swidden agriculture combined with hunting and gathering
- relatively small egalitarian societies
- territorial subgroups that often bear animal names
- territorial subgroups could form alliances with others across linguistic borders
- religion involves shamanism and hallucinogenic substances
- material culture is characterized among others by the *marico*, a crochet carrying net made of the fibres of specific palmtree leaves

- the local fermented alcoholic brew called *chicha*, which is a drink mainly based on maize, yam, manioc or fruits such as banana is mashed, fermented and sifted in a specific way (Caspar 1975: 45, Maldi 1991: 243)

In spite of the lack of genealogical relationships between many languages to the right of the Guaporé River, the cultures of their speakers share many of the traits listed here. These cultural correspondences are probably to a large extent due to the frequent contacts and intermarriage between neighbouring groups. Such interethnic contacts were not just confined to directly connected river basins, but they also took place over land between groups on different headwaters. The headwaters of, for example, the Corumbiara, Branco and Mekens rivers are very close to each other and to those of the Pimenta Bueno River (see Map 1). The combination of cultural traditions and ecological circumstances under which the peoples of this region subsisted did not allow for populations much larger than several thousand individuals. Nevertheless, a number of the cultural traits listed above are also found in other regions.

Note that the Marico cultural complex as defined by Maldi does not match exactly with the eastern part of the Guaporé cultural complex as defined by Galvão, whereas it corresponds more to Lévi-Strauss' Tupí culture area. The Nambikwara (NAMBIKWARAN) were considered neither by Lévi-Strauss nor Maldi to form part of a specific cultural complex. Galvão, however, did include the Nambikwara, who maintained trade relations with Marico culture groups, in the Guaporé complex. Galvão also included the Parecí (ARAWAKAN) and the Irantxe (UNCLASSIFIED), whereas he excluded the Karo (TUPÍ-RAMARAMA), Gavião and Suruí-Paitér (both TUPÍ-MONDÊ) from the Guaporé cultural complex.

The cultures to the left of the Guaporé River, i.e. the Moxo-Chiquito area according to Lévi-Strauss, which may be regarded as an intermediate area between Amazonian and Andean cultures, are characterized by:

- raised field agriculture, hunting, gathering and fishing
- more stratified societies divided into large villages
- religion based on a jaguar cult

Throughout this region thousands of artificial mounds were built, which along with hundreds of rectangular ponds and canals formed part of a complex cultivation and irrigation system. As pointed out by Erickson (2000a, 2000b), William Denevan's 1961 discovery of the massive prehispanic earthworks over vast areas in Moxos led to a complete change in perspective regarding cultural development in the Amazon Basin. A traditional archaeological approach of the Amazon points out the environmental limitations to cultural development, predominance of simple societies (bands and tribes), and subsistence systems based on hunting, gathering, and fishing with some limited slash and burn agriculture (Steward and Faron 1959). The type of prehispanic raised field agriculture as documented in Denevan (1966), however, shows that

intensive agriculture was indeed possible and that large, dense populations were supported in these areas.

Even though the Guaporé and Moxo-Chiquito culture areas are quite distinct, there is archaeological and historical evidence for exchange across the Guaporé River. As an example, the rocker stones used to crush maize, which are encountered among the Karitiana (Storto p.c.) and in the oral traditions of Jabuti-speaking groups (van der Voort 2006), possibly originate from Bolivia, where they are encountered among, for example, the Itonama.

Finally, it is important to consider the Tapajós-Madeira cultural complex (Galvão 1960), also known as the Ji-Paraná/Roosevelt cultural complex (Maldi 1991), which is located more to the northeast of the Marico cultural complex. There were contacts between groups on the headwaters of the Pimenta Bueno River and the Tupí-Mondê speaking Cinta Larga, Suruí-Paitér and Zoró peoples on the headwaters of the Roosevelt River, to the east (see Map 1). Table 1 lists the different culture areas discussed in the present section.

Table 1. Cultural complexes in the Guaporé-Mamoré region

Name	Author	Area
Marico	Maldi (1991)	South-eastern Rondonia
Tupí	Lévi-Strauss (1948)	South-eastern Rondonia, Western Mato Grosso
Chapacura	Lévi-Strauss (1948)	South-western Rondonia
Guaporé	Lévi-Strauss (1948), Galvão (1960)	Southern Rondonia, Western Mato Grosso
Tapajós-Madeira	Galvão (1960),	Border region shared by
Ji-Paraná/Roosevelt	Maldi (1991)	Amazonas, Mato Grosso and Pará
Moxo-Chiquito	Lévi-Strauss (1948)	Beni and northern Santa Cruz departments

1.2 History of contact and present situation

On the left side of the Guaporé or Iténez, i.e. the Bolivian side, contact with Western culture started in the sixteenth century when the first Spanish expeditions with explorers and conquistadors entered the Moxos area in search of El Dorado's gold. They were immediately followed by the slavery expeditions from the Santa Cruz area that came in search of human work force. Although these expeditions usually did not stay on for a prolonged period in the Moxos area, they lasted long enough to contaminate the native populations with epidemic diseases and disturb the demographic balance by taking away young and strong men. By the end of the seventeenth century, the Jesuits, in search of souls, formed the third group to enter Moxos. The often very inhumane

ways, in which the local clergy and administrators who ruled the missions after the expulsion of the Jesuits in 1767 treated the indigenous inhabitants of the missions, has been sufficiently illustrated in, among others, Alcide d'Orbigny's descriptions of his travels in Bolivia (1839). Mission culture came to an end when the rubber boom of the late nineteenth century brought white, criollo and mestizo settlers in sufficient numbers to overwhelm the indigenous culture – in so far as it still existed. Although the natives remained, they became marginalized to the Westernized society that had developed in the core regions, in which fluency in Spanish, affluence, and light skin had become the main values. Obviously, four centuries of contact with Western culture have wiped out traditional cultures to a great extent, causing many of the languages to become obsolete or disappear in the process (cf. Crevels 2002).

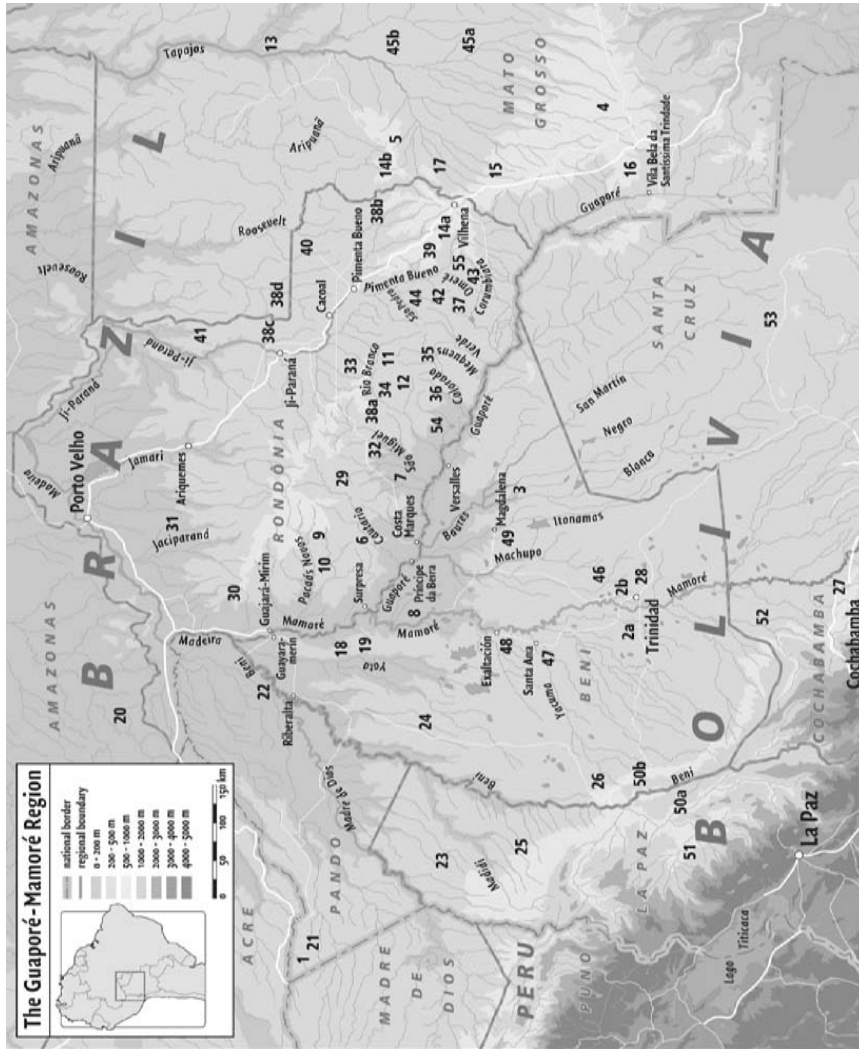
Up to the 1990s the relationship between the indigenous groups and the Bolivian government was based on exclusion from mainstream society, and sometimes even resulted in extermination. In the early 1990s, however, this relationship changed. The State started to take the demands of the native groups into account. The 1990 and 1996 Indian marches held in support of territorial claims have led to the recognition of the fact that the struggle for territory lies at the base of all of the indigenous demands and should, therefore, be treated with priority. By the end of the millennium, Amazonian Bolivia had not only captured the attention of the State but also that of the civilian society. Within the context of the decentralization of the administration on the one hand, and the growing interest of NGOs on the other hand, the attention for regional and local indigenous problems has grown considerably. Nevertheless, it remains to be seen what will be the real impact of these developments in the long run.

The right side of the Guaporé, i.e. the Brazilian side, was confronted with Western culture really only in the beginning of the twentieth century, with the rubber boom. In the 1940s the government planned to open up the Amazon region for settlers from the south and in the 1960s a highway was built that facilitated immigration on a relatively large scale. This has led to considerable population losses among the Indians and consequent loss of traditional cultures, although the memories of traditional culture are still vivid, and eyewitness accounts can still be recorded. The Brazilian side of the region is now deforested for about 50% and many of the traditional indigenous lands are occupied by ranchers. The Indians nowadays live mostly on indigenous reserves in the more remote parts of Rondonia. Even though these parts, including the reserves, are under external political and ecological pressure (the latter both internal due to overpopulation and external due to illegal invasions and destruction of the surrounding forest), the inhabitants have more possibilities to maintain their lifestyle on the reserves than in the slums of the local boomtowns. It is a fortunate development that the Brazilian Indians are becoming interested in protecting and revitalising their traditional cultures and languages. The Brazilian constitution reflects an enlightened view with regard to indigenous peoples, even if the local policies and actual circumstances do not always reflect this view.

## 2. Overview of the languages of the Guaporé-Mamoré region

In the sixteenth century, when the Spaniards first arrived in what today is the Bolivian part of the region, the area was populated by some 400 groups or tribes with an estimated total of 350,000 individuals who spoke about 39 different languages, most of which belonged to the Arawakan family (Baptista Morales 1995: 71). Today only 22 languages are still spoken in this part of the region, of which three belong to the Arawakan family, one to the Chapacuran family, three are Panoan, five Tacanan and two Tupian, both belonging to the Tupí-Guaraní subfamily. The unclassified languages of this part of the region are Leko, Mosetén/Chimane, Yurakaré, Canichana, Movima, Cayubaba, Itonama and Chiquitano (Bolivian lowland languages not mentioned here fall outside of the Guaporé-Mamoré basin or are extinct). Several of these languages only have a handful of speakers, and Canichana has most probably become extinct very recently. In Rondonia and adjacent parts, on the Brazilian side of the river, the majority of the languages (originally about 35) are still extant. There is one Panoan language, three or four Chapacuran languages, 16 Tupian languages that belong to six different subfamilies, four Nambikwaran languages, two Jabutian languages and Rikbak-tsá, which probably belong to the Macro-Ge stock, and four unclassified languages: Aikanã, Kanoê, Kwaza, and Irantxe. Consequently, the original linguistic diversity is still largely intact here, even though speaker numbers may be low.

Recently a number of language descriptions have become available; a fair number are on their way and are partly accessible through recent articles and personal communication with field linguists. Table 2 lists the languages of the Guaporé-Mamoré region with information about their classification, demographic data and speaker numbers. Furthermore, the table lists the comprehensive descriptive output of the researchers that have been working on the languages most recently. Demographic data and speaker numbers are adapted from Crevels and Adelaar (2000–2006). In several parts of Rondonia, traces have been found of existing populations that avoid contact, and it is not known which languages are spoken by them. In the table the unidentified language of one of these groups is listed, but probably there are more. Also one pidgin language is listed, that is used in the contact between the Kanoê and Akuntsu of the Omeré region. Traditionally these groups avoid each other, but due to attempts by surrounding landowners to exterminate them, they were forced to share the protection of a small island of forest and maintain some level of contact. Map 1 shows the probable geographical location of the Rondonian languages at the time of contact, which does not necessarily correspond completely to the present location due to decimation and displacement of many indigenous groups after contact with Western society. The location of the Bolivian languages coincides more or less with the original location at the time of contact, or with the Jesuit missions in which some of the groups were brought together in the seventeenth and eighteenth century. The concentration of different groups in the Jesuit missions resulted in the rapid loss of their tribal organization and linguistic and cultural differences. Nevertheless, nowadays most extant groups are still



Map 1

to be found in the area where the Spaniards first contacted them. The numbers on Map 1 correspond with the languages listed in Table 2.

**Table 2.** Linguistic diversity of the Guaporé-Mamoré region

	Language	Classification	Population	Speakers	Recent (major) output
1	Machineri	Arawakan	155	140	
	Mojo	Arawakan	20,805	<10,000	
2a	Trinitario				Rose (in prep.) Salvatierra 2005
2b	Ignaciano				Olza Zubiri 2002
3	Baure	Arawakan	631	±40	Danielsen 2007
4	Parecí	Arawakan	1,300	?	
5	Enawê-Nawê (Salumã)	Arawakan	320	320	
6	Kaw Ta Yo (Kuyubí)	Chapacuran	50	3	Duran 2000
7	Miguelenho	Chapacuran	50	1	
8	Moré (Itene)	Chapacuran	200	76	Angenot-de Lima 2001
9	Oro Towati (Oro Win)	Chapacuran	82	4	França 2002
10	Wari' (Pakaanova)	Chapacuran	1,300	1,300	Everett & Kern 1997
11	Arikapú	Macro-Ge, Jabuti	15?	2	van der Voort 2005a
12	Djeoromitxi	Macro-Ge, Jabuti	65	40	Pires 1992
13	Rikbaktsá	Macro-Ge	900	900?	Boswood 1971 Silva 2005
14a	Latundê	Nambikwaran,	19	19	Telles 2002
14b	Lakondê	North	7	1	
15	Nambikwara	Nambikwaran, South	820	820	Kroeker 2001
16	Sararé	Nambikwaran, South	80	80	Borella (in prep.)
17	Sabanê	Nambikwaran	140	<10	Araujo 2004
18	Chácobo	Panoan	767	550	Iggesen (in prep.)
19	Pacahuara	Panoan	19	18	
20	Kaxararí (Kaxariri)	Panoan	270	?	
21	Yaminahua	Panoan	171	137	Faust & Loos 2002
22	Ese Ejja	Tacanan	584	502	Vuillermet (in prep.)
23	Araona	Tacanan	90	81	Emkow 2007
24	Cavineña	Tacanan	1,736	1,180	Guillaume 2004
25	Tacana	Tacanan	5,058	1,821	
26	Maropa (Reyesano)	Tacanan	4,118	>10	Guillaume 2005
27	Yuki	Tupian, Tupí-Guaraní	156	125	Villafañe 2004

(Continued)



Table 2. Continued

	Language	Classification	Population	Speakers	Recent (major) output
28	Sirionó	Tupian, Tupí-Guaraní	419	399	Hemmauer 2005
29	Uru-eu-wau-wau (Amondawa, Jupa'ú)	Tupian, Tupí-Guaraní, Kawahib	130	130	Sampaio 2001
30	Karipuna	Tupian, Tupí-Guaraní, Kawahib	11	11	
31	Karitiana	Tupian, Tupí-Arikém	170	170	Storto 1999
32	Puruborá	Tupian	50	2	Galucio 2005
33	Tuparí	Tupian, Tupí-Tuparí	380	230	Alves 1991, 2004
34	Makuráp	Tupian, Tupí-Tuparí	130	50	Braga 1996, 2005
35	Mekens (Sakiráp)	Tupian, Tupí-Tuparí	65	23	Galucio 1996, 2001
36	Wayuru (Ajurú)	Tupian, Tupí-Tuparí	38	10?	
37	Akuntsu (Akũtsũ)	Tupian, Tupí-Tuparí	7	7	
38a	Aruá	Tupian, Tupí-Mondê	36?	12, 20?	
38b	Cinta Larga		<1,000	<1,000	
38c	Gavião		430	440	Moore 1984
38d	Zoró		415	415	
39	Salamãï (Mondé)	Tupian, Tupí-Mondê	10?	2	
40	Suruí-Paitér	Tupian, Tupí-Mondê	920	920	van der Meer 1982 Guerra 2004
41	Karo (Arara)	Tupian, Tupí-Ramarama	170	150	Gabas Jr. 1989, 1999
42	Aikanã (Masaká, Kasupá, Huarí)	Unclassified	200	170?	Vasconcelos 2004
43	Kanoê (Kapixaná)	Unclassified	80?	5	Bacelar 1994, 2004
44	Kwazá (Koaiá)	Unclassified	15	25	van der Voort 2004
45a	Irantxe	Unclassified	250	?	Monserat 2000
45b	Mýky		76	?	
46	Canichana	Unclassified	583	0?	

(Continued)



Table 2. Continued

	Language	Classification	Population	Speakers	Recent (major) output
47	Movima	Unclassified	6,528	1,452	Haude 2006
48	Cayubaba	Unclassified	794	<5	Key 1967
49	Itonama	Unclassified	5,090	<5	Crevels 2007
50a	Mosetén	Unclassified	1,200	585	Sakel 2004
50b	Chimane		5,907	5,316	Gill 1999a,b
51	Leko	Unclassified	80	20	van de Kerke 1998–2006
52	Yurakaré	Unclassified	3,333	2,675	van Gijn 2006
53	Chiquitano (Besiro)	Unclassified	47,086	5,855	Galeote 1996
54	(unknown)	(unknown)	100	100	
55	Akuntsu-Kanoê	Pidgin	–	7	

## 2.1 The state of documentation

Early ethnographic explorers, such as Nordenskiöld (1915), expressed their concern about the state of documentation of the indigenous languages and cultures in the Guaporé-Mamoré region. In view of Western culture encroaching upon the native cultures, Nordenskiöld found that ethnographically interested persons should urgently take up research in the region. Many linguistic publications about the Bolivian side of the region are from this period, such as de Créqui-Montfort and Rivet (1914a) on Movima, (1914b) on Canichana, (1914c, 1917–1920) on Cayubaba, and (1916, 1918) on Itonama. These publications present an analysis of certain grammatical aspects of the respective languages, based on earlier word lists and religious texts from the eighteenth and nineteenth century, of which the reliability is generally doubtful.

In the 1950s, after the devastating epidemics of the previous decades, Becker-Donner (1955) considered the majority of languages and cultures on the Brazilian side to be endangered, with Kanoê, with only a handful of speakers known to her, as a dramatic example. One of the reasons to publish her preliminary research based on incomplete data in her 1955 article was to alert the scientific community about the necessity of documentation. Nevertheless, this article continued to be one of the most important published sources on the cultures of the region and on the Kanoê and Aikanã languages during the following decades.

In the 1960s more research was initiated on the Brazilian side by missionary linguists of the Summer Institute of Linguistics (SIL), some of which led to thorough work, such as Kroeker (2001) on Nambikwara, and by the New Tribes Mission, which resulted in a description of Wari' (Everett and Kern 1997). Other researchers such as Price (1972, 1978) and, in the 1970s, Moore (1984) produced important work on Nambikwara culture and the Gavião language, respectively.

In the Bolivian part of the region, the New Tribes Mission became active in 1942, only two years after its foundation. Today they still work with eight groups in Bolivia, six of which are located in the Guaporé-Mamoré area: Trinitario, Araona, Ese Ejja, Yuki, Chimane, and Yurakaré. SIL researchers became active in the mid 1950s and stayed on until 1985. In this time span they worked on 18 languages in Bolivia, of which 13 pertain to the Guaporé-Mamoré region: Baure, Ignaciano, Chácobo, Pacahuara, Araona, Cavineña, Ese Ejja, Tacana, Sirionó, Cayubaba, Chiquitano, Itonama, and Movima. Resulting publications include, among others, Key (1967, 1975) on Cayubaba, Camp and Liccardi (1967a,b) on Itonama, and Judy and Judy (1962, 1967) on Movima.

Finally, heightened awareness of the issue of endangered languages and subsequent increase of funding brought a considerable number of researchers to the region in the 1990s and at the beginning of the new millennium. The results of their work are now becoming available, such as Gabas (1999) on Karo, Storto (1999) on Karitiana, Galucio (2001) on Mekens, Telles (2002) on Lakondê/Latundê, van der Voort (2004) on Kwaza, Sakel (2004) on Mosetén, Bacelar (2004) on Kanoê, Araujo (2004) on Sabanê, Guillaume (2004) on Cavineña, van Gijn (2006) on Yurakaré, Haude (2006) on Movima, and Danielsen (2007) on Baure.

## 2.2 Reconstructing the history and classification of the Guaporé-Mamoré languages

Table 2 reflects the genealogical classification of the Guaporé-Mamoré languages in so far as there is relatively sound evidence. Due to the scarcity of data at that time available, Greenberg's (1987) attempt at classifying all languages of the Americas resulted in an at times premature classification of the languages of the Guaporé-Mamoré region. For example, the evidence for the relatedness of the Jabutian languages Arikapú and Djeoromitxi is relatively sound, but the value of the claim that they belong to the Macro-Ge stock, which originates from Nimuendajú (2000), was not really substantiated. Only since recent field research has led to a completer documentation of the Jabutian languages, this 'macro' classification can be confirmed (Ribeiro and van der Voort *fc.*). The new data furthermore confirm the suspicion of Caspar (1955) that Mashubi, which used to be considered an extinct third member of the Jabutian family, in reality, is the same language as Arikapú. Moreover, Chiquitano, spoken in the Bolivian lowlands in the Santa Cruz Department and considered an isolate until now, may be Macro-Ge (Adelaar *p.c.*, Ribeiro *p.c.*).

Furthermore, the usual characterization in the literature of Aikanã, Kanoê and Kwaza as isolated languages is, in the absence of detailed evidence, worth just as much as the various proposals for their specific classification. In a recent article, van der Voort (2005b) made an attempt at the classification of these languages on the basis of new evidence, such as provided by Bacelar (2004). The data suggest that there is no conclusive proof for ancient genealogical relationships among these three languages. It seems equally likely that the languages are true isolates that underwent each other's influence during many centuries.

The research by Telles (2002) has further completed the picture of the Nambikwaran family. Even though Latundê is absent in Lowe's recent overview (1999), it is undeniably a member of the family and it belongs with Lakondê and Tawandê to one of the branches of Northern Nambikwaran. The Latundê were first contacted by Aikanã in 1975 and by Westerners in 1976. The question whether the Northern Nambikwaran languages, that include also Mamaindê and Nagarotê, represent distinct languages or a dialect complex, has not yet been settled. In the present chapter we represent Northern Nambikwara only by Latundê and Lakondê.

Furthermore, ongoing research by associates of the Museu Goeldi in Belém is leading to improvements of the internal classification of the Tupian family, for example by inclusion of the language of the recently contacted (1995) Akuntsu in the Tupí-Tuparí subfamily. Together with a handful of Kanoê this small group had been avoiding contact for decades, during which local ranchers committed several massacres against them. In 1995, however, personnel of FUNAI (Fundação Nacional do Índio, the Brazilian governmental department of indigenous affairs) succeeded in establishing permanent contact with them (dos Santos 1996, 2000). Furthermore, the vowel inventories of Karo and Puruborá suggest that these languages could belong to the same subfamily (Galucio and Gabas Jr. 2002). Finally, there may still be speakers of the Arikém language, which belongs to the same subfamily as Karitiana.

When consulting older sources on Rondonia, it is useful to remember that the language of the Karipuna used to belong to the Panoan family. What happened to this group is not known; they may have become extinct due to epidemics following contact with the workers on the Madeira-Mamoré railway. It is possible that the present Tupí-Guaraní speaking Karipuna represent a group that immigrated from Amazonas recently.

Other groups of the region also have become extinct in historical times, or have moved away, destination unknown. One such group is Palmella. Their language represented the only Cariban language of the region and the southernmost Cariban language of the family. The Palmellas were probably brought from the Guianas to Bolivia by Westerners, and are supposed to have moved to the right side of the Guaporé River after the Jesuit mission San Miguel on the Baures River was burned by the Portuguese in 1762. Word lists were recorded by Fonseca (1880–81), Nordenskiöld (n.d.), and Becker-Donner (1956), who met them at Pedras Negras in Brazil, located close to the mouth of the Colorado River, but nothing else is known about the fate of this group. In present-day sociological sources (e.g., Ricardo 1991) the name Paumelenhos is sometimes found for an ethnic group on the right side of the São Miguel River. Apparently there is no contact with this group and their ethnicity and language remain unknown.

There are several other uncontacted groups in the south of Rondonia, about the languages of which nothing is known. One of these groups that avoid contact numbers approximately 100 members, and their material culture is characterized by exceptionally long bows (Ricardo 1991: 447). That is probably the reason why some have speculated that the language belongs to the Tupí-Guaraní family (dos Santos 2000). The Bolivian Sirionó (TUPÍ-GUARANÍ) also have exceptionally long bows and on Nimuendajú's map (1981) some Sirionó pockets are located close to the Guaporé River.

In 1996, the rancher Dalafini in the municipality of Chupinguaia, Southeastern Rondonia, ordered his men to open fire on the members of an uncontacted group on the land he claimed. Hereafter a bulldozer attempted to extinguish the traces of the wrecked village to destroy evidence. Ever since, FUNAI personnel have been attempting to establish contact, but the sole surviving member of the group does not allow them to come close. It is not known what language the man speaks, but the material culture of his original group differs to some extent from the traditional peoples of the region (dos Santos 1998, 2000). This group is not listed in Table 2.

It is said that at the time of the Spanish conquest in the sixteenth century the Toromona, supposedly a Tacanan group led by the mythical cacique Tarona, formed an extremely effective barrier against the incursion of the Westerners into the southern part of Amazonia. The genocide that went hand in hand with the rubber boom (1880–1914) wiped the uncontacted Toromona from the official registers. Until today it still has not become clear whether the group perished during the rubber boom or retreated to inaccessible parts of the forest. Rumours have it that there is a mysterious group roaming through the forest to the south of the Araona territory at Puerto Araona (Iturralde Province, La Paz Department). Whether it actually concerns the ‘phantom’ Toromona or another group continues to be a mystery.

The last speakers of Itonama vividly recall the day that the *Chori* (Sirionó) came out of the forest to seek protection in the village of Magdalena against the aggressive Yanaigua. This must have been around 1930–35. Contrary to the Sirionó, the Yanaigua had small bows. It remains unknown what has happened to this uncontacted group since then. Some rumours have it that there is an uncontacted Tupí-Guaraní group called “Yanaigua” living on the Guarayos Forest Reserve, between the Río Grande and upper San Miguel, but this location is almost 1,000 kilometers to the south of Magdalena. Apart from the Toromona and Yanaigua, it is said that there is yet another nomadic group of Yuki (TUPÍ-GUARANÍ) of about 14 persons that still remains uncontacted.

Finally, the Panoan languages are all closely related, but the internal classification of the family has not been completely settled yet. Following evidence published by, among others, Key (1968, 1979) and Girard (1971), the Panoan languages are nowadays often lumped together with the Tacanan languages (cf. Kaufman 1990, Adelaar with Muysken 2004). Some linguists, however, still show some reticence towards the Pano-Tacanan classification. Loos (1999), for example, states that the lexical and grammatical similarities between these two families might be explained by areal diffusion as well. Fabre (2005) argues that the numerous lexical correspondences between Panoan and Tacanan may point at an old areal contact situation, which seems to be corroborated by the much more subtle morphological correspondences that, according to Fabre, would have a more solid explanation based on areal relations than on genetic ones. Adelaar with Muysken (2004: 419) do mention that certain morphological changes attested by Girard (1971) might be induced by contact between Panoan and Tacanan. If the correspondences between these two families are indeed based on an

earlier contact situation, there is all the more reason to postulate similar cases of areal diffusion within the Guaporé-Mamoré region.

As far as the classification and comparison of the unclassified languages on the Bolivian side of the Guaporé-Mamoré region goes, we are still in an initial phase. It appears that although these languages diverge dramatically at the lexical level, they share a fair amount of structural features. Table 3, in which 14 core lexical items are given for 11 Bolivian and seven Rondonian languages from the area, gives an impression to what extent these languages diverge lexically. The only possible similarity found in the 11 Bolivian languages is the word for 'water' in Baure (ARAWAKAN) and Yuki (TUPÍ-GUARANÍ). Obviously, there are some similarities between the Tupian subbranches, such as the words for 'firewood', and 'maize' in Yuki (TUPÍ-GUARANÍ) and Mekens (TUPÍ-TUPARÍ). The word for 'maize' seems to have spread anyhow in Rondonia, just as the words for 'star' and 'chicha'. Furthermore, Kwaza (UNCLASSIFIED) seems to have a word for 'moon' that is similar to the Mekens form. Note that the orthography of the word forms in Table 3 was homogenized, conforming to the IPA standard, with the exception of ⟨y⟩ (IPA [j]). The smileys indicate that we have not been able to find the relevant item.

In spite of the fact that, for instance, the unclassified languages Cayubaba and Itonama, and Arawakan Baure do not share any lexical items, they do share a considerable number of grammatical 'macro' features, such as, V-initial, head marking, alienable/inalienable distinction, lack of a grammaticalized gender system, verbal number, etc. These languages are geographically adjacent, but, as will be shown in Section 3.2, most of the relevant features are also found throughout the entire Guaporé-Mamoré region.

### 3. Grammatical and areal characteristics

The Amazonian languages tend to share certain specific characteristics, a number of which have been summed up in Derbyshire (1987), Payne (1990: 214), Dixon and Aikhenvald (1999: 8–10), and Campbell (1997: 349–351). Since the Amazonian languages belong to different unrelated families, this suggests that these characteristics define Amazonia as a linguistic area. This area contains a well-known subarea, the Vaupés region of the Northwestern Amazon (Aikhenvald 2002, Sorensen 1967). As suggested by Crevels and van der Voort (2002, 2006), the Guaporé-Mamoré region may harbor one or more linguistic areas as well. There are, for instance, several lexical, phonological and grammatical phenomena that point to areal diffusion in South-eastern Rondonia (van der Voort 2005b). Likewise, Table 4 in Section 3.2 shows us a considerable amount of shared grammatical features between unrelated languages in the Moxos area. In order to explore the full scope of the relations observed in the whole region, it is necessary to consider the basic grammatical characteristics of the languages in question. With respect to the possible genetic relationships of the many

Table 3. Basic lexical items of some languages in the Guaporé-Mamoré area

	ITONAMA	CAYUBABA	MOVIMA	YURAKARÉ	MOSETÉN	LEKO	CHIKUITANO	BAURE	CAVINENA
canoe	okoni	dapa	xurve	poxore	kasko	pele	☺	yafor	kwaba
eye	uxkʔururu	xokore	tʃora	tanti	wex	siri	sitox	mikis	yatuka
firewood	ubari	dore	koʔo	kummæ	sɔn	hamo	bapix hsues	yakis	kwati
house	uku	nika	ro-ya	sibbæ	akaʔ	won	pox	pari	etare
moon	tyaxkaʔkaxka	rare	yextʃo	fuwi	iwā	kurea	panx	kiher	badi
rain	itye	dabo	luʔluʔu	maʃita	aŋe	esa	tax	sowon	nei
smoke	ukuʔtya	naamo	buxru	bombo	xʃʃsaʔ	muswa	naunsix	kotisosokon	wani
star	okitʃi	rauwwa	didinkwa	puʃitʃe	orityāʔ	polea	nostoŋes	wahis	purari
sun	wapatʃa	maka	tinno	puyni	tsi:n	hena	sux	ses	iheti
water	wanuʔwe	kita	tomi	samma	õñiʔ	doa	tux	in	ena
jaguar	utyu	yedawa	rulrul	samu	itsiki	polo	nuitymix	ʃowekon	iba
chicha	uʔwaba	weiki	poʔso	xarru	ʃokdyeʔ	kathi	naxixh	marok	tupari
maize	udame	hiki	kwaxtaʔa	tʃilli	tyārāʔ	ta	noseox	tʃoros	ihike
manioc	tʃamaye	daduhu	hinala	ñowwo	oʔyi	ke	tabax	kahap	kwawe
	CHÁCOBO	YUKI	MEKENS	WARIʔ	ARIKAPÚ	AIKANÁ	KANOÉ	KWAZA	LATUNDÉ
canoe	noti	near	apikat	kanoa	kukəkə	kanowa	āṭapæṭæ	kanwā	☺
eye	biro	resa	o-ebaopap	tokʔ	həkare	ka-muka	ikiy	etyüi	inkin̄in-
firewood	karo	tata	otat	☺	pikə	hine	ini	hi	hih-
house	şobo	tai	ek	tʃirim	reko	keḁa	tɨy	asi	sih-
moon	oş	yasi	pakori	panawoʔ	kupa	ya	mitæ	hakuri	eyn-
rain	oi	ixoi	atsoap	tʃowiʔ	nāy	hane	væ	awe	mih- (V)
smoke	koʔini	tetatʃi	otat miik	tain	tʃio	tʃoni	pwā	hinünü	sin- (V)
star	wiʃtima	yasiriri	parobaro	piyoʔ	warəwarə	yote	warivari	waruwaru	tānkinin-
sun	bari	tēda	kiakop	tʃina	tahā	ya	kwikay	kosa	sün-
water	jini	i	iki	kom	mbi	hane	kuni	hā	nahoh-
jaguar	kamano	yagua	ameko	kopakaoʔ	kura	iʔive	opera	yereʃwa	loh-
chicha	jini	kiagō	tiero	tarakop	tʃuərə	māmāi	tʃero	mīw	nahon-
maize	şki	ibatʃi	atsisi	mapak	tʃitʃi	haki	atiti	atʃitʃi	keyat-
manioc	atsa	dio	tapşit	kop	mbu	nyapuri	tʃue	yo	lin-

unclassified languages of the region, such areal linguistic research is indispensable, since shared features that cannot be explained by diffusion may be candidates for genetic retention.

### 3.1 Criteria for the extent of the area

Since the extent of the suspected linguistic area under discussion is not yet known, the delimitation of the area of research is still provisional. Various criteria have been used for inclusion of certain languages, the geographic definition being the basic one for now: the Amazonian lowlands as drained by the Guaporé and Mamoré river systems.

As an additional, political criterion, we have aimed for a complete coverage of Rondonia, also including languages such as, for example, Kaxararí, Karitiana and Karo, spoken in the upper Madeira basin. Furthermore, the geographical criterion is not completely logical, since many indigenous societies live traditionally on the headwaters of tributaries of the Guaporé and Mamoré rivers, rather than close to these main arteries. Therefore, much contact took place not only between groups on the uppermost headwaters of proximate river subsystems, but also between groups on headwaters of different major river systems, such as lower Madeira tributary rivers and the Juruna-Tapajós and Paraguay rivers (cf. Section 1.1). In this respect, it would be difficult to defend the exclusion of, for example, Enawê-Nawê (ARAWAKAN), Parecí (ARAWAKAN), Rikbaktsá (MACRO-GE) and Irantxe (UNCLASSIFIED), whereas the Nambikwaran languages would be included.

Another important criterion is linguistic: as pointed out at the end of Section 2.2, certain similarities between languages that are not clearly due to genetic relationships are found throughout big parts of the region. This indicates that such similarities are not just the result of a particular history of contact between two languages, but that diffusion of a wider range has been going on for a long time. However, the true range of the linguistic area still has to be established. In the future, supporting linguistic criteria for the area may be envisioned and certain languages or regions probably will have to be included and others excluded with more compelling reason. In view of the distinction between areal traits and genetic correspondences, the study of all nearby unclassified languages is necessary. Therefore the Chiquitano language is included here, despite its southern location.

Finally, an important criterion is cultural: independent linguistic and archaeological information has confirmed the fact that earlier contacts existed between the ethnic groups pertaining to the three cultural areas described above; the Guaporé, the Moxos and the Tapajós-Madeira areas.

### 3.2 Defining features

The genetic linguistic diversity of the Guaporé-Mamoré region is enormous when compared to many other regions of the world. Especially the number of possibly

isolated languages is high. The region probably represents a “residual zone”, in Nichols’ (1992) terms, or, in other words, a region in which linguistic diversity has been accumulating as languages were pushed aside by those of politically and economically powerful populations in adjacent linguistically homogenous “spread zones”. When the distribution of language families and isolates is shown in different colors on a map (e.g., Nimuendajú 1981, Queixalos and Renault-Lescure 2000), the Guaporé-Mamoré region comes across as a focal region of prehistoric population movements. However, from the available linguistic material many similarities between the languages come to the fore. These similarities cross at least ten genetic linguistic borders, the borders of the three culture areas Moxos, Guaporé and Tapajós-Madeira and the geographical border of the Guaporé River itself. They include features such as minimal shared vocabulary, complex verbal morphology, evidentials, directionals, inclusive/exclusive distinction (cf. Crevels and Muysken 2005), and, surprisingly, lack of classifiers. The similarities may be phonological, grammatical, morphological or, sporadically, lexical in nature. The phonological similarities include, for example, nasal harmony. Grammatical similarities concern, for example:

- a high incidence of prefixes
- evidentials
- directionals
- verbal number
- lack of nominal number
- lack of classifiers
- inclusive/exclusive distinction

At the morphological level we find similar forms for grammatical categories. Note that at this point we have included forms from the Bolivian highland languages Uru (URU-CHIPAYAN) and Aymara (AYMARAN) as well:

- locative case KAN *-ni*; KWA *-na*; AIK *-ne*; AYM *-na*
- applicative derivation KAN *-ta-* or *-to-*; KWA *-ta-*; LAK *-ka*; KAR *-ta-*; MOS *-tya-* or *-te-*
- emphatic marker KWA/MEK *-te(te)*; KAN *-kete-*; GAV *tere*; SIR *te*
- semantically empty noun formative root *e-/i-* in KWA/KAN/LAK/CAV/ESE
- specific classifiers such as KWA/KAN/AIK *-mũ* ‘liquid’, KWA/KAN/AIK/ARI/LAK *-nũ* ‘powder’, ‘porridge’, KWA/AIK/NAM/SAB *-su* ‘bone’.

Some of these categories may rather characterize subareas than the entire region. Similar forms are also found at the lexical level.

- banana: AKU/TUP/MEK/KWA *apara*; AIK *dipara*; ITO *upatʃu*; MOV *pere*; MOS *chhi-peʔre*
- arrow: AKU *mambi*; ARI *mbu*; DJE *kubi*; KAN *mapi*; KWA *māʔi*; MEK *mampi*



- maize: AKU/KAN/MAK/TUP/WAY *atiti*; MEK *atsitsi*; ARI/DJE *tʃitʃi*; KWA *atʃitʃi*; ITO *atʃi*
- star: ARI *warəwarə*; DJE *wirəwirə*; KAN *wariwarĩ*; KWA/MEK/TUP/MAK *waruwaru*; URU *waruwaru*; AYM *wara*

A number of these correspondences do not belong to the general Amazonian features as listed in Payne (1990) and Dixon and Aikhenvald (1999). Several of these correspondences may be specific for the Guaporé-Mamoré region, and suggest the existence of a linguistic area, with possible subareas. Consider, for example, the semantically empty noun-formative root in combination with a bound root or classifier in examples (1) and (2), respectively:

- (1) Kanoê (UNCLASSIFIED; Bacelar 2004: 130)  
*i-kuta*  
 Ø-head  
 ‘(its) head’
- (2) Cavineña (TACANAN; Guillaume 2004: 72)  
*e-watçi=hu*  
 Ø-CLF:foot-LOC  
 ‘(it bit me) on the foot’

The same element is also found in this function in Ese Ejja (TACANAN), Kwaza (UNCLASSIFIED) and Latundê (NAMBIKWARAN). Movima (UNCLASSIFIED) has a dummy element *-i-* as well, but its behavior differs somewhat from the above examples. It occurs as a suffix with both verbal and certain nominal lexemes and as a prefix with specific person elements (Haude 2006: 61ff.).

The applicative element *-ta-* could be yet another example of grammatical correspondences across the region:

- (3) Kwaza (UNCLASSIFIED; van der Voort 2004: 359)  
*etay-tjate-wā*      *hoñe-ta-ki*  
 woman-POS-AO    hide-APPL-DECL  
 ‘He hid from his wife.’
- (4) Mosetén (UNCLASSIFIED; Sakel 2004: 322)  
*yāe tī-fāk-e-’*      *mō’ nanasi’*  
 I    APPL-angry-VS-3F she girl  
 ‘I was angry because of the girl’

Similar forms with the same function also occur in Kanoê (UNCLASSIFIED), Karo (TUPI-RAMARAMA), Lakondê/Latundê (NAMBIKWARAN) and have been attested in a wider region as well (cf. Wise 2002).

Certain linguistic correspondences among the Guaporé-Mamoré languages were noticed by earlier researchers who tried to provide genetic linguistic explanation or suggested language contact explanations. For example, Becker-Donner (1955) suggests that the unclassified Kanoê language is distantly related to Tupian. One could

of course hypothesize that the independent development of a language, given enough time, could lead to the invisibility of its genetic relationships with other languages. However, the general typological and lexical differences between Kanoë and Tupian are very big, whereas the observed correspondences are so close that they don't justify the necessary time depth in case they are really genetically related. Subsequently, Price (1972) supposed that Kanoë could be a creolized variant of Nambikwaran, but there is not even circumstantial evidence for this. If creolization were the only explanation for the linguistic correspondences across the region, then one is forced to assume that many adstrate languages have become extinct without even their names ever having been recorded – note that of certain known names of extinct groups, as mentioned in e.g., Caspar (1975: 6–10), it is impossible to tell whether they refer to people speaking extinct languages or extant languages. As pointed out in Section 3, Crevels and van der Voort (2002, 2006) suggested that certain close correspondences call for an areal linguistic explanation. Indeed, the native peoples of the region must have co-existed and intermarried during thousands of years while maintaining their native languages, and until this day multilingualism is not uncommon among the Indians of Rondonia.

As pointed out above, the usual characterization in the literature of Aikanã, Kanoë and Kwaza as isolated languages is worth just as much as the various proposals for their specific classification: since there were only a few word lists, all claims were based on very thin evidence. Van der Voort (2005b) made an attempt at the classification of the 'isolated' languages of Rondonia on the basis of new and comprehensive data collections and descriptive analyses of a number of languages. The evidence presented in the article suggests that genetic relationships among these three languages could be possible, be it that they must be more ancient than those that define the Tupian family for example.

The impression one gathers for the Guaporé-Mamoré region is that speakers of the languages under discussion have been in contact with one another for many centuries, leading to the emergence of a *Sprachbund*. Especially the similarity of both the forms and the structures of classifier systems, the empty noun-formative root and the applicative morpheme suggest this. The cultural and linguistic characteristics of the Guaporé-Mamoré region may be quite different from those of the Vaupés region of Northern Amazonia as described by Sorensen (1967) and Aikhenvald (2002), and of other areas (cf. Aikhenvald and Dixon eds. 2001). As an example of a cultural difference with the Vaupés region, no traditions of obligatory linguistic exogamy have been reported in the Guaporé-Mamoré region. As an example of a linguistic phenomenon, diffusion of bound grammatical morphemes can be observed in the Guaporé-Mamoré region (e.g., the applicative marker and certain classifiers), whereas Aikhenvald and Dixon (2001: 2) suggest that borrowing of grammatical categories, rather than forms, would be more likely to occur.

It is also possible that the Guaporé-Mamoré region presents us with a linguistic puzzle which combines both areal and genetic relationships in an inseparable kind of

‘mesh’, similar to the one described by Fortescue (2003) for the Chukotko-Kamchatkan languages of Siberia.

As already pointed out in Section 2.2, in an attempt to corroborate our claims about the Guaporé-Mamoré region as a linguistic area, we checked the presence or absence of a number of grammatical ‘macro’ features in 24 languages of the area, or more specifically, in 14 Bolivian and 10 Rondonian languages. We included the Bolivian highland languages Quechua, Aymara and Uru as control languages. Although some features may seem dependent on each other, we generally tried to select structural features that are not too strongly correlated. Thus, one might be sceptical about, for instance, the correlation between feature 18 ‘possession marked on possessum’ and feature 5 ‘head marking’. However, there are dependent-marking languages like Mose-tén (UNCLASSIFIED), Cavineña (TACANAN), Chácobo (PANOAN), and Karitiana (TUPÍ-ARIKÉM) that rather mark the possessum instead of the possessor. Reversely, head-marking languages like Kwaza (UNCLASSIFIED), Mekens (TUPÍ-TUPARÍ), and Kanoê (UNCLASSIFIED) mark the possessor rather than the possessum. The results of our survey are listed in Table 4. Note that the categories are listed in the form of binary features, even though the results often represent strong statistical tendencies rather than absolute values. The smileys indicate that we have not yet been able to establish whether the particular feature is present in the language or not, either for lack of data or because we have not been able to interpret the data sufficiently.

The question now is what we can deduce from Table 4 with respect to the status of the Guaporé-Mamoré region as a possible linguistic area. At first sight, many of the tested features seem to be typically Amazonian, but then it appears that the Guaporé-Mamoré languages do not comply with a number of these features. Most of the sample languages (66.7%) indeed show a relatively high degree of synthesis and an even higher percentage (75%) are head-marking languages. However, whereas Amazonian languages, for instance, are generally considered to have elaborated classifier and/or gender systems, only three (12.5%) out of the 24 sample languages have a grammaticalized gender system: Wari’ (CHAPACURAN), Movima and Mose-tén (both UNCLASSIFIED). Furthermore, only eight (33.3%) of the sample languages have classifiers, notably three Bolivian and five Rondonian languages. Another typical Amazonian feature is the asymmetrical morphology of many languages, or, in other words, Amazonian languages tend to have more suffix positions or slots than prefix positions. However, 11 (45.8%) out of the 24 Guaporé-Mamoré languages have a fairly symmetrical morphology in which prefix and suffix positions are relatively balanced.

When looking at features, which have not been labelled as typically Amazonian yet, we see that the Guaporé-Mamoré languages show a high incidence of evidentials (at least 79.2%), verbal number (at least 70.8%), directionals (at least 66.7%), inclusive/exclusive distinction (66.7%), and alienable/inalienable distinction (at least 54.2%). The frequency of the inclusive/exclusive distinction in the Guaporé-Mamoré languages we sampled is higher than the 57% in Crevels and Muysken’s (2005) central-western

Table 4. Some shared grammatical features in the Guaporé-Mamoré area

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Cayubaba	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0
Baure	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0
Mosetén	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1
Lakondê	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	0	0
Kwaza	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	0
Chiquitano	1	1	☺	1	1	1	1	1	0	1	1	1	1	1	☺	1	1	1	0	☺	0
Itonama	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	0	0
Gavião	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0
Quechua	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0	0	1	0
Aynara	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0	0	1	0
Leko	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	0	0	1	0
Chácobo	1	0	1	1	0	1	1	1	1	0	1	1	0	1	0	1	1	1	0	1	0
Movima	1	1	1	1	1	1	0	0	0	1	1	1	0	1	0	0	0	0	1	0	1
Mekens	1	1	1	0	1	1	0	1	1	1	1	☺	0	0	1	0	0	0	0	1	0
Aikaná	1	1	☺	0	0	1	1	1	0	0	☺	☺	1	1	1	0	1	0	1	1	0
Kanoé	1	1	0	0	1	☺	☺	1	1	1	1	0	1	0	1	0	1	0	1	0	0
Warí	1	0	1	1	1	1	☺	0	0	1	1	☺	1	0	0	0	1	1	0	0	1
Yuki	0	1	1	1	1	☺	0	0	1	1	1	1	1	0	1	1	0	0	0	0	0
Cavineña	0	0	1	1	0	0	1	1	1	0	1	1	0	1	0	1	1	0	0	1	0
Uru	☺	1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	0	0	0	0	0
Arikapú	1	1	☺	1	1	0	1	0	1	0	0	☺	0	0	1	0	1	1	0	0	0
Yurakaré	1	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
Karo	1	0	1	0	0	1	0	0	1	1	1	0	0	0	1	0	0	0	1	0	0
Karitiana	0	1	1	1	0	1	0	0	1	1	1	☺	0	0	1	0	0	0	0	0	0

1 = Subordination through nominalization; 2 = Cross-reference; 3 = Evidentiality; 4 = Vowel harmony; 5 = Head marking; 6 = Verbal number; 7 = Directionals; 8 = Polysynthesis; 9 = Postpositions; 10 = Inclusive/exclusive distinction; 11 = Alienable/inalienable distinction; 12 = Verb classification; 13 = Strictly Nom-Acc alignment system; 14 = Asymmetrical morphology; 15 = Nasal harmony; 16 = Nominal number; 17 = Head-Modifier; 18 = Possession marked on possessum; 19 = Classifiers; 20 = Switch reference; 21 = Grammaticalized system.

South-American sample and the 46% of the worldwide distribution sampled by Siewierska and Bakker (2005).

Taking into account the spread of the features within Rondonia and Bolivia separately, we must conclude that some features are more salient on the Rondonian side and others on the Bolivian side. Nominal number, for example, seems to be a typical Bolivian feature, occurring in 12 (85.9%) of the 14 Bolivian sample languages and in only one (10%) of the ten Rondonian languages. Classifier systems seem to be more common in the Rondonian (50%) than in the Bolivian languages (21.4%). Furthermore, we see that at least eight out of the ten Rondonian languages (80%) exhibit verbal number, whereas at least nine out of the 14 Bolivian languages score a slightly lower percentage (64.3%) on this feature. It is possible that this outcome may be slightly biased because of the inclusion of the highland languages Quechua, Aymara and Uru. Disregarding these three languages, the percentage of Bolivian languages exhibiting verbal number increases considerably to 81.8% (nine out of eleven languages).

The overall picture that emerges from Table 4 suggests that the Guaporé-Mamoré region is a linguistic area with two distinct subareas.

#### 4. Conclusion

The great diversity of languages in the Guaporé-Mamoré region must have accumulated during many centuries. As Nichols (1992: 22), in the context of prehistoric population movements, points out, “residual zones” are not necessarily explained as areas of refuge under the pressure of populations from other regions. They are more likely regions that have attracted populations for certain reasons. In the Guaporé-Mamoré region, the attracting factors could have been fertile lands and fishing grounds. Then again, in historic times (from the seventeenth century onwards) Rondonia did actually function as an area of refuge. As Leonel (1995) has shown on the basis of historical documents from the seventeenth century onwards, populations from the lower Madeira region through time migrated south and drove the existing populations even further south, and formed part of a push-chain mechanism that may have been created ultimately by the encroachment of Western civilization from the northeast, channeled through the Amazon River and its tributaries. Furthermore, Arawakan groups arrived at a later stage in the Bolivian lowlands than the traditional savanna groups Canichana, Cayubaba, Movima and Itonama. The Arawakan groups may have had an intermediary role in the trade relations between the Bolivian lowlands and the Andean highlands (cf. Métraux 1942, Denevan 1966, Hornborg 2005).

The distribution of formal and functional similarities across genetic linguistic boundaries indicates that the Guaporé-Mamoré region is a linguistic area. Although a small number of linguistic features show the clustering of certain languages in geographic subregions, the overall picture is that of a single linguistic area.

In the absence of abundant detailed historical documents, thorough archaeological reconstructions of the entire region and complete descriptions of all languages and cultures involved, our ideas about the historical context of the Guaporé-Mamoré languages will remain somewhat speculative. The complex linguistic setting of the Guaporé-Mamoré region makes it impossible anyway to paint its long term development with a large brush. The only way to know more seems to be by studying every particular aspect of this complex setting, both from linguistic viewpoints as well as from anthropological, archaeological and historical perspectives. The case of the Jabutian languages, which show a clear connection with geographically quite remote languages may serve as an illustration of the importance of historical linguistics in all of this. Fortunately, the long-term interdisciplinary study that is required in order to arrive at a more general picture has started already a century ago with the likes of Nordenskiöld.

## Acknowledgements

We are very grateful to Swintha Danielsen, Nilson Gabas Jr., Vilacy Galucio, Rik van Gijn, Antoine Guillaume, Katharina Haude, Simon van de Kerke, Denny Moore, Jeanette Sakel, and Luciana Storto for helping us interpreting their data. We are also indebted to Pieter Muysken for his useful comments. The excellent map was made by Willem Doelman. Needless to say, none of these persons are responsible for any errors, infelicities, or misinterpretations in this paper.

## Abbreviations

AIK = Aikanã; AKU = Akuntsu; ARI = Arikapú; AYM = Aymara; AO = animate object; APPL = applicative; CAV = Cavineña; CLF = classifier; DECL = declarative; DJE = Djeoromitxi; ESE = Ese Ejja; F = feminine; GAV = Gavião; ITO = Itonama; KAN = Kanoê; KAR = Karo; KWA = Kwaza; LAK = Lakondê/Latundê; LOC = locative; MAK = Makuráp; MEK = Mekens; MOS = Mosetén; MOV = Movima; POS = possessive; SIR = Sirionó; TUP = Tuparí; URU = Uru; VS = verbal stem marker; WAY = Wayuru; 3 = third person; Ø = zero morpheme.

## References

- Adelaar, W.F.H., with the collaboration of P.C. Muysken. 2004. *The languages of the Andes*. Cambridge: CUP.
- Aikhenvald, A.Y. & Dixon, R.M.W. 2001. Introduction. In *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, A.Y. Aikhenvald & R.M.W. Dixon (eds.). Oxford: OUP.

- Aikhenvald, A.Y. 2002. *Language contact in Amazonia*. Oxford: OUP.
- Alves, P.M. 1991. Análise fonológica preliminar da língua Tuparí. MA thesis, Universidade de Brasília.
- Alves, P.M. 2004. O léxico do Tuparí: Proposta de um dicionário bilingüe. PhD dissertation, Universidade Estadual Paulista “Júlio de Mesquita”.
- Angenot de Lima, G. 2001. *Description phonologique, grammaticale et lexicale du moré, langue amazonienne de Bolivie et du Brésil*, 2 vols. Porto Velho: Editora da Universidade Federal de Rondônia.
- Araujo, G.A. de. 2004. *A grammar of Sabanê, a Nambikwaran language*. Utrecht: LOT Publications.
- Bacelar, L. 1994. Fonologia segmental da língua Kanoê: Uma análise preliminar. *Signótica: Revista do Mestrado em Letras e Linguística* 6: 59–72.
- Bacelar, L. 2004. Gramática da língua Kanoê. PhD dissertation, Katholieke Universiteit Nijmegen.
- Baptista Morales, J. 1995. Los Misioneros Jesuitas de Mojos. *Yachay* 21: 71–90, Cochabamba: Universidad Católica Boliviana.
- Becker-Donner, E. 1955. Notizen über einige Stämme an den rechten Zuflüssen des Rio Guaporé. *Archiv für Völkerkunde*, Band X: 275–343. Vienna: Wilhelm Braumüller Universitäts-Verlag G.M.B.H.
- Becker-Donner, E. 1956. Archäologische Funde vom mittleren Guaporé (Brasilien). *Archiv für Völkerkunde*, Band XI: 202–249. Vienna: Wilhelm Braumüller Universitäts-Verlag GMBH.
- Borella, C. In preparation. Grammatical description of Sararé. PhD dissertation, Vrije Universiteit Amsterdam.
- Boswood, J. 1971. Phonology and morphology of Rikbaktsa and a tentative comparison with languages of the Tupi and Jê families. MA thesis, Reading University.
- Braga, A. de Oliveira. 1996. A fonologia segmental e aspectos morfofonológicos da língua Makurap (Tupi). *Moara: Revista dos cursos de pós-graduação em letras da UFPA* 4: 7–22 [Estudos de línguas indígenas].
- Braga, A. de Oliveira. 2005. Aspects morphosyntaxiques de la langue makurap. PhD dissertation, Université Toulouse II.
- Camp, E.L. & Liccardi, M.R. 1967a. Itonama. *Bolivian Indian grammars* 1, E. Matteson (ed.), 257–352.
- Camp, E.L. & Liccardi, M.R. 1967b. *Itonama: castellano e inglés* [Vocabularios Bolivianos 6]. Riberalta: Instituto Lingüístico de Verano.
- Campbell, L. 1997. *American Indian languages: The historical linguistics of native America* [Oxford Studies in Anthropological Linguistics 4]. New York NY: OUP.
- Caspar, F. 1955. A expedição de P.H. Fawcett à tribo dos Maxubi em 1914. *Anais do XXXI Congresso Internacional de Americanistas*, São Paulo, 113–120.
- Caspar, F. 1975. *Die Tupari: Ein Indianerstamm in Westbrasilien* [Monographien zur Völkerkunde herausgegeben vom Hamburgischen Museum für Völkerkunde VII]. Berlin: Walter de Gruyter.
- Créqui-Montfort, G. de & Rivet, P. 1914a. Linguistique bolivienne. La langue Mobima. *Journal de la Société de Américanistes de Paris* 11: 183–211.
- Créqui-Montfort, G. de & Rivet, P. 1914b. La langue kanichana. *Mémoires de la Société de Linguistique de Paris* 18: 354–377.
- Créqui-Montfort, G. de & Rivet, P. 1914c. Linguistique bolivienne. La langue Kayuvava. *Le Museon* 315: 121–162.

- Créqui-Montfort, G. de & Rivet, P. 1916. La langue Itonama I. *Mémoires de la Société de Linguistique de Paris* 19: 301–322.
- Créqui-Montfort, G. de & Rivet, P. 1917–1920. Linguistique bolivienne. La langue Kayuvava. *International Journal of American Linguistics* 1(4): 245–265.
- Créqui-Montfort, G. de & Rivet, P. 1918. La langue Itonama II. *Mémoires de la Société de Linguistique de Paris* 20: 26–57.
- Crevels, M. 2002. Why speakers shift and languages die: An account of language death in Amazonian Bolivia. In *Current studies on South American languages*, M. Crevels et al. (eds.), 9–30.
- Crevels, M. 2007. A Grammar of Itonama, an Amazonian language from northeast Bolivia. Radboud Universiteit Nijmegen. Ms.
- Crevels, M. & Adelaar, W.F.H. 2000–2006. South America. *UNESCO Red Book of Endangered Languages*. University of Tokyo. Available at: [http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/SAmerica/SA\\_index.cgi](http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/SAmerica/SA_index.cgi).
- Crevels, M., van de Kerke, S., Meira, S. & van der Voort, H. (eds.). 2002. *Current studies on South American languages* [ILLA 3]. Leiden: CNWS Publications.
- Crevels, M. & Muysken, P. 2005. Inclusive/exclusive distinctions in the languages of central-western South America. *Clusivity: Typology and case studies of inclusive-exclusive distinction*, E. Filimonova (ed.), 311–338.
- Crevels, M. & van der Voort, H. 2002. A Mamoré-Guaporé linguistic area? Paper presented at the LRTC (Language Typology Resource Centre) Annual Meeting, September 28, 2002, Utrecht.
- Crevels, M. & van der Voort, H. 2006. The Guaporé-Mamoré region: A linguistic area? Paper presented at the DGfS (Deutsche Gesellschaft für Sprachwissenschaft) Annual Meeting, February 22–24, 2006, Bielefeld.
- Danielsen, S. 2007. *Baure: An Arawak language of Bolivia* [ILLA 6]. Leiden: CNWS Publications.
- Denevan, W.M. 1966. *The aboriginal cultural geography of the Llanos de Mojos of Bolivia*. Berkeley CA: University of California Press.
- Derbyshire, D.C. 1987. Morphosyntactic areal characteristics of Amazonian languages. *International Journal of American Linguistics* 53(3): 311–326.
- Dixon, R.M.W. & Aikhenvald, A.Y. 1999. Introduction. In *The Amazonian languages*, R.M.W. Dixon & Alexandra Y. Aikhenvald (eds.), 1–21.
- Dixon, R.M.W. & A.Y. Aikhenvald (eds.). 1999. *The Amazonian languages*. Cambridge: CUP.
- Duran, I.R. 2000. Descrição fonológica e lexical do dialeto Kaw Tayo da língua Moré. M.A. thesis, Universidade Federal de Rondônia.
- Emkow, C. 2007. A grammar of Araona, an Amazonian language of northern Bolivia. PhD dissertation, La Trobe University, Victoria.
- Erickson, C. 2000a. An artificial landscape-scale fishery in the Bolivian Amazon. *Nature* 408: 190–193.
- Erickson, C. 2000b. Lomas de ocupación en los Llanos de Moxos. Arqueología de Tierras Bajas, A.D. Coirolo & R. Bracco Boksar. Montevideo, Uruguay: Comisión Nacional de Arqueología, Ministerio de Educacion y Cultura, 207–226.
- Everett, D.L. & Kern, B. 1997. *Wari': The Pacaas Novos language of Western Brazil*. London: Routledge.
- Fabre, A. 2005. *Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos*. Pano-Takana. On-line available at: <http://butler.cc.tut.fi/~fabre/BookInternetVersio/Dic=PanoTakana.pdf>.



- Faust, N. & Loos, E. 2002. *Gramática del idioma yaminahua* [Serie Lingüística Peruana 51]. Peru: Instituto Lingüístico de Verano.
- Filimonova, E. (ed.). 2005. *Clusivity: Typology and case studies of inclusive-exclusive distinction* [Typological Studies in Language 63]. Amsterdam: John Benjamins.
- Fonseca, J.S. da. 1880–1881. *Viagem ao redor do Brasil, 1875–1878*. 2 vols. Rio de Janeiro.
- Fortescue, M. 2003. Diachronic typology and the genealogical unity of Chukotko-Kamchatkan. *Linguistic Typology* 7: 51–88.
- França, M.C.V. de. 2002. Aspectos da fonologia lexical e pós lexical da língua Oro Towat'i (Oro Win). PhD dissertation, Universidade Federal de Rondônia.
- Gabas Jr., N. 1989. Estudo fonológico da língua Karo (Arara de Rondônia). MA thesis, Dept. of Linguistics, IEL-UNICAMP.
- Gabas Jr., N. 1999. A grammar of Karo, Tupi (Brazil). PhD dissertation, University of California at Santa Barbara.
- Galeote Tormo, J. 1996 [1993]. *Manityana auki besiro. Gramática moderna de la lengua Chiquitana y vocabulario básico*, 2<sup>nd</sup> ed. Santa Cruz de la Sierra: Centro de Estudios Chiquitanos.
- Galucio, A.V. 1996. Mekens Syntax: A preliminary survey. MA thesis, University of Chicago.
- Galucio, A.V. 2001. The Morphosyntax of Mekens (Tupi). PhD dissertation, University of Chicago.
- Galucio, A.V. 2005. Puruborá: Notas etnográficas e lingüísticas recentes. *Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas* 1(2) 159–192. Belém.
- Galucio, A.V. & Gabas Jr., N. 2002. Evidências de agrupamento genético Karo-Puruborá, tronco Tupi. Paper presented at the XVII Encontro Nacional da ANPOLL, Gramado. (Abstract in: ANPOLL Boletim Informativo 31: A pós-graduação em letras e lingüística no Brasil: memórias e projeções. p. 163).
- Galvão, E. 1960. Áreas culturais indígenas do Brasil; 1900–1959. *Boletim do Museu Paraense Emílio Goeldi*, n.s. [Antropologia 8] (repr. in: Galvão, Eduardo. 1979. *Encontro de sociedades: índios e brancos no Brasil*. Rio de Janeiro: Paz e Terra, 193–228), and in *Indians of Brazil in the twentieth century*, J. Hopper (ed.). 1967. Washington, D.C.: Institute for Cross-Cultural Research, 167–205).
- Gill, W. 1999a. Diccionario Tsimane-Castellano. Cochabamba: Misión Nuevas Tribus. Ms.
- Gill, W. 1999b. A pedagogical grammar of the chimane (tsimane') language. Cochabamba: The New Tribes Mission. Ms.
- Girard, V. 1971. *Proto-Takanan Phonology* [University of California Publications in Linguistics 70]. Berkeley CA: University of California Press.
- Greenberg, J. 1987. *Language in the Americas*. Stanford CA: Stanford University Press.
- Gijn, R. van. 2006. A grammar of Yurakaré. PhD dissertation, Radboud Universiteit Nijmegen.
- Guerra, M. de Lacerda. 2004. Aspects of Suruí phonology and phonetics. MA thesis, Université Libre de Bruxelles.
- Guillaume, A. 2004. A Grammar of Cavineña, an Amazonian language of northern Bolivia. PhD dissertation, La Trobe University, Victoria.
- Guillaume, A. 2005. Diccionario Maropa-Castellano con apuntes gramaticales. CNRS-Université Lumière Lyon 2/ELDP. Ms.
- Haude, K. 2006. A Grammar of Movima. PhD dissertation, Radboud Universiteit Nijmegen.
- Hemmuier, R. 2005. Studien zur historischen Morphosyntax des Sirono. MA thesis, Ludwig-Maximilians-Universität, Munich.
- Hornborg, A. 2005. Ethnogenesis, regional integration, and ecology in prehistoric Amazonia: Toward a system perspective. *Current Anthropology* 46(4): 589–620.

- Iggesen, O. In preparation. A grammar of Chácobo. RCLT/La Trobe University, Victoria.
- Judy, R.A. & Judy, J. 1962. *Movima y castellano* [Vocabularios Bolivianos 1]. Cochabamba: Instituto Lingüístico de Verano.
- Judy, R.A., & Judy, J. 1967. Movima. *Bolivian Indian grammars* 2, E. Matteson (ed.), 353–408.
- Kaufman, T.S. 1990. Language history in South America: What we know and how to know more. In *Amazonian linguistics. Studies in Lowland South American languages*, Doris L. Payne (ed.), 13–73.
- Kerke, S. van de. 1998. Verb formation in Leko: Causatives, reflexives and reciprocals. *Typology of verbal categories* [Linguistische Arbeitsberichte 382], L. Kulikov & H. Vater (eds.), 195–203. Tübingen: Max Niemeyer Verlag.
- Kerke, S. van de. 2000. Case marking in the Leko language. In *Indigenous languages of Lowland South America* [ILLA 1], H. van der Voort & S. van de Kerke (eds.), 25–39. Leiden: CNWS Publications.
- Kerke, S. van de. 2002. Complex verb formation in Leko. In *Current studies on South American languages*, M. Crevels et al (eds.), 241–255.
- Kerke, S. van de. 2006. Object cross-reference in Leko. In *What's in a verb?* [LOT Occasional Series 5], G.J. Rowicka & E.B. Carlin (eds.), 171–188. Utrecht: Netherlands Graduate School of Linguistics (LOT).
- Key, H. 1967. *Morphology of Cayuvava* [Janua Linguarum, series practica 53]. The Hague: Mouton.
- Key, H. 1975. *Lexicon-dictionary of Cayuvava-English*. Huntington Beach CA: Summer Institute of Linguistics.
- Key, M.R. 1968. *Comparative Tacanan phonology; with Cavineña phonology and notes on Pano-Tacanan relationship* [Janua Linguarum, series practica 50]. The Hague: Mouton.
- Key, M.R. 1979. *The grouping of South American Indian languages*. Tübingen: Gunter Narr.
- Kroeker, M. 2001. A descriptive grammar of Nambikuara. *International Journal of American Linguistics* 67(1): 1–87.
- Leonel, M. 1995. *Etnodicéia Uruéu-au-au: O endocolonialismo e os Índios no centro de Rondônia, O direito à diferença e à preservação ambiental*. São Paulo: EDUSP/IAMA/FAPESP.
- Lévi-Strauss, C. 1948. Tribes of the right bank of the Guaporé river. In *Handbook of South American Indians* 3, J.H. Steward (ed.), 370–379. Washington DC: Smithsonian Institution.
- Lévi-Strauss, C. 1955. *Tristes tropiques*. Paris: Librairie Plon.
- Loos, E.E. 1999. Pano. *The Amazonian languages*, R.M.W. Dixon & A.Y. Aikhenvald (eds.), 227–250.
- Lowe, I. 1999. Nambiquara. In *The Amazonian languages*, R.M.W. Dixon & A.Y. Aikhenvald (eds.), 268–291.
- Maldi, D. 1991. O Complexo Cultural do Marico: Sociedades Indígenas dos Rios Branco, Colorado e Mequens, Afluentes do Médio Guaporé. *Boletim do Museu Paraense Emílio Goeldi*, Antropologia 7(2): 209–269. Belém.
- Matteson, E. (ed.). 1967. *Bolivian Indian grammars* [Summer Institute of Linguistics Publications in Linguistics and Related Fields 16], 2 vols. Norman: Summer Institute of Linguistics of the University of Oklahoma.
- Meer, T. van der. 1982. Fonologia da língua Suruí. MA thesis, Universidade Estadual de Campinas.
- Métraux, A. 1942. *The native tribes of Eastern Bolivia and Western Matto Grosso* [Bureau of American Ethnology Bulletin 134]. Washington DC: Smithsonian Institution.

- Monserat, R.M.F. 2000. A língua do povo Mynky. PhD dissertation, Universidade Federal de Rio de Janeiro.
- Moore, D. 1984. Syntax of the language of the Gavião Indians of Rondônia, Brazil. PhD dissertation, The City University of New York.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago IL: The University of Chicago Press.
- Nimuendajú, C. 1981. *Mapa Etno-Histórico de Curt Nimuendaju*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística (IBGE).
- Nimuendajú, C. 2000 [1923–1942]. *Cartas do Sertão de Curt Nimuendajú para Carlos Estevão de Oliveira*, T. Hartmann (ed.). Lisbon: Museu Nacional de Etnologia/Assírio & Alvim.
- Nordenskiöld, E. 1915. *Forsknningar och Äventyr i Sydamerika*, Stockholm: Albert Bonniers Förlag.
- Nordenskiöld, E. N.d. Palmella. Notebook from the 1913–1914 expedition, archive of the Världskulturmuseet, Göteborg. Ms.
- Olza Zubiri, J. 2002. *Gramática Moja Ignaciana (morfosintaxis)*. Caracas: Publicaciones UCAB.
- Orbigny, A. d'. 1839. *Voyage dans l'Amérique méridionale*, vol. 4, 1<sup>st</sup> part *L'homme américain (de l'Amérique méridionale), considéré sous ses rapports physiologiques et moraux*. Paris.
- Payne, D.L. 1990. Morphological characteristics of Lowland South American languages. In *Amazonian linguistics: Studies in Lowland South American languages*, D.L. Payne (ed.), 213–241.
- Payne, D.L. (ed.). 1990. *Amazonian linguistics: Studies in Lowland South American languages*. Austin TX: University of Texas Press.
- Pires, N.N. 1992. Estudo da gramática da língua Jeoromitxi (Jabuti) - aspectos sintáticos das cláusulas matrizes. MA thesis, Departamento de Linguística, Universidade Estadual de Campinas.
- Price, P.D. 1972. Nambiquara society. PhD dissertation, University of Chicago.
- Price, P.D. 1978. The Nambiquara linguistic family. *Anthropological Linguistics* 20: 14–37.
- Queixalós, F. & O. Renault-Lescure (eds.). 2000. *As línguas amazônicas hoje*. São Paulo: Instituto Socioambiental.
- Ribeiro, E. & van der Voort, H. Forthcoming. Nimuendajú was right: The inclusion of the Jabuti language family in the Macro-Jê stock. *International Journal of American Linguistics*.
- Ricardo, C.A. (ed.). 1991. *Povos Indígenas no Brasil 1987/88/89/90*. São Paulo: Centro Ecumênico de Documentação e Informação (CEDI).
- Rose, F. In preparation. A grammar of Trinitario. CELIA/CNRS-IRD, Villejuif.
- Sakel, J. 2004. *A Grammar of Mose'tén* [Mouton Grammar Library 33]. Berlin: Mouton de Gruyter.
- Salvatierra, C. 2005. Morfología trinitaria. BA thesis, Universidad Mayor San Simón, Cochabamba (Bolivia).
- Sampaio, W.B. de Araujo. 2001. As línguas tupi-kawahib: Um estudo sistemático filogenético. PhD dissertation, Universidade Federal de Rondônia.
- Santos, M. dos. 1996. Índios acossados em Rondônia. In *Povos Indígenas no Brasil 1991/1995*, C.A. Ricardo (ed.), 550–553. São Paulo: Instituto Socioambiental.
- Santos, M. dos. 1998. Proposta para interdição isolado Tanaru. Typescript, 5 pp. Vilhena.
- Santos, M. dos. 2000. Os “Isolados” de Rondônia. In *Povos Indígenas no Brasil 1996/2000*, C.A. Ricardo (ed.), 594–595. São Paulo: Instituto Socioambiental.
- Siewierska, A. & Bakker, D. 2005. Inclusive and exclusive in free and bound person forms. In *Inclusivity: Typology and case studies of inclusive-exclusive distinction*, E. Filimonova (ed.), 151–178.
- Silva, L. de Jesus. 2005. Aspectos da fonologia e morfologia da língua Rikbátsa. MA thesis, Universidade de Brasília.

- Sorensen, A.P. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69: 670–684.
- Steward, J.H. & Faron, L.C. 1959. *Native peoples of South America*. New York NY: McGraw-Hill.
- Storto, L. 1999. Aspects of a Karitiana grammar. PhD dissertation, MIT.
- Telles, S. 2002. Fonologia e gramática Latundê/Lakondê. PhD dissertation, Vrije Universiteit Amsterdam.
- Vasconcelos, I.P. 2004. Aspectos fonológicos e morfofonológicos da língua Aikanã. PhD dissertation, Universidade Federal de Alagoas.
- Villafañe, L. 2004. Gramática Yuki: Lengua Tupí-Guaraní de Bolivia. PhD dissertation, Katholieke Universiteit Nijmegen.
- Voort, H. van der. 2004. *A grammar of Kwaza* [Mouton Grammar Library 29]. Berlin: Mouton de Gruyter.
- Voort, H. van der. 2005a. Arikapu dictionary with Djeoromitxi equivalents. Radboud Universiteit Nijmegen. Ms.
- Voort, H. van der. 2005b. Kwaza in a comparative perspective. *International Journal of American Linguistics* 71(4): 365–412.
- Voort, H. van der. 2006. Os povos Jabutí. In *Povos indígenas. Lista de povos*. São Paulo: Instituto Socioambiental. Available at: <http://www.socioambiental.org/pib/english/whwhhow/wichpe.shtm>.
- Vuillermet, Marine. In preparation. A grammar of Ese Ejja. PhD dissertation, Université Lumière Lyon 2.
- Wise, M.R. 2002. Applicative affixes in Peruvian Amazonian languages. In *Current Studies on South American Languages*, M. Crevels et al (eds.), 329–344.



# An integrated areal-typological approach

## Local convergence of morphosyntactic features in the Balkan Sprachbund

Olga Mišeska Tomić

University of Leiden Centre for Linguistics

The author examines some Balkan Sprachbund features that appear in neighbouring dialects of individual languages, but are absent from the other dialects of the languages to which these dialects belong; as well as some features that function in neighbouring Balkan languages of different families, but are absent from the other Balkan languages of the same family. Displaying individual typological features of the Balkan languages in interaction with other features in the structure of the DP or the clause and taking into consideration sociolinguistic factors and dialectal variation, she tries to identify structural similarities, establish links between structural differences, and shed light upon a number of pertinent Balkan Sprachbund questions.

### 1. Prologue

Thought often only remotely genetically related (and sometimes totally unrelated), the languages of the Balkans exhibit remarkable structural similarities. These structural similarities were first signalled by Jernej Kopitar (1829), who pointed out that the languages spoken south of the Danube have analogous forms expressed through “different language material.” Miklosich (1861) examined the facts more closely and singled out a number of distinct common features of the languages in the area. Nikolai Trubetzkoy (1928: 17–18), observing that the languages of the Balkans (a) have similar sentence-structure but show no systematic sound correspondences and (b) have a great number of common “cultural” words, though their basic vocabularies may be diametrically different, referred to relationships such as the ones exhibited among these languages as “Sprachbund”.<sup>1</sup>

---

1. Note that Weinreich (1953: 378) argues against the use of the term “Sprachbund”, its fault being that “it implies a unit, as if a language either was or was not a member of a given Sprachbund.” More recently, Reiter (1994) attempts to do away with the term “balkanism” and “language union”. Hindrichs (1999) also tends to avoid the term “Balkan”, opting for the term “Südeuropa”.

## 2. The area

### 2.1 Geographic and Ethnic Balkans

The modern Balkan states share a geographical unity and historical heritage dating back to inhabitation during the Lower Palaeolithic times, 200,000–100,000 BC (cf. Carter 1977: 1). In the course of the first millennium of the modern era, however, due to the uneven influence of Rome in the territories in and around the Balkans, which the empire had conquered, two different civilisations developed on the peninsula.

During the period before Christ, Roman influence on the Balkans was chiefly along the Adriatic and Ionic coast, concentrating in coastal towns. In the first century AD the Romans began pushing their frontiers inland across the peninsula. As they advanced, they established forts and small towns and built roads to connect them with the coast. The countryside beyond the forts and the towns remained populated with indigenous population, and experienced little or no Roman influence.

In AD 324, the emperor Constantine chose to live in the East of the Roman empire and established a new centre, Constantinople. This marked the beginning of the end of the centralised Roman rule, which was precipitated after the death of emperor Theodosius I, who divided the empire between his two sons, into Old Rome in the west and New Rome in the East. From the fifth century onwards, two distinct spheres of influence developed on the Balkans—Latin in the western part of the peninsula and Greek in its Eastern part, eventually each with its own Christian church. The empire remained “Roman” and, despite the divisions of its territory, it was always seen as a single unit.<sup>2</sup> The citizens of both the western and Eastern part called themselves Romans, though Greek came to be the predominant language in the east.<sup>3</sup>

After the Slavic invasions of the late sixth and seventh centuries cut off east-west communications of the empire, the differences between the two parts increased rapidly. The dividing line between Old Rome and New Rome ran from Sremska Mitrovica – a Serbian city on the river Sava, close to the border with Croatia, to Scadar – a Montenegrin city on a lake bordering with Albania, close to the Adriatic sea (see Map 1). This is basically the line dividing the Catholic from the Orthodox sphere of influence at the present.

When the Austro-Hungarian and Ottoman empires came into the picture, the ethnic borders of the Balkans, distinct from its geographical borders, were definitely shaped. The western borders were moved eastwards – what is now Slovenia and part of Croatia (west of the river Una) “went out” of the “ethnic” Balkans, whereas in the north and north-east the boundary extended far beyond the Danube, to encompass the Serbian

---

2. The term *Byzantine* was an invention of Renaissance scholars after the fall of the empire; it was not used by its contemporaries. (cf. Fine 1991: 15–16)

3. The term *Hellene* ‘Greek’ connoted a pagan; Christian Greeks called themselves “Romans” (cf. Fine 1991: 15–16).





province of Vojvodina and most of present-day Romania (see Map 1). It is through these “ethnic Balkans” that the Balkan Sprachbund linguistic features spread.

## 2.2 Membership

There are quite a number of languages spoken in the modern Balkan states. The “Balkan Sprachbund” features spread in those languages which have been spoken since the early middle ages – the Slavic languages Macedonian, Bulgarian and Serbo-Croatian, the Eastern Romance languages Romanian, Aromanian and Megleno-Romanian, Albanian, Modern Greek, the Balkan Romani dialects, and to some extent in Judeo-Spanish and Turkish.<sup>4</sup> Macedonian is spoken as a majority language in Macedonia, and as a minority language in a large area in northern Greece, as well as in areas in Bulgaria and Albania adjacent to Macedonia. Bulgarian is spoken as a majority language in Bulgaria and as a minority language in a restricted area in Serbia, adjacent to Bulgaria. Serbo-Croatian<sup>5</sup> is spoken as a majority language in Bosnia and Hercegovina, Croatia and Serbia and Montenegro, and as a minority language in areas in Hungary and Romania, adjacent to Croatia and Serbia. Romanian is spoken as a majority language in Romania, and as a minority language in areas in Serbia, adjacent to Romania and northern Bulgaria. Aromanian is spoken in central and western Macedonia, southern Albania and central and north-western Greece. Megleno-Romanian is spoken by a few thousand people originally living in several villages in what is now south-eastern Macedonia and the adjacent part of Greece. Albanian is spoken as a majority language in Albania and the (southern Serbian) province of Kosovo and Metohia, and as a minority language in western Macedonia and north-western Greece. Modern Greek is spoken as a majority language in Greece and as a minority language in southern Albania. Balkan Romani is spoken by scattered communities throughout the Balkans. Balkan Turkish is spoken as a majority language in the Balkan part of Turkey and as

---

4. In the period when the Balkan Sprachbund features were developing, Turkish was the most widely spoken language on the Balkans. Nevertheless, the Balkan Sprachbund features are less prominent in it. Partly because of its non-Indo-European structure, which makes it difficult to find correspondent exponents for the phenomena exhibited in the other languages (all of them Indo-European), and partly because of its dominant social status, Turkish was an instigator for the development of the Sprachbund, more than a participant in it.

5. Upon the disintegration of Yugoslavia, Serbo-Croatian dissolved into Serbian, Croatian and Bosnian. The grammatical structures of the three “successors” do not, however, substantially differ from one another. I am using the term “Serbo-Croatian” when speaking of the grammatical structure of the language/languages, and “Serbian” or “Croatian”, when referring to the socio-linguistic categories “standard language” or “dialect”.



Map 2. The Balkan States

a minority language by substantial communities in southern Bulgaria and by restricted communities in Macedonia and Kosovo and Metohia. Judeo-Spanish was spoken extensively in the major cities in Bosnia and Hercegovina and Macedonia; but after World War Two the number of its speakers is very small. Turkish is spoken in Macedonia, north-eastern Greece and in a substantial part of southern and south-eastern Bulgaria. (See Map 2)

Not all Balkan languages have an equal share in the Balkan Sprachbund; moreover, different analysts have different rankings of membership. Weigand (1928) sees Albanian, Romanian and Bulgarian as typically Balkan languages, taking Greek, Serbian and Turkish to be only “geographically Balkan”. Sandfeld (1930) finds that the Balkan features are typically exhibited in Greek, Bulgarian, Albanian and possibly Serbian, while Turkish has many lexical concordances with each and everyone of them. For Schaller (1975), Albanian, Romanian, Bulgarian and Macedonian are Balkan languages of first degree, Greek and Serbian – Balkan languages of second degree, while Turkish is a Balkan language of third degree. For Birnbaum (1968), Romanian and Aromanian are “most Balkan”; then come Bulgarian, Macedonian, Modern Greek and Albanian, in this order. Solta (1980: 7) maintains that, when individual dialects are taken into consideration, the northern Greek and southeastern Serbian dialects could also be treated as true Balkan language codes. Lindstedt (2000) singles out twelve grammatical Balkanisms and examines their attestation in five language groups of the Sprachbund: Greek, Albanian, Balkan Romance – including Romanian, Aromanian and Megleno-Romanian, Balkan Slavic – including Bulgarian, Macedonian and the Torlak Serbian dialects, and Balkan Romani.<sup>6</sup> Giving one point for each “full” attestation and half a point for a partial attestation, he calculates the Balkanization indices of the language groups discussed and receives the following scores: Balkan Slavic 11.5; Albanian 10.5; Balkan Romance and Greek 9.5; Balkan Romani 7.5. According to this computation, the Balkan Slavic languages are the most Balkanized ones.

### 3. Methodology

#### 3.1 Lists

Apart from discussing the origin of Balkanisms, Balkan linguistic studies have traditionally been concerned with compiling lists of similarities and making parallels between lexical and grammatical forms.<sup>7</sup> In a book considered a classic, Sandfeld (1926/30) registers over one hundred Balkan Sprachbund morphosyntactic concordances (i.e., properties or features), some of which are general, while others apply to a more limited number of languages. Aware of the disadvantage of the mechanical

---

6. Balkan Romani has not been discussed by the previous authors. Recent research, however, shows that Balkan Romani exhibits many of the properties that characterize the Balkan Sprachbund (cf. Boretzky 1996; Matras and Bakker 1997).

7. The lists are copied and recopied without attention being paid to the time when they were collected and mistakes are perpetuated.

listing of possible correspondences,<sup>8</sup> many Balkanologists have subsequently sought to limit the set of Balkan linguistic features with which they work.

Using as criteria the typological character of a feature and its widespread occurrence, most authors have agreed on the existence of the following grammatical features: (1) prepositional cases; (2) postpositive definite articles; (3) dative/genitive merger; (4) location/direction merger; (5) vocative case marker; (6) pronominal clitic doubling of objects; (7) loss of the infinitive and its substitution by subjunctive clauses; (8) analytic expression of futurity with a “will” auxiliary; (9) analytic perfect with a “have”-auxiliary; (10) evidentials. Individual authors, have, however, concentrate on different arrays of features. Schaller (1975), making a distinction between primary and secondary Balkanisms, lists seven morphosyntactic phenomena among the latter: (a) merge of the dative and the genitive; (b) postpositive articles; (c) analytic comparison; (d) loss of the infinitive and its replacement by subordinate clauses; (e) use of the “will”-auxiliary in the construction of periphrastic future tenses; (f) doubling of objects (by “short” pronominal forms); (g) use of “short pronominal form” for the expression of possessiveness. Solta (1980) concentrates on six grammatical Balkanisms, which he classifies in three groups: (a) “morphological Balkanisms”, comprising the existence of the postpositive article and the merge of the genitive and the dative; (b) “syntactic Balkanisms” including the loss of the infinitive and the synthetic expression of futurity; (c) “special Balkanisms” (*sonstiges*) containing the vocative “as a living category” and the periphrastic comparison of adjectives. Gołąb (1984) singles out one negative similar or identical structural pattern – lack of the infinitive and its replacement by subjunctive clauses – and ten positive similar or identical structural patterns: (a) “will” future + subjunctive mood of a given verb; (b) future in the past, “derived from the future tense by the replacement of present tense markers by past tense ones” (Gołąb 1984: 6); (c) present optative-subjunctive mood, formed by a modal particle + the present tense of the verb; (d) imperfect optative-subjunctive mood, formed in the same manner as the present optative-subjunctive mood, but with the imperfect tense of the verb; (e) compound perfect, formed by the present tense of the “have” auxiliary + an indeclinable form of the perfect participle; (f) compound pluperfect – a derivative of the compound perfect, through the replacement of the present tense of the auxiliary by its imperfect tense; (g) future perfect – a derivative of the compound perfect, through the replacement of the present tense of the auxiliary by its future tense; (h) postpositive definite articles; (i) dative-possessive as a single morphosyntactic category. Lindstedt (2000) registers twelve grammatical Balkanisms, which fall into three groups: (1) argument-marking

---

8. Even Sandfeld was aware of the disadvantages of his methodology, and wrote that, “as concerns the non-lexical Balkan parallels, one must content oneself with simple attestation of the facts” (Sandfeld 1930: 101).

Balkanisms – (a) enclitic articles, (b) object reduplication, (c) prepositions instead of case endings, (d) dative/possessive merger, (e) goal/location merger, and (f) relativum generale; (2) Balkanisms pertaining to the verbal system – (g) Aux (+ Comp) + finite verb, (h) volo future, (i) past future as conditional, (j) habeo perfect, (k) evidentials; (3) analytic comparison (of adjectives and adverbs). Table 1 sums up the registering of the ten most widely discussed Balkan Sprachbund features by Sandfeld, Schaller, Solta, Gołąb and Linstedt.

**Table 1.** Morphosyntactic Balkan Sprachbund features discussed by individual authors (Sand = Sandfeld; Scha = Schaller; Solt = Solta; Gołą = Gołąb; Lind = Lindstedt)

	Sand	Scha	Solt	Gołą	Lind
postpositive articles	+	+	+	+	+
prepositional cases					+
Dative/Genitive merger	+	+	+		+
location/direction merger	+				+
clitic doubling of objects	+	+			+
Dative clitics in NP		+		+	
Vocative case markers			+		
loss of infinitives (and use of subjunctives)	+	+	+	+	+
“will” future	+	+	+	+	+
“will” future in the past + subjunctive				+	
“have” perfect				+	+
“have” pluperfect				+	
future perfect as conditional				+	+
evidentials					+
general relativizer					+
periphrastic comparison			+		

The listing of typical Balkanisms has reduced the study of Balkan morphosyntax to search for and identification of the items of the list in individual languages<sup>9</sup> and has led to tautological definitions of Balkan linguistics as a discipline that describes the features of the Balkan Sprachbund (cf. Desnickaja 1979). One should, however, bear in mind that the linguistic status of given features in individual languages varies. The Balkan languages differ rather dramatically in the extent to which they show the presence of a given feature and the divergences concerning the presence of a feature deserve as much emphasis as the general convergence or any of the specific convergences. Moreover, the “core languages” for the Sprachbund as a whole are not necessarily

9. Duridanov (1977: 21) writes: “Balkan linguistics aims above all to describe the substantial features of the Balkan Sprachbund.”

“core”-languages for each feature.<sup>10</sup> Yet, though not all the Balkan features are evenly distributed among the Balkan languages, the development of these features is parallel, and that is what makes the Balkan Sprachbund unique.

### 3.2 Dialects versus standard language

There have been claims that the source of the Balkan Sprachbund features was an ancient substrate – Thracian, Dacian and/or Illyrian (cf. Kopitar 1829; Miklosich 1861; Weigand 1928); a more recent sub- or superstrate – Latin (cf. Solta 1980); or Greek – the language “which has never ceased to transmit a civilisation superior to that of its neighbours” (cf. Sandfeld 1930). These claims are, however, unsubstantiated: First, most of the Balkan Sprachbund features developed during the Middle Ages<sup>11</sup> and after, and it is unrealistic to assume that any feature of any substratal language had been dormant for centuries, to become active long after the language itself had disappeared. Second, in Latin the Balkan Sprachbund features are hardly present, while in Greek they are post-classical innovations. As pointed out by Lindstedt (2000), these innovations are not simple transfers from a single substrate, nor from one of the languages themselves, but rather represent a shared drift. The Balkan Sprachbund is a typological phenomenon, which developed through convergence of dialects in a multilingual environment. As pointed by Joseph (2001), it is the dialects that represent more recent, even current, contact situations of bi- and multilingualism, which to a certain extent approximate those that led to the emergence of the particular Sprachbund effects observed in the modern standards. Accordingly, any typological analysis of Balkan morphosyntax should treat language as a diasystem of regional and social dialects and take into consideration the functioning of the “Balkan features” in individual dialects.

Preference for non-standard forms was implied already in Sandfeld’s monograph, where texts in Balkan dialects are used as the source. Birnbaum (1968) and Gallis (1969) argue that the problem of the geographical boundaries of the Balkanisms can

---

10. With respect to loss of the infinitive and its replacement by subjunctive constructions, Macedonian, Modern Greek, Bulgarian and the Torlak Serbian dialects are central, while Romanian, Albanian and standard Serbian or Croatian are peripheral (cf. Joseph 1983). The highly specific phonetic innovations combined with the preservation of grammatical archaisms in the Bulgarian Rodopi dialect (inflective nominal morphology, use of morphological infinitive) makes this dialect peripheral within the Balkan Sprachbund (Sobolev 2004). Alternatively, the east Serbian Timok dialect gains full membership as far as the same features are concerned. Many Albanian linguistic properties that are highly idiosyncratic (e.g., the grammaticalization of the future tense with “have” auxiliaries, the way modal or phrasal verb are related to its object, the position of pronominal clitics relative to imperative verbs) may indicate the special status of this language among the languages of the Balkans (cf. Soboljev 2004).

11. The oldest Albanian and Romanian documents are from the 16th century. The history of the Slavic languages and Greek are, however, well-documented and most of the Balkan Sprachbund features do not show up before the fifteenth century.

only be solved if dialect data are taken into consideration. Mladenov and Steinke (1978) point out that the use of dialect lead to discovery of new Balkanisms and make it possible to reconstruct the different developmental stages of a Balkanism and trace their diffusion. Mladenov and Steinke (1978) reason that the rules of standard grammars are congruent only with the rules in some dialects, but in the process of standardization become common and suggest inter-regionality, which in reality exists only in the written language. According to Friedman (1991), each Balkan standard language is realised in reality as a variety of different subsystems, based on the mother dialect of each speaker. Yet, Balkan Sprachbund morpho-syntactic parallels have predominately been made on the basis of the standard languages.

In an attempt to enrich Balkan studies with an areal component, comparing data from a number of different dialects, a group of St. Petersburg linguists have undertaken work on a "Small Dialectological Atlas of the Balkan Languages," aiming to register the spatial distribution of the fundamental grammatical and lexical features of representative Balkan dialects south of the Danube (Domosileckaja et al. 1998; Sobolev 2001). According to Sobolev (2004), the use of the geolinguistic method and the mapping of grammatical and lexical features makes it possible to single out the main linguistic sub-areas of the Balkan Peninsula, and leads to dialectological studies of convergent language communities. Each investigated feature should be understood as a link in the framework of a corresponding dialectal system that allows for comparison and mapping. Yet, though aiming at compiling a representative corpus of "balkanologically relevant data", the "Small Dialectological Atlas of the Balkan Languages" depicts a limited number of maximally divergent dialects, many of whose features are not present in the other dialects of the given language. Thus, it offers a garden variety of idiosyncrasies.

### 3.3 An integrated areal-typological approach

The features characterizing individual Balkan languages should not be neglected. But they should not be analyzed along with the features that appear in a number of these languages. Independent internal developments should actually counter-balance contact-induced convergence.

Already Belić (1936) pointed out that the common Balkan features in each language combine with different local features and thus play different systemic roles. More recently, Feuillet (1986) insisted that the Balkanisms have to be discussed in the framework of the phonological, morphosyntactic and lexical systems prevailing in the Balkan languages. Friedman (2004) argues that form, content, narrative structure, and social factors all have roles to play, and a typologically informed areal explanation of the Balkan Sprachbund phenomena appears to be the most adequate.

Rather than making parallel lists of convergent phenomena, and perpetuating a picture of uniformity, Balkan linguistics should display individual "Balkan Sprachbund" features in interaction with other features in the structure of the DP or the sentence of individual languages. The analysis of the grammatical sub-systems in

which Balkanisms operate should take into consideration sociolinguistic factors and dialectal variation at various levels of the grammar; in particular, when the standard language – taken as a diasystem of regional and social dialects and not as a normative standard – does not embody a typological feature of the area. In order to uncover the design of linguistic similarity and diversity and gain insight into morphosyntactic convergence and divergence, modern comparative synchronic and diachronic methods should be applied.<sup>12</sup> This would make possible the identification of morphosyntactic similarities, and will help find out what syntactic differences are linked to each other in the structure of the clause and the nominal phrase.

In the following sections, I examine some Balkan Sprachbund features that appear in neighbouring dialects of individual languages, but are absent from the other dialects of the languages to which these dialects belong; as well as some features that function in neighbouring Balkan languages of different families, but are absent from the other Balkan languages of the same family.<sup>13</sup> The comparison of the Balkan features in these dialects/languages with corresponding features in the languages to which the dialects belong, or with corresponding features in the other Balkan languages of the same family, will help display the working of convergent tendencies characterizing Balkan Sprachbund morphosyntax.

Section 4 deals with features of the nominal case systems, while in section 5 pronominal case and gender/number mergers are discussed. Section 6 focuses upon analogy of conditions for pronominal clitic doubling. Section 7 deals with DP possessive clitics relating to possessors denoting family relations, and section 8 with constructions with reflexive clitics and datives that are not part of the argument structure of the verb. Section 9 focuses on “have”-perfects, while in section 10 the spread of evidentials is analyzed. Section 11 draws some general conclusions.<sup>14</sup>

12. Since an areal explanation is by definition diachronic, a counter-argument must make use of diachronic facts.

13. Let us remind ourselves that there is no *a priori* difference between “language” and “dialect”. Each of the Chinese dialects has more speakers than any of the Balkan languages and intelligibility among the Chinese dialects is much smaller than intelligibility among the Balkan Slavic or the Balkan Romance languages. The Balkan Slavic languages are actually sections of a dialect continuum, while the Balkan Romance languages are often referred to as dialects of the same language.

14. The examples are taken from Tomic (2006). In the glosses of the examples, the following abbreviations are used: 1/2/3 = 1<sup>st</sup>/2<sup>nd</sup>/3<sup>rd</sup> person; Ac = accusative (case); Adm = admiring; Aor = aorist; Asp Mk = aspect marker; Aux = auxiliary; Cl = clitic; Da = dative (case); Evid mk = evidential marker; Eth = ethical dative; F = feminine; Impf = imperfect; Imps = impersonal; Inv = inverted; *l*-Part = participle ending on *-l* – for masculine singular (in Serbo-Croatian *-o*), *-la* – for feminine singular, *-lo* – for neuter singular, *-li* – for all persons plural (In Macedonian clauses with clitics, the behaviour of the *l*-participle is analogous to that of tensed verbs (cf. Tomić 1996, 1997)); M = masculine; N = neuter; Non-fin = non-finite Part = participle; Past – past (tense); Perf = perfective (aspect) perfect (tense); Pl = plural; Pres = present (tense); Prox = proximate; Ref = reflexive; Repet = repetitive Sg = singular; Subj = subjunctive; Su.Mk = subjunctive marker.



#### 4. Convergence of nominal case systems

##### 4.1 General oblique case markers

While Serbo-Croatian has distinct nominative, genitive, accusative, dative, instrumental and vocative case forms for all nominal types,<sup>15</sup> in Macedonian and Bulgarian, vocative case forms are the only productive oblique nominal forms. In the northern Macedonian (N-MACEDONIAN) and southeastern Serbian (SE-SERBIAN) dialects, however, accusative case forms are used as general oblique case markers.<sup>16</sup> As shown in (1b–c), when marking relationship other than those of direct object, nouns with accusative case markers are accompanied by prepositions.

- (1) N-MACEDONIAN
- a. *Vide gu sestru vu.*  
see.S.Aor 3S.F.Ac.Cl sister.F.Ac 3S.F.Da.Cl  
'I/you/(s)he saw his/her sister.'
  - b. *Dade vu na profesorku.*  
give.1S.Aor 3S.F.Da.Cl to professor.F.Ac  
'I gave it to the professor.'
  - c. *On otide kude profesorku.*  
he go.3S.Aor at professor.F.Ac  
'He went to the professor.'
- (2) SE-SERBIAN
- a. *Nesām videl sestru ti*  
not+ be.1S.Aux.Cl seen.M.S sister.Ac 2S.Da.Cl  
'I haven't seen your sister.'
  - b. *Dado ga na ovčara.*  
give.1S.Aor 3S.M.Ac.Cl to shepherd.Ac  
'I gave them to the shepherd.'
  - c. *Skaramo se sās sina mu.*  
quarrel.1Pl.Aor Ref.Cl with son.Ac 3S.M.Da.Cl  
'We quarrelled with his son.'

As pointed out above, the South Slavic languages actually represent a dialectal continuum. The loss of morphological case – a Balkan Sprachbund typological feature – intensifies as one moves southwards. The northern Macedonian and southern Serbian dialects are interim stages of this process. The fact that the general oblique case marker

15. The locative case is represented by prepositional phrases with nouns in the dative case, when expressing location, and by prepositional phrases with nouns in the accusative case, when expressing direction.

16. It is noteworthy that, while in Balkan Romance, Greek and Albanian dative case morphology is most persistent, in Balkan Slavic it is the accusative case markers that are last dispensed with.

functions in both groups of neighbouring dialects shows that the features developed through areal contact, which intensifies as one moves southwards.

#### 4.2 Loss of dative case markers

In Romanian, indirect objects carry dative case markers, whereas in the majority of Aromanian dialects they carry dative case markers and are introduced by the preposition *a* 'of/to':

(3) ROMANIAN

- a1 *Petru i-a dat o floare fetei.*  
 Petru 3S.Da.Cl-have.3S given a.F.S flower girl+the.F.S.Da  
 a2 *Petru i-a dat o floare unei fete.*  
 Petru 3S.Da.Cl-have.3S given a.F.S flower a.F.S.Da girl.F.S.Da

AROMANIAN

- b1 *Petre lji deade lilice a fatilj(e)i.*  
 Petre 3S.Da.Cl give.3S.Aor flower to girl+the.F.S.Da  
 b2 *Petre lji deade lilice a unei feată*  
 Petre 3S.Da.Cl give.3S.Aor flower to a.F.Da girl  
 'Petru/Petre gave a flower to the girl/a girl.'

Nevertheless, in the western dialects of Aromanian (W-AROMANIAN), spoken in areas where Macedonian – a Balkan Slavic language without morphological nominal case – is spoken by a great majority of speakers, dative case relationships are featured by prepositional phrases in which nouns in the nominative case are introduced by the preposition *al* 'to/of.M' if the noun is masculine, and *alji* 'to/of.F' if the noun is feminine:

(4) W-AROMANIAN

- a. *Petre lji deade cartea al fcior.*  
 Petre 3S.Da.Cl give.3S.Aor book+the.F.S to+the.M boy  
 'Petre gave the book to the boy.'  
 b. *Petre lji deade lilice alji feată.*  
 Petre 3S.Da.Cl give.3S.Aor flower to+the.F girl  
 'Petre gave a flower to the girl.'

In Megleno-Romanian (M-ROMANIAN) – a Balkan Romance language in very intensive contact with Macedonian – all nominal case markers have disappeared and indirect object relationships are expressed by nouns in the nominative case introduced by the prepositions *la* 'to':

(5) M-ROMANIAN

- a. *Ăi ded si mancă la calu meu.*  
 3S.Da.Cl give.1S.Aor Su.Mk eat.3S.Subj to horse+the.M.S my.M.S  
 'I gave food to my horse.' (lit. 'I gave my horse to eat.')  
 b. *Petre ăi deadi la feata flor.*  
 Peter 3S.Da.Cl give.3S.Aor to girl+the.F.S flour  
 'Peter gave the girl a flower.'

Indirect object case markers are also lost in the northern dialects of Modern Greek, which are also in intensive contact with Macedonian. In standard and southern Modern Greek (ST-GREEK) indirect objects can be expressed either in the genitive,<sup>17</sup> which can, but need not be clitic-doubled, or as proposition plus an accusative DP, which never clitic-doubles.

- (6) ST-GREEK
- |    |                 |                          |               |               |            |                |
|----|-----------------|--------------------------|---------------|---------------|------------|----------------|
| a. | ( <i>Tin</i> )  | <i>edhosa</i>            | <i>tis</i>    | <i>Marias</i> | <i>to</i>  | <i>vivlio.</i> |
|    | 3S.F.Ac.Cl      | give.1S.Aor              | the.F.S.Gen   | Maria.Gen     | the.N.S.Ac | book           |
| b. | ( <i>*Tin</i> ) | <i>edhosa</i>            | <i>sti(n)</i> | <i>Maria</i>  | <i>to</i>  | <i>vivlio.</i> |
|    | 3S.F.Ac.Cl      | give.1S.Aor              | to+the.F.S.Ac | Maria         | the.N.S.Ac | book           |
|    |                 | 'I gave Maria the book.' |               |               |            |                |

In the northern, Macedonian Greek dialects (N-GREEK), however, indirect object is expressed either by an Ac DP which always clitic-doubles or by a preposition plus an Ac DP, which can also clitic-double:

- (7) N-GREEK
- |    |                |                          |               |              |            |                |
|----|----------------|--------------------------|---------------|--------------|------------|----------------|
| a. | <i>Tin</i>     | <i>edhosa</i>            | <i>ti(n)</i>  | <i>Maria</i> | <i>to</i>  | <i>vivlio.</i> |
|    | 3S.F.Ac.Cl     | give.1S.Aor              | the.F.S.Ac    | Maria        | the.N.S.Ac | book           |
| b. | ( <i>Tin</i> ) | <i>edhosa</i>            | <i>sti(n)</i> | <i>Maria</i> | <i>to</i>  | <i>vivlio.</i> |
|    | 3S.F.Ac.Cl     | give.1S.Aor              | to+the.F.S.Ac | Maria        | the.N.S.Ac | book           |
|    |                | 'I gave Maria the book.' |               |              |            |                |

Loss of nominal morphological case markers is one of the distinct typological features of the Balkan Sprachbund. It did not, however, affect all cases equally. Indirect object case markers have survived in the nouns of most Balkan languages, though not in the Macedonian and Bulgarian nouns. In direct areal contact with Macedonian, the Balkan Romance and Modern Greek dialects have also lost their direct object nominal case markers.

It is interesting to note that the use of the Ac case in both direct and indirect objects in the northern Greek dialects, was reflected in the neighbouring south-eastern Macedonian dialects through accompaniment of direct objects by the preposition which throughout the Macedonian speaking area signals indirect objects. Compare the Standard Macedonian (ST-MACEDONIAN) sentences in (8) to their south-eastern Macedonian (SE-MACEDONIAN) counterparts:<sup>18</sup>

- (8) ST-MACEDONIAN
- |    |            |            |                   |                 |
|----|------------|------------|-------------------|-----------------|
| a1 | <i>Ī</i>   | <i>go</i>  | <i>dade</i>       | <i>na Jana.</i> |
|    | 3S.F.Da.Cl | 3S.N.Ac.Cl | give.3S.Perf.Past | to Jana         |

17. In Modern Greek there are no Dat case markers; Gen case markers are used instead.

18. Whereas in Standard Macedonian the clitics occur clause-initially, in south-eastern Macedonian they do not. See Section 6.

## SE-MACEDONIAN

- a2 *Dade i go na Jana.*  
 give.3S.Aor 3S.Da.Cl 3S.N.Ac.Cl to Jana  
 '(S)he gave it to Jana.'

## ST-MACEDONIAN

- b1 *Ja vide Jana.*  
 3S.F.Ac.Cl see.3S.Aor Jana

## SE-MACEDONIAN

- b2 *Vide ja na Jana.*  
 see.3S.Aor 3S.F.Ac.Cl to Jana  
 '(S)he saw Jana.'

## 5. Convergence of pronominal case systems

### 5.1 Loss of morphological datives

Macedonian has distinct dative and accusative third person full and clitic personal pronominal forms for both masculine and feminine gender. In the eastern Macedonian dialects, however, the dative form of the full pronominal form is lost – as it is in Bulgarian, with whose western dialect the eastern Macedonian dialects border. Thus, while in standard and western Macedonian pronouns in indirect object position appear in the dative case, in eastern Macedonian (E-MACEDONIAN) and Bulgarian, they appear in the accusative case and are introduced by the preposition *na* 'to'.<sup>19</sup>

## (9) ST-MACEDONIAN

- a. *Nejze i go dade.*  
 her.3S.F.Da 3S.F.Da.Cl 3S.N.Ac.Cl give.3S.Perf.Past

## E-MACEDONIAN

- b. *Na nejza i go dade.*  
 to her.3S.F.Ac 3S.F.Da.Cl 3S.N.Ac.Cl give.3S.Perf.Past

## BULGARIAN

- c. *Na nejza (i) go dade.*  
 to her.3S.F.Ac 3S.F.Da.Cl 3S.N.Ac.Cl give.3S.Perf.Past  
 'It is to her that (s)he gave it.'

### 5.2 Gender and number merger

In some southwestern Macedonian dialects, namely in the dialect of Bitola, as well as in the Macedonian spoken in Lerin (Florina) and Voden (Edesa) in northern Greece,

19. In Macedonian, as well as in Bulgarian, the (full) pronoun is used in emphatic (focused) phrases. The use of the doubling clitic becomes optional as one moves eastwards in the Macedonian-speaking area.

the dative clitic has lost distinctions for number and gender in favour of the masculine/neuter singular form *mu*. Thus, while in standard Macedonian, feminine, masculine and plural indirect objects are doubled by distinct pronominal clitics, in the south-western Macedonian dialects (SW-MACEDONIAN), all indirect objects are doubled by one and the same form:

(10) ST-MACEDONIAN

- a. *Mu go dade na brat ti.*  
 3S.M.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to brother 3S.Da.Cl  
 '(S)he gave it to your brother.'
- b. *Ĭ go dade na sestra ti.*  
 3S.F.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to sister 3S.Da.Cl  
 '(S)he gave it to your sister.'
- c. *Im go dade na decata*  
 3Pl.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to children+the  
 '(S)he gave it to the children.'

(11) SW-MACEDONIAN

- a. *Mu go dade na brat ti.*  
 3S.M.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to brother 3S.Da.Cl  
 '(S)he gave it to your brother.'
- b. *Mu go dade na sestra ti.*  
 3S.M.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to sister 3S.Da.Cl  
 '(S)he gave it to your sister.'
- c. *Mu go dade na decata*  
 3S.M.Da.Cl 3S.M/N.Ac.Cl give.3S.Perf.Past to children+the  
 '(S)he gave it to the children.'

In the area of Bitola, Lerin (Florina) and Voden (Edesa) Aromanian is intensively spoken, and it is in areal contact with this language, in which there are no gender distinctions in the Da pronominal clitics, whereas plural distinctions are minimal, that the Macedonian third person Da pronominal clitics have lost their gender and number distinctions.

## 6. Convergence of conditions for clitic doubling

### 6.1 Obligatory clitic doubling

Clitic doubling of indirect and direct objects is a prominent typological feature of the Balkan Sprachbund. In Bulgarian, Albanian, Modern Greek and Romanian clitic-doubling involves discourse factors. In Bulgarian both direct and indirect objects are clitic-doubled if topicalised, though in some cases in situ non-focused objects can also be clitic-doubled and there is a specificity, definiteness and possibly humanness effect. In Albanian, indirect object clitic-doubling is fully grammaticalized, i.e., a Da pronominal clitic accompanies obligatorily each indirect object, while direct object clitic-doubling is

contingent on occurrence in non-focused positions, as well as on specificity. In Modern Greek clitic doubling of both direct and indirect objects is contingent on the occurrence of direct objects in non-focused positions and specificity, though in colloquial speech direct object clitic-doubling tends to be fully grammaticalized, i.e., a doubling clitic tends to accompany every specific direct object. In Romanian, clitic-doubling of indirect objects is typically conditioned on specificity (though non-specific indirect objects can also be clitic-doubled), whereas direct object clitic-doubling is contingent on topicalization, specificity and humanness. In Macedonian, however, all definite direct objects and all specific indirect objects are clitic-doubled and only indefinite direct objects and non-specific indirect objects can appear without a doubling clitic:

(12) MACEDONIAN

- a. *Jana mu go dade pismoto na*  
 Jana 3S.M.Da.Cl 3S.N.Ac.Cl give.3S.Perf.Past letter+the.N.S to  
*deteto/ na edno dete.*  
 child+the.N.S to a.N child  
 'Jana gave the letter to the child/to a (specific) child.'
- b. *Jana (\*mu) (\*go) dade (edno) pismo na*  
 Jana 3S.M.Da.Cl 3S.N.Ac.Cl give.3S.Perf.Past a letter to  
*(edno) dete.*  
 a.N child  
 'Jana gave a letter to a(n unspecified) child.'

In Megleno-Romanian and Aromanian – the Balkan Romance languages in close areal contact with Macedonian – clitic doubling occurs under conditions analogous to those of clitic doubling in Macedonian. Compare (12) to (13) and (14):

(13) M-ROMANIAN

- a. *Petre i-u ari dat floarea la*  
 Petre 3S.Da.Cl-3S.F.Ac.Cl have.3S given.Part flower+the to  
*feata/ la una feată.*  
 girl+the.F.S to a.F girl  
 'Petre has given the flower to the girl/to a (specific) girl.'
- b. *Petre (\*i-u) ari dat (ună) floari la*  
 Petre 3S.Da.Cl-3S.F.Ac.Cl have.3S given.Part a.F flower to  
*(una) feată.*  
 a.F girl  
 'Petre has given a flower to a(n unspecified) girl.'

(14) AROMANIAN

- a. *Petre lĵ-u deade cartea a featiljei/*  
 Petre 3S.Da.Cl-3S.F.Ac.Cl give.3S.Past book+the.F.S to girl+the.Da  
*a unei feată.*  
 to a.F.Da girl  
 'Petre has given the book to the girl/to a (specific) girl.'

- b. *Petre (\*lj-u) deade una carte a*  
 Petre 3S.Da.Cl-3S.FAc.Cl give.3S.Past a.F book a  
*(unei) feată.*  
 a.F.Da girl  
 'Petre has given a book to a(n unspecified) girl.'

Thus, conditions for clitic-doubling in Aromanian and Megleno-Romanian differ from the conditions for clitic-doubling in the third and major Balkan Romance language, Romanian, where, as shown in (15), non-topicalized, non-human direct objects are, as a rule, not clitic-doubled even when they are definite:<sup>20</sup>

- (15) ROMANIAN  
*Am mâncat pește.*  
 3S.M.Ac.Cl-have.1S eaten fish+the.M.S  
 'I ate/have eaten the fish.'

## 6.2 Gradual change of conditions for clitic doubling

In the Balkan Slavic dialectal continuum, the conditions for clitic doubling change gradually. Whereas in Standard Serbian there is no clitic doubling whatsoever, in all south-Eastern Serbian dialects pronouns are clitic-doubled. The clitics in these dialects are syntactically and phonologically second-position clitics, as they are in Serbo-Croatian in general.

- (16) SE-SERBIAN  
 a. *Vikaše ni nas.*  
 call.3S.Imperf 1Pl.Ac.Cl us  
 '(S)he was calling us.'  
 b. *Mene me je zemnja pritisnula.*  
 me.Ac 1S.Ac.Cl is land pressed.F.S.I-Part  
 'I have to ply the land.' (lit. 'The land has pressed me.')

In the southeasternmost fringes of the southeastern Serbian dialect area (SE-SE-SERBIAN), the clitics are still second position clitics, but they can optionally double any definite direct object and any specific indirect object:

- (17) SE-SERBIAN  
 a. *Nesăm (ga) videl ovčara(toga).*  
 not+be.1S.Aux.Cl 3S.M.Cl.Ac.Cl seen.M.S.I-Part shepherd+the.M.S.Ac  
 'I haven't seen the shepherd.'

20. Examples of clitic-doubled Romanian objects are given in Tomić (2004).

- b. *Dala*                      *săm*                      (*mu*)                      *cveće*                      *na*                      *šefa/*                      *edno dete.*  
 given.F.S.I-Part    am.Aux.Cl    3S.M.Da.Cl    flowers    to    chief.Ac    a.N    child  
 'I gave (the) flowers to the chief/a child.'

In the northern Macedonian dialects the occurrence of clitic-doubling increases, to become obligatory for all definite direct objects and all specific indirect objects as one reaches the capital of Macedonia, Skopje.

In the dialect of Skopje and the dialects of western and west-central Macedonia (which are actually the base for standard Macedonian) all definite direct objects and all specific indirect objects are clitic-doubled. While syntactically pre-verbal, the clitics are in many cases phonologically proclitic, and thus can occur clause-initially.

- (18) ST-MACEDONIAN  
*Mu*                      *go*                      *dade*                      *pismoto*  
 3S.M.Da.Cl    3S.N.Ac.Cl    give.3S.Past    letter+the.N.S to  
*na deteto/*                      *na*                      *edno*                      *dete.*  
 child+the.N.S    to    a.N    child  
 '(S)he gave the letter to the child/to a (specific) child.'

In the Eastern Macedonian dialects, the conditions for clitic-doubling are the same as those in western and central Macedonia, but the phonological behaviour of the clitics changes as one moves Eastwards: the clitics become phonologically enclitic and obey the Tobler-Mussafia rule for non-occurrence in clause-initial position. The following examples are from the easternmost Macedonian dialects of Pirin Macedonia (PIR-MACEDONIAN) – an area within the boundaries of the Republic of Bulgaria:

- (19) PIR-MACEDONIAN  
 a. *Mene nikoj*                      *ne*                      *mi*                      *nosil*                      *kitka.*  
     me    nobody    not    1S.Da.Cl    bring.M.S.I-Part    flower  
     'Nobody has (ever) brought me flowers.'  
 b. *Ostail*                      *go*                      *na*                      *mira*                      *Petreto.*  
     leave.M.S.I-Part    3S.M.Da.Cl    to    peace.Ac    Peter+the.N.S  
     'He did not bother Peter (anymore).' (lit. 'He left Peter at peace.')

The southernmost Bulgarian dialects behave analogously. In the majority of Bulgarian dialects, however, clitic-doubling is contingent on specificity and focusing. Thus we have the following contrast:<sup>21</sup>

- (20) BULGARIAN  
 a. *Jana*    (*\*mu*)                      *pomaga*    *na*                      *DETETO/*                      *EDNO DETE.*  
     Jana    3S.N.Ac.Cl    help.3S    to    child+the.N.S    a.N.S    child  
     'Jana is helping the child/a child.'

21. The capitalized words are focused. As we see, focused constituents cannot be clitic-doubled.



- b. *JANA mu pomaga na deteto/ edno dete.*  
 Jana 3S.N.Ac.Cl help.3S to child+the.N.S a.N.S child  
 'It is Jana that is helping the child/a child.'

The gradual changes in the conditions for clitic-doubling in the Balkan Slavic dialectal continuum actually move along a vertical north-south axis and a horizontal east-west axis. As one moves from north to south, the restrictions on the type of object that can be clitic-doubled are relaxed, while the distance between the clitics and the verb is tightened. As one moves from east to west, the conditions for clitic-doubling are relaxed and the Tobler-Mussafia rule for non-occurrence of the clitics in clause-initial position gradually peters out. In the southwesternmost Balkan Slavic dialects – the dialects of Ohrid, the clitics are both syntactically and phonologically pre-verbal and clitic doubling is obligatory with all definite direct and all specific indirect objects. The same conditions for clitic-doubling obtain in the Aromanian and Megleno-Romanian dialects, which are in close areal contact with the southwestern Macedonian dialects, though the phonological behaviour of the clitics is somewhat different.

## 7. Convergence of types of hosts for possessive clitics

In Bulgarian, Macedonian, Modern Greek and Romanian, the pronominal clitics that function as doubled indirect objects can also function as possessive nominal modifiers. In Modern Greek, Bulgarian and Romanian, these possessive clitics can refer to any type of definite nouns, though in Romanian their usage is restricted to the formal style.<sup>22</sup> In Macedonian, however, only human nouns denoting family relationships can be modified by possessive clitics.<sup>23</sup> Thus, the Macedonian counterpart to the Bulgarian sentence (21a) is well-formed, whereas the Macedonian counterpart to (22a) is not:

- (21) BULGARIAN  
 a. *Dade podarāci na sina im.*  
 give.3S.Aor presents to son.Ac 3Pl.Da.Cl  
 '(S)he gave presents to their son.'

22. Many native speakers characterize them as "outdated".

23. Possessive clitics relating to human nouns denoting family relationships were also used in Serbo-Croatian literary writings of the 19th century, and are currently used in some dialects, especially if the nouns are in the Nom case:

## MACEDONIAN

- b. *Mu dade podaroci na sin im.*  
 3S.Da.Cl give.3S.Perf.Past presents to son 3Pl.Da.Cl  
 '(S)he gave presents to their son.'

## (22) BULGARIAN

- a. *Izplaši kokoškite ni.*  
 frighten.3S.Aor hens+the.F.Pl 1Pl.Da.Cl  
 '(S)he/it frightened our hens.'

## MACEDONIAN

- b. *\*Gi uplaši kokoškite ni.*  
 3Pl.Ac.Cl frighten.3S.Past hens+the.F.Pl 1Pl.Da.Cl

In Aromanian and Megleno-Romanian, personal pronominal clitics do not function as possessive nominal modifiers, though possessive non-clitic pronouns often occur post-nominally – in the position where the possessive clitics occur in Bulgarian and Macedonian. The first and second person singular possessive pronouns can have clitic versions:<sup>24</sup>

## (23) AROMANIAN/M-ROMANIAN

- sor-mea/ta*  
 sister-my.F.S.Cl/your.F.S.Cl  
 'my/your sister'

In Arli Balkan Romani, nouns denoting family relationships are also modified by possessive clitics related to lexical possessive modifiers:<sup>25</sup>

## (24) ARLI

- a. *Dikhlja me dajaja.*  
 see.1S.Past my.1S.Obl.Cl mother.Instr  
 'I saw my mother.'
- b. *So šuži stadik pletingja ti daj!*  
 what beautiful.F.S hat knit.3S.Past your.2S.F.Cl mother  
 'What a beautiful hat has your mother knitted for you!'

- 
- i. *%Došli su gospodin Petrović i kćer* SERBO-CROATIAN  
 come.Pl.Part be.3.pl mister Petrović and daughter  
*mu.*  
 3S.M.Da.Cl  
 'Mr Petrović and his daughter have come.'

24. For third person singular possessive relationships, forms related to the impersonal pronominal clitics are used:

- i. *tetă-sa* AROMANIAN/M-ROMANIAN  
 aunt-his/her.F.S.Cl  
 'his/her aunt'

25. Like the articles, these clitics are placed to the immediate left of the noun.

In Tosk Albanian (T-ALBANIAN), on the other hand, the same function is performed by agreement clitics, placed to the immediate left of the nouns to which they refer:

- (25) T-ALBANIAN
- a. *Erdhi nga e ëma.*  
 come.3S.Aor from F.S.Agr.Cl mother+the.F.S  
 ‘(S)he came from his/her mother’s.’
- b. *Erdhi nga i ati.*  
 come.3S.Aor from M.S.Agr.Cl father+the.M.S  
 ‘(S)he came from his/her father’s.’

Whereas, the Aromanian, Megleno-Romanian, Tosk Albanian and Arli Balkan Romani paradigms have only singular members and can refer only to singular possessors,<sup>26</sup> in the southwestern Macedonian dialects, which are in direct contact with Aromanian and Tosk Albanian, the paradigm of possessive clitics has both singular and plural forms and can refer to any noun denoting family relationship, whether in the singular or in the plural number. Accordingly, one might be led to conclude that the development of the Aromanian, Megleno-Romanian, Tosk Albanian and Arli Balkan Romani possessive clitics was instigated by Macedonian. Nevertheless, the fact that possessive clitics analogous to those functioning in Aromanian and Megleno-Romanian also appear in Romanian, and, more importantly, the fact that in all Balkan Romance languages there are postnominal possessive pronouns, indicates that, while the Arli and Albanian possessive clitics might have developed under the influence of Macedonian, the Balkan Romance possessive clitics are an internal development, that, along with the Greek personal pronominal clitics, actually influenced the development of the Macedonian and Bulgarian possessive clitics.

## 8. Semantic convergence of impersonal constructions with datives

The impersonal, passive and middle clauses with impersonal clitics (as well as the Albanian and Modern Greek verbs with medio-passive morphology) often include dative personal pronominal clitics and/or lexical dative phrases. According to Rivero (2004), there are two types of impersonal/passive constructions with datives (cf. Tomić 2004). The first construction, which has an anti-causative core, is found in all Balkan languages, along parallel lines:

- (26) MACEDONIAN
- a. *Mu se skrši prozorecot.*
- ALBANIAN
- b. *I- u thye dritarja.*

26. In Albanian, they can hardly be used with nouns other than *ëma* ‘mother’ and *ati* ‘father’.

## ROMANIAN

- c. *I se sparse ferestrea.*  
 3S.Da.Cl Imps.Cl broke.3S window+the

1. 'His window got broken.'
2. 'He involuntarily caused the window to break.'

The second construction is the desiderative construction, which occurs in the South Slavic languages and Albanian, but not in Romanian and Modern Greek, where structurally analogous constructions have a passive interpretation.<sup>27</sup>

## (27) MACEDONIAN

- a. *Ne mu se raboteša vo nivata.*  
 not 3S.F.Da.Cl Imps.Ac.Cl work.3S.Impf.Past in field+the.FS  
 'He didn't feel like working in the field.'

## ALBANIAN

- b. *I u punua këtu.*  
 3S.Da Imps.Cl work.3S.Impf here  
 '(S)he felt like working here.'

## ROMANIAN

- c. *I se lucra în fabrică.*  
 3S.Da.Cl Imps.Ac.Cl work.3S.Impf in factory  
 1. 'One/people worked in the factory on his/her behalf.'  
 2. 'One/people worked in his/her factory.'  
 Not: '(S)he felt like working in the factory.'

In Aromanian and Megleno-Romanian, however, the structural counterparts of the constructions in (27) have desiderative interpretations:

## (28) AROMANIAN

- a. *Al Petko nu-lj si lucrea.*  
 to Petko not-3S.Da.Cl Ac.Ref.Cl work.3S.Impf

27. The South Slavic and Albanian datives in passive *se*-constructions with direct objects may also have an "ordinary" passive reading, in which the dative is an indirect object. Thus, (i) and (ii) have two readings each:

- i. *Ni se peea pesni.* MACEDONIAN  
 1Pl.Da.Cl Ac.Ref.Cl sing.3Pl.Perf.Past songs  
 1. 'We felt like singing songs.'  
 2. 'Songs were being sung to us.'
- ii. *Na Petko ne mu se gradi kuća.*  
 to Petko not 3S.F.Da.Cl Ac.Ref.Cl build.3Sg/Impers house  
 1. 'Petko doesn't feel like building a house.'  
 2. 'No house is being built for Petko.'

## M-ROMANIAN

- b. *La Petku nu ăi ți sirbea.*  
 to Petko not 3S.Da.Cl Ac.Ref.Cl work.3S.Impf  
 'Petko didn't feel like working.' (lit. 'To Petko it didn't feel like working.')

In close areal contact with Macedonian, the Aromanian and Megleno-Romanian impersonal/passive constructions with datives have acquired the meaning of analogous South Slavic constructions.

## 9. Convergence of constructions for the expression of the perfect

While the Romance Balkan languages, Albanian and Modern Greek have "have"-perfects, the South Slavic languages, like the Slavic languages in general, as a rule, have "be"-perfects with inflecting I-participles. Nevertheless, in the western Macedonian dialects, as well as in the southernmost rim of the eastern Macedonian dialects and the Rodopi Bulgarian dialects, all of which are in direct contact with non-Slavic Balkan languages, there is a "have"-perfect.<sup>28</sup> Whereas in Balkan Romance, Modern Greek and Geg Albanian (G-ALBANIAN) the "have"-perfects are used (a) to express events that had begun prior to a reference point of time and last until that point, as well as (b) events that occurred at a definite point of time in the past,<sup>29</sup> in Tosk Albanian (T-ALBANIAN) they can only have the former usage.<sup>30</sup>

28. The "have"-perfect is also operative in standard Macedonian, which is based on the west-central Macedonian dialects. As a matter of fact, in standard Macedonian two verb subsystems intersect. Subsystem A has a (simple) past tense (with distinct paradigms for perfective and imperfective verbs, reflecting the paradigms of the aorist and the imperfect tenses, respectively), a "be"-perfect used to express resultativeness and experience as well as evidentiality, and a "be"-pluperfect. Subsystem B has a simple past tense (with distinct paradigms for perfective and imperfective verbs), a "be"-perfect used exclusively to express evidentiality, and "have" present and past perfects (cf. Tomić 1988).

29. Whereas in Romanian the perfect has almost ousted the use of the aorist and the imperfect, in Aromanian and Megleno-Romanian it is in free variation with these tenses:

- i. a. *Si dușea ăn căsăba ier.* M-ROMANIAN  
 Ac.Ref.Cl walk.3S.Imperf in town+the.FS yesterday  
 b. *Si ari dus ăn căsăba ier.*  
 Ac.Ref.Cl have.3S walk.Part in town+the.FS yesterday  
 '(S)he walked through town yesterday.'
- ii. a. *Vinj di Ohărda aseara.* AROMANIAN  
 come.1S.Aor from Ohrid yesterday  
 b. *Am vinită di Ohărda aseara.*  
 have.1S come.Part from Ohrid yesterday  
 'I came from Ohrid yesterday.'

30. Actions/events that occurred at a definite point of time in the past are most often expressed by the aorist, the use of which increases as one moves from north to south on the peninsula.

## (29) ROMANIAN

- a. *Au sosit toți musafirii.*  
 have.3Pl.Cl arrive.Part all.M.Pl guests+the.M.S  
 'All the guests (have) arrived.'

## GREEK

- b. *Dhen ton exo dhi.*  
 not 3S.M.Ac.Cl have.1S see.Act.Perf.Part  
 'I haven't seen him./I didn't see him.'

## G-ALBANIAN

- c. *Baba im e ka vizitu Tiranën.*  
 father 1S.Poss.Cl 3S.Ac.Cl have.3S visit.Non-Fin Tirana+the.F.S.Ac  
 'My father (has) visited even Tirana.'

## T-ALBANIAN

- d. *Kam erdhur të bisedojmë.*  
 have.1S come.Part Su.Mk talk.1Pl  
 'I have come to talk (to you)./\*I came to talk (to you).'

Since in the Slavic Balkan languages the “have” perfect appears only in areas in which it is in direct areal contact with neighbouring Balkan (non-Slavic) languages, we can safely conclude that it has developed through this contact. This conclusion has material support in the dialectal differentiation of the form and substance of the Balkan perfects. Thus, in the southwestern Macedonian dialects, which border with Tosk Albanian, the “have”-perfect exclusively denotes events that had began prior to a reference point of time and last until that point,<sup>31</sup> while the forms of the common Slavic “be” perfect with l-participles have acquired a new usage – they are exponents of evidentiality (cf. the next section).

## (30) SW-MACEDONIAN

- a. *Go nemam viden.*  
 3S.M.Ac.Cl not+have.1S see.Past.Part  
 'I haven't seen him/\*I didn't see him.'
- b. *Ja imam pročitano knjiga.*  
 3S.F.Ac.Cl have.1S read.Past.Part book+the.F.S.Prox  
 'I have read this book/\*I read the book.'

In the Ohrid-Struga southwestern dialects, which seem to be the center of the innovation, the “have”-auxiliary can take as complements not only past participles of lexical verbs, but also past participles of the “be” and “have” auxiliaries:

## (31) SW-MACEDONIAN

- a. *Imam imano vakov fustan.*  
 have.1S have.Past.Part such.M.S dress  
 'I have had a dress like this one.'

31. This is the usage of the “have”-perfect in standard Macedonian, but there it intersects with the “be”-perfect.

- b. *Imam bideno vo ovoj grad.*  
 have.1S be.Past.Part in this.M.S town  
 'I have been in this town.'

In the southernmost Macedonian dialects (spoken in what is now Greece), which are in direct contact with Greek, the Macedonian "have"-perfect expresses not only events that had began prior to a reference point of time and lasts until that point, but also events that occurred at a definite point of time in the past. As one moves eastwards, the use of the "have"-perfect decreases, while the use of the "be"-perfect increases (see Map 3). In some west-central dialects to the west of Prilep (PR-MACEDONIAN), which border with the eastern Macedonian dialects, where only "be"-perfects with *l*-participles exist, there are hybrid perfect constructions with "be"-auxiliaries which take as complements not *l*-participles, but rather non-inflecting -no passive participles, such as the ones that the "have"-perfects take:

- (32) PR-MACEDONIAN  
*Ne ti je Stojan umreno.*  
 not 2S.Eth.Da be.3S Stojan.M.S died.Past.Part  
 'Your Stojan has not died.'

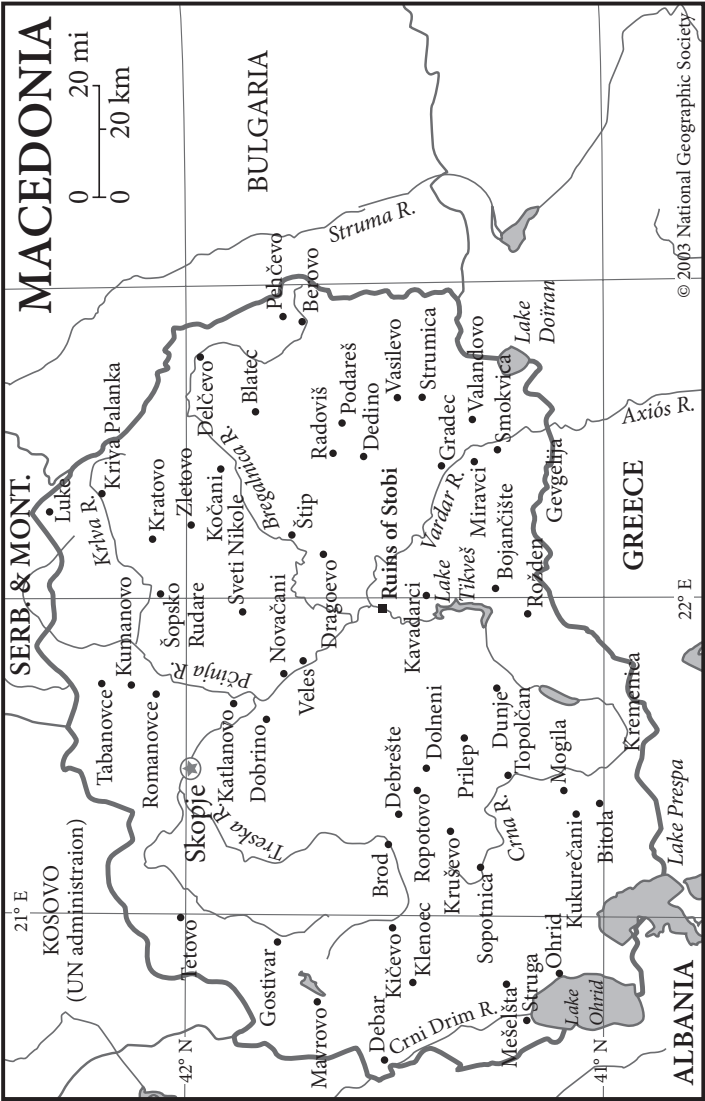
In the southwestern Macedonian dialects, there is a "be"-perfect with inflecting passive participles, which has formal counterparts with analogous cases in Aromanian and Megleno-Romanian:

- (33) SW-MACEDONIAN  
 a. *Dojdena<sup>32</sup> sum da si pozboruvame.*  
 come.F.S.Pass.Part be.1S.Aux.Cl Su.Mk Da.Ref.Cl talk.1Pl.Repetit
- AROMANIAN  
 b. *Escu vinită tă să zburăm.*  
 be.1S come.F.S.Pass.Part that.Mod Su.Mk talk.1Pl
- M-ROMANIAN  
 c. *Sam vinit si zăprikăjom.<sup>33</sup>*  
 be.1S come.M.S.Pass.Part Su.Mk talk.1Pl  
 'I have come to talk to you.'

32. In construction with inflecting passive participles the Macedonian "be"-auxiliary clitic *can* and preferably does follow the verb, while in construction with *l*-participles it has to precede it. Compare (33a) to (i):

- i. *Sum ti go dala.* MACEDONIAN  
 be.1S.Aux.Cl 2S.Da.Cl 3Sg.N.Acc.Cl given.F.S.I-Part  
 'I have (reportedly) given it to you.'

33. This verb, a borrowing from Macedonian, has a repetitive meaning. Note that in the Macedonian word *poprikažuvame* 'talk.1Pl.Repet', which was the source for this Megleno-Romanian word, the repetitive meaning is represented by the prefix *po-*. Thus, the Slavonic aspect has reflexes in Megleno-Romanian.



NATIONAL GEOGRAPHIC marcopolo  
**XPEDITIONS**  
www.nationalgeographic.com/xpeditions

Map 3. Macedonia and Adjacent Areas



Since the Macedonian constructions are most frequently used in areas of intensive areal contact with Aromanian and Megleno-Romanian, one might be led to conclude that they have developed under the influence of these two Balkan Romance languages. The more so since “be”-perfects with inflecting passive participles also occur in Romanian – a Balkan Romance language which is not in direct areal contact with any other language that has this type of perfect. As observed, the Macedonian clause (34a) has a verbatim Romanian counterpart:

## (34) SW-MACEDONIAN

- a. *Jadena sum, možam da izdržam.*  
eat.F.S.Pass.Part be.1S.Aux.Cl can.1S Su.Mk resist.1S

## ROMANIAN

- b. *Sunt măcată, pot să rezist.*  
be.1S.Aux eat.F.S.Pass.Part can.1S Su.Mk resist.1S  
‘I have eaten and can last.’

Yet, none of the Western Romance languages use “be” as an auxiliary in active constructions, and the occurrence of “be”-perfects in all the three Romance Balkan languages must be due to contact with Slavic.<sup>34</sup>

The Macedonian “be”-perfect extends to other western Macedonian dialect and is used in standard Macedonian, though only in constructions with intransitive verbs. In these dialects, the “be”-perfect actually supplements the “have”-perfect: Whereas the “have”-perfect expresses experience with intransitive verbs (cf. 35a) and experience or resultativeness with transitive verbs (cf. 35b and 35c, respectively), the “be”-perfect expresses resultativeness with intransitive verbs (cf. 35d):

## (35) MACEDONIAN

- a. *Imam dojdено овде повеќе pati.*  
have.1S come.Past.Part here more times  
‘I have come here more than once.’
- b. *Go nemam vidено од ponedelnikot.*  
3S.M.Ac.Cl not+have.1S see.Past.Part from Monday+the.M.S  
‘I haven’t seen him since last Monday.’
- c. *Ja imam pročitano knjigata.*  
her.Ac.Cl have.1S read.Past.Part book+the.F.S  
‘I have read the book.’

34. While Manoliu-Manea (1994: 270) offers conflicting hypotheses concerning Slavic influence versus native development, Gołąb (1984: 135), ascribes the use of the “be”-auxiliary in the Balkan Romance perfects to Slavic influence.

- d. *Dojdena*                      *sum*                      *ovde i*                      *porano*.  
 come.F.S.Pass.Part   be.1S.Aux.Cl   here   and   earlier  
 ‘I have been here before.’ (lit. ‘I have come here before.’)<sup>35</sup>

Developed through areal contact with neighbouring non-Slavic Balkan languages, the Macedonian “have”- and “be”-perfects have thus adapted their uses in order to coexist.<sup>36</sup>

## 10. Formal devices for expressing evidentiality

The existence of formal devices for the expression of evidentiality is often listed among the typical typological properties of the Balkan Sprachbund. The spread of morphologically marked evidentiality is, however, limited: it is characteristic only of Balkan Slavic, Albanian and Turkish. Moreover, the devices for the expression of evidentiality in Balkan Slavic differ from those in Albanian – both in form and substance.

### 10.1 Balkan Slavic evidentials

The Balkan Slavic evidentials are “be”-auxiliary clitics followed by l-participles, originally used as exponents of the perfect. In the western Macedonian dialects, where the “have”-perfect is fully operative, these constructions are exclusively used to express evidentiality. In Bulgarian and the eastern Macedonian dialects, the same forms are also used as “be”-perfects. In the Macedonian standard, two subsystems intersect – one in which the “be”-auxiliary clitics followed by l-participles are exclusively used as exponents of evidentiality, and another in which they are also used as exponents of the perfect (cf. Tomić 2004).

---

35. Since the form of the “be”-perfect is analogous to the form of passive constructions, which are used only with transitive verbs, the south-western “be”-perfects with transitive verbs are interpreted by speakers of the other Macedonian dialects as passives. Thus, (34a) would be interpreted as:

i. ‘I am eaten and can last.’

Being nonsensical, interpretations such as (i) are often sources for jokes.

36. Note that in the southwestern Macedonian dialects of Bitola and Kruševo, in which there are many native speakers of Aromanian, both the “have”- and the “be”-perfects have the meaning of the Aromanian “have” and “be”-perfects: The “have”-perfect is used to express not only actions/events that had begun some time in the past and last until the moment of speaking, but also actions/events that occurred at a definite point of time in the past, and the “be”-perfect is used not only with intransitive, but also with transitive verbs.

Typically the Macedonian and Bulgarian evidential forms impart the speaker's non-awareness of the event to which the verb refers. Consider the following sentences:

## (36) MACEDONIAN

- a. *Veli deka si me videl.*  
say.3S that be.2S 1S.Ac.Cl seen.M.S.I-Part  
'(S)he says that you have seen me.'
- b. *Se gleda deka si zboruvala so Ana.*  
Imps.Ac.Cl see.3S/Imps that be.2S speak.F.S.I-Part with Ana  
'One can see that you have spoken to Ana.'
- c. *Ovaa forma se upotrebuva za dejstva što se izvršile pred neкое drugo dejstvo vo minatoto.*  
this.F form Imps.Ac.Cl use.3S/Imps for actions that Imps.Ac.Cl  
accomplish.PL.I-Part before some other action in past+the.N.S  
'This form is used to express actions that have taken place before some other actions in the past.'

In (36a) we have a reported event; in (36b) a circumstantial deduction; in (36c) the evidential form expressed by the *I*-participle *izvršile* is used in a definition.

Further, the Macedonian and Bulgarian evidential forms can express surprise, admiration, amazement or irony. Very often, with appropriate intonation, the same sentence can have all of these meanings:

## (37) MACEDONIAN

- Ama ti si bil junak!*  
but you be.2S be.M.S.I.Part hero  
1. 'You are such a hero!' (admiration)  
2. '(I am surprised to see that) you are a hero.' (surprise)  
3. 'You a hero!' (irony)

The Macedonian evidential forms, though not their Bulgarian counterparts, are also commonly used in connected narratives:<sup>37</sup>

## (38) MACEDONIAN

- a. *Sum (ti) stanala rano i sum*  
be.1S 2S.Eth.Da.Cl get up.F.S.I-Part early and be.1S

37. Thus, the Macedonian sentence (38a) has no adequate equivalent in Bulgarian, while in the Bulgarian equivalent of (38b), given in (i), instead of the *I*-participle of the "be"-auxiliary, preceded by an impersonal clitic, we have the *I*-participle of the "have"-auxiliary:

i. *Imalo edin car i toj imal dvama sina.* BULGARIAN  
have.N.S.I-Part a king and he have.M.S.I-Part two.M sons  
'There was a king and he had two sons.'

- (ti) se dala na rabota.  
 2S.Eth.Da.Cl Imps.Ac.Cl give.F.S.I-Part on work  
 ‘Lo and behold, I got up early and got to task.’ (lit. ‘I got up early  
 and gave myself to work.’)
- b. Si bil eden car i si  
 Imps.Da.Cl been.M.S.I-Part a.N.S king and Imps.Da.Cl  
 imal dva sina.  
 have.M.S.I-Part two sons  
 ‘There was a king and he had two sons.’

## 10.2 Albanian evidentials

The Albanian evidentials – two sets of synthetic and four sets of analytic active and medio-passive evidential forms – developed through inverting the order of the auxiliary and the participle in its perfect forms. They are constructed by suffixing forms of the verb *kam* ‘have’ to the root of the lexical verb or (in the case of complex tenses) to the participle of the auxiliaries *kam* ‘have’ or *jam* ‘be’.

Typically, the Albanian evidential forms express admiration,<sup>38</sup> surprise or irony:

- (39) ALBANIAN  
 TI (me të vërtetë) flitka kinezçe.  
 you in Pl.Agr.Cl truth speak.3S.Adm Chinese  
 ‘You really speak Chinese!’

The Albanian evidential forms are also used to assert the occurrence of an action or the existence of a state:<sup>39</sup>

- (40) ALBANIAN  
 a. Ai (vërtetë) flitka kinezçe.  
 he truth speak.3S.Adm.Pres Chinese  
 ‘He does speak Chinese!’  
 b. Ai (vërtetë) e paska përdorur thikën.  
 he truth 3S.Ac.Cl have.3S.Pres.Adm used.Part knife+the.M.S.Ac  
 ‘He indeed used the knife!’

38. Albanian grammars actually label the evidential forms “admirative”.

39. It is noteworthy that, while in Balkan Slavic the evidential forms always have some sort of past reference, even when the apparent meaning is present, the Albanian evidential system has a true present. Mark the difference between the reading of (40a) and (40b). Note also that, while the Slavic evidential forms are used only in declarative sentences, the Albanian evidentials can be used in questions, as well (cf. Friedman 2004):

- i. Ku qenka mjeshtri? ALBANIAN  
 where be.3S.Adm boss+the.M.S  
 ‘Where is the boss?’

The Albanian evidentials are, however, never used in narratives.

By inference, the Albanian evidential forms are used to express disagreement, or uncertainty:

(41) ALBANIAN

- a. *Thonë se lakam teshë çdo ditë.*  
 say.3S that wash.1S.Pres.Adm clothes.S+the.FS every day  
 ‘They say that I wash/am washing clothes every day (but I don’t).’
- b. *(Më duket) sikur po afroheshkan retë.*  
 1S.Da.Cl seem.Imps as if Aff.Part approach.3Pl.Pres.Adm clouds  
 ‘(It seems to me) that the clouds are approaching.’

### 10.3 Turkish as the source of the Balkan evidentials

The formal and conceptual differences between the Balkan Slavic and Albanian evidentials preclude the possibility that Macedonian or Bulgarian are responsible for the development of the Albanian evidentials or, vice versa, that Albanian is responsible for the development of the Balkan Slavic evidentials. Moreover, there is no textual evidence that formally marked evidentiality existed on the Balkans prior to the time when features which are nowadays referred to as “Balkan Sprachbund features” began to develop. Since this development (a) followed the Turkish invasion and (b) the oldest (eighth century) Turkish texts already display an opposition between the *-di* (assertive) and *-miş* (evidential) pasts, much like that of modern Turkish (Tekin 1968: 192–193), while no evidentiality was encoded in Ancient Greek, Latin, or Old Church Slavonic, it can be postulated that the grammaticalization of evidentiality was due to contact with Turkish. The more so since, outside the Balkans, the grammaticalization of evidentials seems to take place predominantly in regions with some sort of Turkic contact, and dialects that are especially influenced by a Turkic language (e.g., Istanbul Judesmo, Ingilo Georgian or Tajik; cf. Friedman 1988).

Friedman (2004) argues that the grammatical encoding of evidentiality began in urban centers, where Turkish was widely spoken and had high prestige, and spread from there to the countryside.<sup>40</sup> He mentions three such centers: Elbasan in central Albania, Ohrid in south-western Macedonia and Preslav (later Tărnovo) in northeastern Bulgaria. Elbasan is actually in the center of the Albanian area contiguous with southwestern Macedonia, where the evolution of the Albanian evidentials from inverted perfects into classic admiring sets of paradigms is complete. Ohrid is the center of the area where the forms of the Slavic “be” perfect are exclusively used as exponents of evidentiality, while Tărnovo is in the area from

40. This prestige was associated with Ottoman restrictions on urban residency (cf. Akan 2000).

where the use of the *l*-participle of the Slavic “be” auxiliary as exclusive marker of evidentiality spread out.<sup>41</sup>

It is noteworthy, however, that the Macedonian dialect of Bitola – the most important administrative center of the Ottoman empire on the Balkans – has no morphological exponents of evidentiality. Since the Bitola Macedonian dialect is in intensive contact with Aromanian, the absence of morphological marking of evidentiality in this dialect could be explained through contact with Aromanian. As for Aromanian and Modern Greek, Friedman argues that the non-appearance of evidentials in these languages “may be connected with patterns of rural transhumancy for much of Aromanian, and competing urban prestige both for Greek as a language of religion and commerce, and, in some urban centers such as Bitola, resistance of Aromanian vis-à-vis Macedonian, which during the Ottoman period was more associated with the countryside” (Friedman 2004: 134). But the countryside around Bitola does not have evidentials, either! The absence of morphologically marked evidentiality from most Aromanian dialects can hardly be ascribed to resistance vis-à-vis Macedonian. If the lack of evidentials in Modern Greek is due to the prestige of Greek as the language of commerce and religion, the resistance to its development in Aromanian could plausibly be ascribed to the fact that most urban Aromanians considered Greek as their second language.

---

41. In Macedonian, the exponents of evidentiality interact with the “have”-perfect to render evidential perfects. For example:

- i. *Reče deka si me imala* MACEDONIAN  
 say.3S.Past that be.2S 1S.Ac.Cl have.F.S.*l*-Part  
*videno.*  
 see.Past.Part  
 ‘(S)he said that you had seen me.’

In Bulgarian, on the other hand, evidential perfects are expressed by constructions in which present tense forms of the “be”-auxiliary are followed by *l*-participles of “be” and *l*-participles of lexical verbs:

- ii. *Kazva če sme bili pātuvali* BULGARIAN  
 say.3S.Pres that be.1Pl be.Pl.*l*-Part travel.Pl.*l*-Part  
*mnogo časove kogato avtobusāt sprjal vāv*  
 many hours when bus+the.M.S stop.M.S.*l*-Part in  
*njakakav grad.*  
 some kind town  
 ‘He says that we had been travelling for many yours when the bus stopped in some town.’

Just as the use of the Macedonian “have”-perfect does not extend to the eastern Macedonian dialects, the use of the Bulgarian *l*-participle of “be” does not extend to the western Bulgarian dialects. Accordingly, in the dialectal continuum of the eastern Macedonian and western Bulgarian dialects, there is no evidential perfect.

#### 10.4 Areal influence of Macedonian and Albanian on Balkan Romance

While absent from most of the Aromanian dialects and practically absent from Romanian,<sup>42</sup> the evidentials are explicitly present in Megleno-Romanian. There are actually two sets of forms with evidential function in this language: the forms of the inverted perfect and the forms of the inverted past perfect. Whereas in the inverted perfect invariable past participles of lexical verbs are followed by present tense forms of the “have” auxiliary, in the inverted past perfect inverted perfect forms of the “have” auxiliary appear to the right of invariable past participles of lexical verbs.

Typically, the Megleno-Romanian inverted perfect renders reported or circumstantially deduced actions, while the inverted past perfect expresses non-witnessed events or actions that have begun at a point in the past and last until the moment of speaking:

(42) M-ROMANIAN

- a. *Zisi ca ȝ-li vutau dat.*  
 say.3S that 2S.Da.Cl+3Pl.F.Ac.Cl have.3S.Inv.Perf give.Part  
 ‘(S)he says that (s)he gave/has given them to you.’
- b. *Vutau vizut vrină feată ăn drumu.*  
 have.3S.Inv.Perf see.Part some girl in road  
 ‘They had reportedly seen some girl on the way.’

The Megleno-Romanian inverted perfect can also express surprise, admiration, amazement or irony and can be used in story-telling:

(43) M-ROMANIAN

- a. *Tu fostai ăncrilată!*  
 you be.3S.Inv.Perf clever  
 ‘You are clever!’ (admiration)  
 ‘(I am surprised to see that) you are so clever!’  
 ‘You clever!’ (irony)
- b. *Ăsh vutau ună mumă trei feati.*  
 Imps.Da.Cl have.3S.Inv.Perf a.F mother three girls  
 ‘(Once upon a time) a mother had three daughters ...’

The meaning of the Megleno-Romanian evidentials is analogous to that of their Macedonian counterparts and has most probably developed under the influence of Macedonian, with which Megleno-Romanian has been in direct areal contact. The structure of the forms of the Megleno-Romanian evidentials is, however, related to the structure of the Albanian evidential forms, which is difficult to be given an areal

42. Dimitriu (1979: 269–271) treats the Romanian “will/would” plus the ‘be’ infinitive *fi* plus present participle constructions as present presumptive and the “will”/“would” plus the ‘be’ infinitive *fi* plus past participle constructions as past presumptive constructions. Yet, evidential-like uses of similar modal constructions in Western Romance (cf. Squartini 2001) can be adduced as an argument against the independent status of these constructions.

explanation since the area where Megleno-Romanian is spoken is not contagious with any Albanian dialect. The developments might be parallel, but independent. After all, in all the Balkan languages the evidential forms developed from the perfect.

The development of the forms of the evidentials in the Farsheriote Aromanian dialect spoken in the village Gorna Belica in southwestern Macedonia, can, however, clearly be explained through areal contact with Albanian.<sup>43</sup> The evidential forms in this dialect are constructed by adding a suffix to a masculine plural participial base. Moreover, the suffix is borrowed from Albanian – it actually is the third person singular present tense form of the Albanian auxiliary *kam* ‘have’, *ka*. Examples:

(44) AROMANIAN

- a. *Abe nās ancupratsca cera!*  
Interj he buy.Part+Evid.Mk car+the.F.S  
‘(I am surprised to find out that) he did buy the car!’
- b. *Abe tini avusca mults pāradz!*  
Interj you have.Part+Evid.Mk much money  
‘(I am surprised to find out that) you do have so much money!’

These forms are for the most part used to express surprise. They are not used in questions, as in Albanian or in narratives as in Macedonian. Thus, the meaning of the Farsheriote Aromanian evidentials is restricted to the most prominent usage of the evidentials, which is characteristic for both Albanian and Macedonian.

## 11. Conclusions

Displaying the individual typological features of the Balkan languages in interaction with other features in the structure of the DP or the clause of individual languages and taking into consideration sociolinguistic factors and dialectal variation – in particular, when the standard language does not embody a typological feature present in other Balkan languages, one is able to identify structural similarities and establish links between structural differences. The examination of the Balkan Sprachbund typological features that appear in neighbouring dialects of individual languages, but are absent from the other dialects of the languages to which these dialects belong, and the comparison of these features with corresponding features in the languages to which these dialects belong or in other Balkan languages of the same family, has shed light upon a number of pertinent Balkan Sprachbund questions.

1. An examination of a typological feature against the background of the synchronic and diachronic systems of the languages which are in areal contact, may indicate an

---

43. The Macedonian village of Gorna Belica is on the Albanian border. The Farsherites actually emigrated to Gorna Belica from Central Albania at the beginning of the twentieth century.



initial instigator for the development of that feature, but the above analyses have shown that a scrutiny of the meaning and distribution of that feature may lead to the conclusion that the initial instigator is not always responsible for its spread. Thus, the Balkan Slavic possessive clitics developed under the joint influence of the Greek possessive clitics and the Balkan Romance postnominal possessive pronouns. The appearance of such clitics in Albanian and Arli Balkan Romani is, however, due to contact with Macedonian. Even when the emergence of a feature in a language or languages is (or may be) due to sociolinguistic factors, its spread does not have to be stimulated by such factors. The grammatical encoding of evidentiality in Albanian and Macedonian may have been instigated by Turkish and, as argued by Friedman (2004), may have begun in urban centers, where Turkish was widely spoken and had high prestige. But its spread in Macedonian was related to the “division of labour” with the “have perfect”, while its appearance in Megleno-Romanian and Farsheriote Aromanian was due to areal contact with Macedonian and Albanian.

2. The formal devices for the expression of a Balkan feature most often come from within the language, but the nature of the native material differs. In the case of the evidentials, in Bulgarian and Eastern Macedonian, the new meaning of evidentiality was added to the existing meaning of the perfect (“be”-clitics plus *l*-participles); in western Macedonian, in south-western Macedonian in particular, the shomeurized (left without a job through the appearance of the “have”-perfect) “be”-clitics plus *l*-participles were given a new employment; in Albanian and Megleno-Romanian, the elements of existing material (the “have”-auxiliary forms plus participles), employed as exponents of the perfect, were rearranged and then underwent some morphological changes. While the means for expressing evidentiality in all languages were provided by the perfect, the clitics expressing possessive relationships came from different sources: in Balkan Slavic they are equivalent to the dative clitic forms of the personal pronouns, in Balkan Romance and Arli Balkan Romani they are cliticized forms of possessive modifiers, whereas in Albanian the adjectival agreement clitics acquire an additional usage. The formal devices for the expression of a feature can also be borrowed from a language with which the given languages is in direct areal contact, as is the case with the evidential marker in Farsheriote Aromanian.

3. Direct areal contact can lead to meaning equivalence of typological features. Thus, clitic doubling in Macedonian, Aromanian and Megleno-Romanian occurs under same conditions, though the forms of the clitics and their phonological behaviour remain different. Also, Aromanian and Megleno-Romanian impersonal constructions with datives have the meaning of analogous Balkan Slavic and Albanian constructions, though their forms are equivalent to those of the Romanian impersonal constructions with datives.

4. Last but not least, the examination of the presence of given Balkan Sprachbund typological features in neighbouring dialects (of same or different languages) makes it possible to observe the gradual spread of that feature and draw conclusions about the focus of its operation. By examining the case systems in the south Serbian and northern

Macedonian dialects we see that they are impoverished when compared to the systems of the other Serbian dialects, but still exhibit morphonological case endings absent from the majority of the Macedonian dialects. The progressive loss of morphological case markers from north to south culminates in the area where southwestern Macedonian, western Aromanian and northern Tosk Albanian intersect. This is also the area where clitic doubling is most strongly grammaticalized and where evidentiality is marked by distinct markers. Thus, a conclusion might be drawn that the area where Macedonian, Aromanian and Albanian intersect is the “epicenter” of Balkanization.

Lindstedt (2000) argues that the Balkan languages in the middle of the prestige hierarchy – the Balkan Slavic languages, Balkan Romance languages and Albanian, show extraordinary mutual convergence owing to extraordinary mutual multilingualism, whereas Greek and Balkan Romani, which are at the top and bottom of such a hierarchy, respectively, are less convergent.<sup>44</sup> The results of the analysis of the Balkan Sprachbund features in neighbouring dialects of individual languages, and their comparison with the Balkan Sprachbund features in the standard languages to which these dialects relate or else in other Balkan languages of the same family, lends support to Lindstedt’s arguments. Yet, since there are local areal convergences which do not follow the general direction of Balkanization,<sup>45</sup> a relativization of this argumentation is in order.

## References

- Akan, E.B. 2000. Shadow genealogies: Memory and identity among urban muslims in Macedonia. PhD dissertation, The American University.
- Belić, A. 1936. La linguistique balkanique aux congrès internationaux des linguistes. *Revue Internationale des Etudes Balkaniques*. II-ème année. Tom I–II (3–4), 167–171. Beograd: Imprimerie Minerva.
- Birnbaum, H. 1968. Slavjanskije jazyki na Balkanah i ponjatije tak nazyvajemyh jazykovyh sojuzov (The Slavic languages on the Balkans and the notion of the so-called language unions). *Glossa* 2(1): 70–92.
- Boretzky, N. 1996. Arli. Materialien zu einen südbalkanischen Romani-Dialect. *Grazer Linguistische Studien* 46 (Herbst 1996): 1–30.
- Carter, F.W. 1977. *A historical geography of the Balkans*. London: Academic Press.

---

44. Since Greek has high prestige as the language of commerce and religion, its speakers do not endeavour to learn and speak the other languages spoken on the Balkans as much as the speakers of these languages endeavour to learn and speak Greek. Balkan Romani speakers, on the other hand, always speak the other Balkan languages in their environment, but the speakers of these languages do not even attempt to speak Balkan Romani.

45. For example, in the Macedonian and Aromanian in the south-western Macedonian city of Bitola, evidentiality is not morphologically expressed. The meaning of the “have”-perfect in this city was strongly influenced by Modern Greek, where there are no evidentials.

- Desnickaja, A. 1979. O sovremennoj teorii balkanističeskikh issledovanij. K izučeniju grammatičeskikh sootvetstvij jazykov balkanskogo areala (On the contemporary theory of Balkan studies. Towards a study on the grammatical correspondences in the languages of the Balkans). In *Problemy sintaksisa jazykov balkanskogo areala* (Syntactic Problems of the Languages of the Balkans), A. Desnickaja (ed.), 3–15. Leningrad: Nauka.
- Dimitriu, C. 1979. *Gramatica limbii române explicată* (An Explanatory Grammar of Romanian). Iași: Junimea.
- Domosileckaja, M.V., Plotnikova, A.A. & Sobolev, A. 1998. Malyj dialektologičeskij atlas balkanskih jazykov (A small dialectological atlas of the Balkan languages). *Slavjanskoje jazykoznanije. XII Meždunarodnyj sjezd slavistov. Doklady Rossijskoj delegacii*, 196–211. Moskva: Nauka.
- Duridanov, I. 1977. Zum Begriff des Sprachbundes. *Balkansko ezikoznanie* 20: 17–22. Sofija: Izdatelstvo na Bălgarskata akademija na naukite.
- Feuillet, J. 1986. La linguistique balkanique [Cahiers balkaniques 10]. Paris.
- Fine, J. 1991 [1983]. *The early medieval Balkans*. Ann Arbor MI: The University of Michigan Press.
- Friedman, V. 1988. The category of evidentiality in the Balkans and the Caucasus. In *American contributions to the Tenth International Congress of Slavists; Linguistics*, A. Schenker (ed.), 121–139. Columbus OH: Slavica.
- Friedman, V. 1991. Glagolskite kategorii, morfološkite paradigmi i opredeluvanjeto na izoglosite na makedonskata jazična teritorija (The verbal categories, the morphological paradigms and the determination of the isoglosses of the Macedonian linguistic territory). *Studia linguistica polono-jugoslavica. Akademija nauka i umjetnosti Bosne i Hercegovine. Posebna izdanja*. XCIII (26): 33–40. Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine.
- Friedman, V. 2004. The typology of evidentiality and areal linguistics. In *Balkan syntax and semantics* [Linguistik Aktuell 67], O. Mišeska Tomić (ed.), 101–134. Amsterdam: John Benjamins.
- Gallis, A. 1969. Die Balkanlinguistik – Probleme und Forschung. *Norsk Tidsskrift for Sprogvidenskap* 23: 37–58.
- Gołąb, Z. 1984. *The Arumanian dialect of Kruševo in SR Macedonia, SFR Yugoslavia*. Skopje: Macedonian Academy of Sciences and Arts.
- Hindrichs, U. (ed.) 1999. *Handbuch der Südeuropa-Linguistik*. Wiesbaden: Harrassowitz Verlag.
- Joseph, B. 1983. *The synchrony and diachrony of the Balkan infinitive – A study in areal, general and historical linguistics*. Cambridge: CUP.
- Joseph, B. 2001. Is Balkan comparative syntax possible? In *Comparative syntax of Balkan languages*, M.L. Rivero & A. Ralli (eds.), 17–43. Oxford: OUP.
- Kopitar, J. 1829. Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur* Band 46: 59–106. Wien.
- Lindstedt, J. 2000. Linguistic Balkanization: Contact-induced change by mutual reinforcement. In *Languages in contact*, D. Gilbers, J. Nerbonne & J. Schaeken (eds.), 231–246. Amsterdam: Rodopi.
- Manoliu-Manea, I.M. 1994. *Discourse and pragmatic constraints on grammatical choices: A grammar of surprises* [North Holland Linguistic Series 57]. Amsterdam: Elsevier.
- Matras, Y. & Bakker, P. 1997. *The typology and dialectology of Romani*. Amsterdam: John Benjamins.
- Miklosich, F. 1861. Die slavischen Elemente im Rumunischen. *Denkschriften der Wiener Akad. Phil.-hist. Cl.* Vol. 12.

- Mladenov, M. & Steinke, K. 1978. Die Ergebnisse der neueren bulgarischen Dialektforschung im Lichte der Balkanologie. *Zeitschrift für Balkanologie* 14: 68–82. Wiesbaden: Harrassowitz.
- Reiter, N. 1994. *Grundzüge der Balkanologie. Ein Schritt in die Eurolinguistik*. Berlin: Harrassowitz Verlag.
- Rivero, M.L. 2004. Datives and the non-active voice: Reflexive clitics in Balkan languages. In *Balkan syntax and semantics* [Linguistik Aktuell 67], O. Mišeska Tomić (ed.), 237–267. Amsterdam: John Benjamins.
- Sandfeld, K. 1926/1930. *Balkanfilologien*. University of Copenhagen, 1926. (Translated into French under the title *Linguistique balkanique: problèmes et résultats*. Paris: Champion, 1930).
- Schaller, H.W. 1975. *Die Balkansprachen. Eine Einführung in die Balkanphilologie*. Heidelberg.
- Sobolev, A. 2001. Das Sankt Petersburger Projekt eines dialektologischen Atlases der Balkansprachen. In *Grundfragen eines Südosteuropa-Sprachatlas. Geschichte. Problematik. Method. Pilotprojekt*, H. Schaller (ed.), 55–68. Marburg: Biblion.
- Sobolev, A. 2004. On the areal distribution of syntactic properties in the languages of the Balkans. In *Balkan Syntax and Semantics* [Linguistik Aktuell 67], O. Mišeska Tomić (ed.), 59–100. Amsterdam: John Benjamins.
- Solta, G.R. 1980. *Einführung in die Balkanlinguistik mit besonderer Berücksichtigung des Substrats und des Balkanlateinischen*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Squartini, M. 2001. The internal structure of evidentiality in Romance. *Studies in Language* 25: 297–334.
- Tekin, T. 1968. *A Grammar of Orkhon Turkic*. Bloomington IN: Indiana University.
- Tomić, O. Mišeska. 1988. Language contact, language system and language code. In *Yugoslav general linguistics*, M. Radovanović (ed.), 361–381. Amsterdam: John Benjamins.
- Tomić, O. Mišeska. 1996. The Balkan Slavic clausal clitics. *Natural Language and Linguistic Theory* 14: 811–872.
- Tomić, O. Mišeska 1997. Non-First as a default clitic position. *Journal of Slavic Linguistics* 5(2): 301–323.
- Tomić, O. Mišeska 2004. The Balkan Sprachbund properties: Introduction. In *Balkan Syntax and Semantics* [Linguistik Aktuell 67]. O. Mišeska Tomić (ed.), 1–55. Amsterdam: John Benjamins.
- Tomić, O. Mišeska 2006. *The Balkan Sprachbund morphosyntactic features*. Dordrecht: Springer.
- Trubetzkoy, N. 1928. Proposition 6. *Actes du premier congrès international des linguistes*. Leiden.
- Weigand, G. 1928. Texte zur vergleichenden Syntax der Balkansprachen. *Balkan Archiv* IV: 53–70.
- Weinreich, U. 1953. *Languages in contact*. The Hague: Mouton.



# Zhuang: A Tai language with some Sinitic characteristics

## Post-verbal ‘can’ in Zhuang, Cantonese, Vietnamese and Lao

Rint Sybesma

Leiden University Centre for Linguistics, Department of Chinese Studies\*

This paper deals with an areal feature shared by languages spoken in Indo-China and Southern China, namely, the presence of a modal element in a for these languages a-typical post-verbal position. In this paper we investigate the properties of this element in Cantonese, Lao, Vietnamese and Zhuang. One of the questions we focus on is the fact that the element in question can be translated with ‘can’, but the languages differ in that only in a subset of them, it can mean ‘can-ability’. It turns out that the interpretational properties of the element is tightly interrelated with both syntactic and interpretational properties of the resultative construction.

### 1. Introduction

Most languages in the focus area of this paper (roughly, Indo-China and southern China) feature an element, glossed and referred to here as “ACQ”, which performs

---

\* Many people have provided crucial help, which I acknowledge thankfully. First of all, I would like to thank Nick Enfield, for doing such inspiring work as well as for his untiring willingness to help with the Lao data. I am also grateful to Nigel Duffield, for answering numerous questions about Vietnamese. LAN Qingyuan provided invaluable help in sorting out the Zhuang data; without his assistance, I would not have been able to write this paper. I also thank QIN Xiaohang for answering questions about Zhuang. In sorting out the Cantonese data, I relied on Lisa Cheng, Joanna Sio, Sze-Wing TANG and Leo Wong; thanks! I would like to stress that none of these people can be held responsible for what I did with the data they provided; see footnote 1.

Nick Enfield, LAN Qingyuan, Pieter Muysken, Leo Wong and You Rujie took the time and trouble to read through the first draft of this paper and to give me comments, which led to important improvements. Thank you all!

Lisa Cheng and Pieter Muysken deserve special mention, as they have contributed to the development of this paper in many different ways. I thank them for it.

Some of the material reported on here was presented at the Chinese University of Hong Kong in May 2005. Questions and comments from individuals in the audience have also led to several improvements, as did the comments Stephen Matthews made afterwards.

numerous seemingly unrelated functions in each individual language: besides functioning as a lexical verb meaning ‘acquire’, for instance, it operates as a modal element meaning ‘can’. ACQ is a truly areal feature, because virtually all languages in the area have it, despite the fact that they belong to several different language families (Enfield 2003). While the similarities among the languages with respect to the ACQ-phenomenon are startling, we concentrate on where they differ. In particular, we focus on how two Tai languages, Zhuang and Lao, differ in their use of ACQ as a modal element in post-verbal position. Our main problem can be stated as follows: Why is it the case, as (1) illustrates, that in Zhuang, in (1a), post-verbal ACQ cannot express ability (wherever it is in the sentence, as the curly brackets indicate), while it can have this reading in Lao in (1b)?<sup>1</sup>

- (1) a. \**De gangj {ndaej} vah Yeznanz {ndaej}*.  
       3s speak ACQ language Vietnam ACQ  
       INTENDED: ‘S/he is able to speak Vietnamese.’
- b. *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>*. (cf. Enf 102 (95))  
       3s speak language Lao ACQ  
       ‘S/he is able to speak Lao.’

To get a proper perspective on this question, we do not restrict our attention to post-verbal modal ACQ in these two languages. We not only include two more languages

---

This paper was planned during the time that I was involved in Pieter Muysken’s Spinoza-project, “Lexicon and syntax”. The actual research was conducted in the context of my own “Vernieuwingsimpuls”-project on syntactic variation in southern China, co-funded by the Dutch Organization for Scientific Research NWO, Universiteit Leiden (main sponsors) and the International Institute for Asian Studies, IIAS.

I dedicate this article to the memory of Rudolf de Rijk, a good man and a great scholar.

1. A note on the data. In taking sentences from published sources I have generally taken the entire example, that is, including the glosses and the translation (which I sometimes had to retranslate into English). However, I have often silently changed things in the glosses or the translations, always for the sake of consistency and ease of explication (e.g., I have consistently glossed the ACQ elements as “ACQ”). Here and there I made other changes as well. Although I have made these changes in good faith, and generally checked with native speakers or specialists, I may have distorted the data. As a consequence, when quoting examples from the present article, one must always check it against the original source if there is one.

Some frequent sources are abbreviated as follows: Duff: Duffield 2003; Enf: Enfield 2003; Qin: Qin 1995; Z&Q: Zhang and Qin 1993; numbers in brackets refer to examples, non-bracketed numbers refer to pages.

Lao examples are all due to Enfield; the transcription used is his. For Cantonese I use the system designed by the Linguistic Society of Hong Kong, even if the written source uses a different system, and Mandarin is transcribed using standard pinyin; tones have been added if the original source did not represent them. Vietnamese and Zhuang are given in their official orthographies.

from the area, Cantonese and Vietnamese, both genetically unrelated to the other three languages (they are Sinitic and Mon-Khmer respectively), but we include many more uses of ACQ as well. By investigating more languages and structures, we will get a better picture of possible correlations between different syntactic patterns.

The area, the languages we investigate and the basic properties of ACQ are briefly introduced in §2. A detailed overview (with some preliminary analyses) of the ACQ facts in the four languages follows in §3. §4 is devoted to an account of the variation we find with respect to ACQ as a post-verbal modal element and §5 presents some general conclusions.

## 2. The area, the phenomenon and the languages

### 2.1 The area

The ACQ area is large. Languages with ACQ are spoken in a region which comprises the Chinese speaking parts of China as well as most of mainland Southeast Asia. In this paper, we deal with languages spoken in the central part of this area, namely the provinces of Guangdong, Guangxi and Yunnan in China, and Laos and Vietnam on the other side of the international border. In this central area we find languages of different families, such as Sinitic, Tibeto-Burman, Tai and Mon-Khmer. Typologically, the languages spoken here have much in common: most (not all!) have tones, they tend to have monosyllabic morphemes and generally qualify as “isolating”, they make use of classifiers, and most (not all!) are SVO. The languages we discuss in this paper, Lao and Zhuang (Tai), Cantonese (Sinitic) and Vietnamese (Mon-Khmer) do share these typological properties to a large extent.<sup>2</sup>

### 2.2 The phenomenon

The element ACQ can function in various ways:

- as a lexical verb meaning ‘acquire, come to have’ (§3.1);
- as a lexical verb meaning ‘be okay’ (§3.4);
- as a pre-verbal modal, with mostly deontic meanings (§3.2);
- as a post-verbal modal meaning ‘can’ (§3.4);
- as an element comparable to *-able* in English *doable* (§3.3.2.);
- as an aspectual-like element directly preceding the main verb, expressing that certain conditions have to be met, before the event denoted by the main verb can take place (§3.2);

---

2. For an excellent overview of the language situation in the area, see Enfield (2003; 2004).



- as an aspectual element in post-verbal position, marking achievement (§3.3.3); and
- as a linker between the main verb and (what look like) adverbial expressions (§3.3.1).

Not only does one element perform these functions (and more!), it does so in many different, genetically unrelated languages: the ACQ-phenomenon is a truly areal phenomenon. The fact that there is one element which performs most if not all of these functions in virtually every individual language in the area, regardless of genetic affiliation and typological characteristics, is already fascinating in itself. What makes it even more fascinating (as Enfield 2003 rightly stresses) is that, although the form of the element looks similar in some cases – e.g., *ndaej* in Zhuang, *daj*<sup>4</sup> in Lao, *dak*<sup>1</sup> in Cantonese, *đượ* in Vietnamese – suggesting wholesale borrowing (but see Enfield 2001), in others, like Khmer, we find the clearly unrelated *baan*, performing the very same array of functions (Enfield 2003; Matisoff 1991). ACQ is simply everywhere.

Another indication of the all-pervasiveness (or intrusiveness) of the phenomenon is that we find ACQ as a modal element ('can') in post-verbal position, despite the fact that in the languages in question, modal verbs typically precede the main verb, as the primed examples in (2) illustrate.

(2) ZHUANG

- a. *De benz ndaej hwnj go faex neix.* (cf. Qin 267 (10))  
 3s climb ACQ up CL tree that  
 'S/he can climb up that tree.'

- a'. *Gou siengj youq neix ninz haj haemh.* (Qin 3 (3))  
 1s want at here sleep 5 evening  
 'I would like to sleep here for five nights.'

LAO

- b. *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>.* (= (1b))  
 3s speak language Lao ACQ  
 'S/he can speak Lao.' (one of the possible readings)

- b'. *laaw<sup>2</sup> jaak<sup>5</sup> khaa<sup>5</sup> kaj<sup>1</sup>.* (Enfield to app. (9))  
 3s want kill chicken  
 'S/he wants to kill a chicken.'

CANTONESE

- c. *keoi<sup>5</sup> lo<sup>2</sup>-dak<sup>1</sup>-hei<sup>2</sup> li<sup>1</sup>-seung<sup>1</sup> syu<sup>1</sup>.* (Cheng and Sybesma 2004 (1b))  
 3s take-ACQ-up this-box book  
 'S/he can lift this box of books (can get them up).'

- c'. *ngo<sup>5</sup> wui<sup>5</sup> heoi<sup>3</sup>.*  
 1s will go  
 'I will go.'

VIETNAMESE

- d. *Tôi lái xe đượ.* (Duff (39a))  
 1s drive car ACQ  
 'I can drive.'

- d. *Chét! Tòi phải ra bưu-diện trước năm giờ.* (Nguyễn 1997, 160)  
 die 1s must go.out post office before five o'clock  
 'Whoops! I must go to the post office before 5 o'clock.'

What this implies is that the development of ACQ in the individual languages seems to have gone against certain basic grammatical properties of these languages. One of the goals of the present paper is to try to understand this particular aspect of the ACQ-phenomenon better. How does the post-verbal modal fit into the grammatical system of the different languages?

It is not clear in which language the phenomenon originated; it may have been a Tai language, it may have been a variety of Chinese (Simpson 2001; Enfield 2001; 2003). What seems uncontroversial, however, is that the Tai are responsible for its spread all through Southeast Asia. Edmondson and Solnit (1997, 1, 12) state that "[n]early all historians agree that the Tai *Urheimat* lies in southern China", more particularly, according to Chinese historians, in Lianshan county in Guangdong province. From there the Tai migrated to the south and to the west as far as Assam and the surrounding areas in India, taking ACQ with them and leaving it behind wherever they went. I refer the reader to Enfield (2001; 2003), Matisoff (1991) and Simpson (2001) for discussion and details regarding origin, spread and distribution of ACQ.

## 2.3 The languages

In this paper we concentrate on particular differences between Zhuang and Lao, two Tai languages. The other two languages we study are Vietnamese (Mon-Khmer) and Cantonese (Sinitic; Mandarin, the now dominant variety of Sinitic, is also mentioned now and then). Cantonese is included because the Zhuang speaking area is part of the modern state of China. For centuries it has been dominated, both culturally and politically, by speakers of Cantonese and other Sinitic languages. It is possible that language contact is one of the reasons for the differences between Zhuang and Lao. The speakers of Zhuang are separated from the speakers of Lao by the political border between China and Laos, which has marked the limits of Chinese influence in the region for many centuries.<sup>3</sup> Vietnamese has been included in this study because it is a Mon-Khmer language, genetically unrelated to Sinitic and Tai (although it has been deeply influenced by both families).

### 2.3.1 *Zhuang*

The least known language of the four included in this study is Zhuang.<sup>4</sup> Zhuang is a Tai language spoken in a continuous area comprising, roughly, the western half of the

3. Southeast Asia is often seen as divided into an "Indosphere" and a "Sinosphere" (see Enfield 2004, 184–186).

4. "Zhuang" is the Chinese name for the language. In Zhuang, it is called "Cuengh" (or "vah Cuengh" 'Zhuang language'). As much as I would like to use "Cuengh", I use "Zhuang" because it is the term most commonly used in international publications.

Chinese province of Guangxi, Wenshan county in the neighboring province of Yunnan (to the west), and bordering areas in the provinces of Hunan and Guizhou (to the north). There is one discontinuous patch: Lianshan county and vicinity in Guangdong province (East), the likely place of origin of Zhuang (possibly of all Tai languages) (Wei and Qin 1980, 1; Zhang and Wei 1997, 78; Edmondson and Solnit 1997). The Zhuang area is mountainous and rich in water, and over the past millennia, it has been peopled by many different ethnic groups. According to Zhang and Wei (1997, 78), archaeological evidence shows bronze techniques and rice cultivation were at high levels of development here, when the (Chinese) Qin emperor sent half a million soldiers into the area in the 3rd century B.C. (Yue-Hashimoto 1991; Chappell 2001a, 7; see also Li 2002).

It is not obvious that what is called “Zhuang” is really one language. All sources on the language agree that there are two major dialect areas, “north” and “south”, each consisting of several subvarieties, seven in the north, five in the south (e.g., Zhang and Wei 1997; Wei and Qin 1980). Ramsey (1989, 236) notes that “[t]he speech of these two dialect areas is divergent enough to be considered two separate languages, even by the relaxed criteria of mutual intelligibility used in China.” SIL’s *Ethnologue* ([www.ethnologue.com](http://www.ethnologue.com)) has separate entries for “Northern Zhuang” and “Southern Zhuang.” Of course, miss-classifications happen; Buyi, for instance, is officially recognized as a separate language while, “[b]y Chinese standards, the Buyi language could be called a dialect of Northern Zhuang” (Ramsey 1989, 243; confirmed by Lan Qingyuan, p.c.).

Because Northern Zhuang is more widely spoken (two-thirds of the Zhuang speakers speak a northern variety; Zhang and Wei 1997, 79) and because the northern varieties are more unified, the variety of the town of Wuming in the north has been picked as the basis for the official standard; “Standard Zhuang” means “Northern Zhuang” (Zhang 1987, 268; Ramsey 1989, 236). Wuming Zhuang is also the variety that the official spelling is based on. Wuming Zhuang may officially be the standard language, but there is no incentive for people from outside Wuming to learn it. This not only has the consequence that literacy in Zhuang is very low, but also that Zhuang speaking people from different dialect areas communicate with each other in Chinese (either South-Western Mandarin, which they call “Guanhua”, or Cantonese, which they refer to as “Baihua”).

The Zhuang nationality has more than 15.5 million people (census of 1990; Huang 2000, 42), but the number of Zhuang speakers is hard to assess. Figures in Huang (2000) are not always consistent, and different sources contradict each other. Whatever the exact figures, there is no doubt that the vast majority of Zhuang people who do speak Zhuang is bilingual, some variety of Chinese being the other language. Monolingual speakers (most likely all older than 50) may still be found in rural areas in the western hills.

Chinese languages are omnipresent; the Zhuang language area may be by and large continuous, but the Zhuang people often do not constitute the majority. And even if they do, Chinese is prominent. In Wuming, for instance, where the Zhuang make up

more than 80% of the population and the Han Chinese “are able to speak Zhuang for the purposes of communication . . . since [it] is the center of political, economic, and cultural activities, Wuming Mandarin is the main means of communication” (Huang 1997, 58).

Many centuries of contact with the Chinese have left their mark on the Zhuang language.<sup>5</sup> I have only found systematic accounts of Chinese influence on the lexicon, but phonology and syntax have also been influenced (Dai 1992). Until the middle of the 20th century, Chinese loanwords were almost exclusively Cantonese (but see Qin 2004, 13–14). Since then, South-Western Mandarin has taken over the role of main supplier. Wang (1962, 257) reports that folk stories and folk songs from before the Communist take-over in 1949 contained about 30% Chinese loanwords. According to Wang, the percentage of loans has only increased since 1949; he estimates the percentage of loanwords in everyday conversation at his time of writing, 1962, “especially when the conversations involve modern life and work”, as more than 40%. In heavily political-administrative texts he counted more than 80% Chinese loanwords (Wang 1962, 257). Qin (2004, 10) confirms this picture for contemporary Zhuang.

### 2.3.2 *Cantonese, Lao and Vietnamese*

The other three languages we investigate are Cantonese, Lao and Vietnamese.

Cantonese, or Yue, is a group of Sinitic languages, spoken in the Chinese province of Guangdong, in Hong Kong, as well as in the eastern parts of the Zhuang province of Guangxi to its west.<sup>6</sup> It is also spoken in many other places around the world. Within China, it is spoken by about 60 million people (5% of the population; Ramsey 1989, Chappell 2001a). *Ethnologue* estimates the number of Yue speakers outside China at 20 million. The variety spoken in the cities of Guangzhou (“Canton”) and Hong Kong is taken to be Standard Cantonese. Due to the language policy of the central government, promoting Mandarin, more and more Cantonese speakers are bilingual (Chen 1999). Among the “Chinese dialects”, Cantonese counts as “strong”: not only does Mandarin not push Cantonese out of the home or the work place, speakers from other areas in China (such as Peking, Shanghai) learn Cantonese (Chen 1999). Although Cantonese is the only variety of Chinese (besides Mandarin) with what could be called a written tradition (Bauer 1988; Chen 1999), literacy in Cantonese is low and a typical situation of diglossia persists: virtually all writing is done in Mandarin, the written standard.

Cantonese/Yue is generally assumed to have a Tai substrate (Yue-Hashimoto 1991, 296; see also LaPolla 2001, 233; Chappell 2001b, 335), which would be reflected in

---

5. The local Chinese varieties have also been influenced by Zhuang (Huang 1997).

6. Normally, the term “Yue” is used for the language variety as a whole, whereas “Cantonese” is reserved for the standard. I use the latter for both.

syntax, morphology and lexicon.<sup>7</sup> It is important to keep in mind that similarities between Zhuang and Cantonese are not necessarily due to sinification of Zhuang but may as well be due to Taiization of Cantonese.<sup>8</sup>

While Zhuang and Cantonese are regional languages in China, Lao and Vietnamese are national languages. Lao is the national language of Laos, with more than 4 million speakers, but for a national language it is not very unified, neither in grammar nor in pronunciation (Enfield 2000; Ivarsson 2000). Lao is a Tai language and it is very close, and similar, to Thai in many respects. Enfield (2003, 62) speaks of a “high degree of convergence in grammar and lexicon” between the two languages and Ivarsson (2000, 69) mentions “minor variations with respect to tones and the pronunciation of specific vocals and combinations of vocals.” Especially in recent years, influence of Thai has been strong, through television, radio and popular culture (Enfield 2000; Ivarsson 2000). No Chinese influence is mentioned by anyone.

Vietnamese is a Mon-Khmer language, though it differs significantly from the other languages in this family in being tonal. As the national language of Vietnam, it is the first language of the majority (close to 90%; *Ethnologue*) of the population of that country (of more than 80 million) and the second of the rest; it is also spoken by communities of Vietnamese people all over the world. Vietnamese has been profoundly influenced by Chinese (Luong 1994; Nguyễn 1997). What is now Vietnam was ruled over by the Chinese between 111 B.C. and 939 A.D. Even after that date, until the 20th Century, Classical Chinese had a status comparable to that of Latin in Europe in the Middle Ages: it was the main literary language, used in higher education, etc. In the 20th century, the fact that both China and Vietnam had communist regimes assured the inflow of Chinese loanwords into the Vietnamese language. According to Luong (1994, 12), the Vietnamese lexicon consists for 50–70% of Chinese loanwords, although, according to Nguyễn (1997, 37), the loanwords from the first millenium CE have been “thoroughly Vietnamized” and are “hardly recognizable” as Chinese. Not only the lexicon has been influenced, though; other aspects of the language have been affected considerably as well. Enfield states that Tai influence on the Vietnamese language has also been argued to be significant (cf. Enfield 2003, 66, fn. 9).

### 3. The data

This section presents an overview of the different uses of the ACQ elements in Zhuang, Lao, Vietnamese and Cantonese. Although it is not exhaustive, the overview will be

---

7. Other southern varieties of Sinitic, such as Min, spoken in Fujian province north of Guangdong, have been shown to have an Austroasiatic substrate (Norman 1988; 1991). Austroasiatic is the family that Mon-Khmer, including Vietnamese, belongs to.

8. In a series of papers in the 1970s, Mantaro Hashimoto proposed that “Chinese” is divided in a north and a south branch, the former being influenced by Altaic, the latter by Tai; for the references, see Chappell (2001b).

extensive enough to make one realize how versatile and pervasive the ACQ element is; as a result, not every example sentence will be directly relevant for the main topic of the paper. At the same time, the overview is meant to show that, despite the all-pervasiveness of the phenomenon, there is variation when we look at the different languages as individual systems, regardless of the genetic affiliation. This section is loosely organized according to usage type. Starting with ACQ as a lexical verb ‘acquire’, we will gradually work our way towards our main topic of interest, the use of ACQ as a post-verbal modal element translatable into English as ‘can.’<sup>9</sup>

### 3.1 Lexical verb ‘come to have’

All four languages have ACQ as a lexical verb meaning something like ‘acquire’; as it is always non-agentive (in any case, not subject controlled), Enfield (2003) glosses it as ‘come to have’.

With respect to ACQ ‘come to have’, Zhuang finds itself in between Cantonese on the one hand and Lao and Vietnamese on the other. Grammatically, its ACQ is like the ACQs in Lao and Vietnamese, in that it can function as the main verb of a sentence by itself, while ACQ in Cantonese, fixed expressions aside, must be followed by *dou*<sup>2</sup>, a suffixal achievement marker. From the perspective of language use, on the other hand, Zhuang *ndaej* is limited to a number of set collocations, similar to the ones we find with Cantonese *dak*<sup>1</sup>-*dou*<sup>2</sup> (which is even more limited), whereas *daj*<sup>4</sup> and *được* in respectively Lao and Vietnamese have a much broader use.

In actual use, Zhuang *ndaej* seems to be restricted to objects related to ranking, age and profits.

- (3) a. *De ndaej daih-it mingz ha?* (Qin 124 (32))  
3s ACQ ORD-one place Q-PRT  
‘Did s/he become first (e.g., in the class)?’
- b. *O! Gou ndaej gigwz lo!* (Z&Q 74)  
INTERJ 1s ACQ pass PRT  
‘Wow! I passed!’ (lit. got a pass)
- c. *Ngoenzneix gou gai byaek ndaej ngeih-cib maenz*  
today 1s sell vegetables ACQ two-ten unit  
*ngaenz.* (based on Z&Q 275 (9))  
money  
‘Today I earned 20 yuan selling vegetables.’
- d. *Beix gou ndaej roek-cib bi lo, hoeng de lij*  
brother 1s ACQ six-ten year PRT but 3s still  
*bae guh hong.* (Z&Q 90)  
go do work  
‘My brother is (lit. got) sixty years old, but he still goes out to work.’

9. Undefined, general terms and labels, such as “pre-verbal” and “modal element”, are used loosely and have been picked mainly for referential convenience.

The non-subject control nature is clear from the fact that none of these sentences is combinable with agentive adverbs. Example (4), an expansion of (3a), though it is not judged as ungrammatical, is made out to be “strange” or “very awkward”.

- (4) ??*Ngoenzcog gou caenhliengh ndaej daih-it mingz.*  
 tomorrow 1s to.one's.extreme ACQ ORD-one place  
 ‘Tomorrow, I’ll do my utmost to be first (e.g., in the class).’

Cantonese *dak*<sup>1</sup> can also be used as a verb ‘come to have’, though it must be suffixed by *dou*<sup>2</sup>. Tang (2002 (1)) has the following example without *dou*<sup>2</sup>, but other native speakers consulted judge it “hardly Cantonese”, “too Mandarin”:

- (5) *di<sup>1</sup> naan<sup>3</sup>-man<sup>4</sup> dak<sup>1</sup>-zo<sup>2</sup> jat<sup>1</sup>-go<sup>3</sup> gwaai<sup>3</sup> beng<sup>6</sup>.*  
 CL refugees ACQ-PRF one-CL strange disease  
 ‘The refugees contracted a strange disease.’

More idiomatically, we find *dak*<sup>1</sup>, with *dou*<sup>2</sup>, in similar contexts as where we find Zhuang *ndaej*, though *dak*<sup>1</sup> is even more restricted than *ndaej*: we do find it with rankings and marks in school, but not in the context of age or making profits.

- (6) a. *keoi<sup>5</sup> dak<sup>1</sup>-dou<sup>2</sup> sap<sup>6</sup> fan<sup>1</sup>.*  
 3s ACQ-SUCC ten points  
 ‘S/he got a ten points.’  
 b. *ngo<sup>5</sup>-dei<sup>6</sup> dak<sup>1</sup>-dou<sup>2</sup> dai<sup>6</sup>-jat<sup>1</sup> ming<sup>4</sup>.*  
 1p ACQ-SUCC ORD-one name  
 ‘We got first.’

The use of Lao *daj*<sup>4</sup> and Vietnamese *đuợc* is broader than *ndaej* in Zhuang and *dak*<sup>1</sup> in Cantonese. *Daj*<sup>4</sup> and *đuợc* are not limited to a few collocations and can take objects denoting many different kinds of entities that one can “come to have”. Here are some Lao examples.<sup>10</sup>

- (7) a. *kuu<sup>1</sup> daj<sup>4</sup> cot<sup>2</sup>-maaj<sup>3</sup> tòon<sup>3</sup> sip<sup>2</sup> moong<sup>2</sup> saw<sup>4</sup>.* (Enf 82 (14))  
 1s ACQ letter time ten o'clock morning  
 ‘I got a/the letter at ten o'clock in the morning.’  
 b. *daj<sup>4</sup> luuk<sup>4</sup> nùng<sup>1</sup>.* (Enf 84 (29))  
 ACQ child one  
 ‘(I) got one child.’  
 c. *caw<sup>4</sup> han<sup>5</sup> daj<sup>4</sup> thii<sup>2</sup> thò<sup>0</sup>-daj<sup>3</sup>.* (Enf 87 (48))  
 2s that ACQ ORD amount-which  
 ‘What ranking (i.e., how-manieth) did you get (in English class)?’

10. For corresponding examples in Vietnamese, not reproduced here for reasons of space: ‘get a letter’: Duff (9c); ‘get a son-in-law’: Enf 176 (45); ‘get health’: Duff (9d); ‘become third in class’: Enf 178 (59).

- d. *khòj<sup>5</sup> sanaq<sup>1</sup> (na<sup>j2</sup>) kaan<sup>3</sup>-khèng<sup>1</sup>.khan<sup>3</sup> ñok<sup>1</sup>-nam<sup>4</sup>.nak<sup>2</sup> - daj<sup>4</sup>*  
 1s be.victorious in NOM-compete lift-weight ACQ  
*laang<sup>2</sup>van<sup>2</sup> thii<sup>1</sup>-nùng<sup>1</sup>.* (Enf 96 (76))  
 prize ORD-one  
 'I was victorious in the weight-lifting competition – I got first prize.'

Lao *daj<sup>4</sup>* has an additional use which Enfield (2003) describes as meaning 'have a procedural knowledge of'. In this use it has a noun phrase as its object, like stories and songs ('know how to tell/sing a story/song well') or languages: *daj<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> viat<sup>5</sup>* /ACQ language Vietnam/ 'know how to speak Vietnamese' (Enfield 2003, 91). In addition, Lao *daj<sup>4</sup>* can be used as an intransitive verb meaning 'succeed', interpretable as 'win' in sporting contexts (Enfield 2003, 94–97). Both uses are also found with Vietnamese *được* (Enfield 2003, 178–182), e.g., *được trận* 'win a battle' (Matisoff 1991, 422). In as far as I have been able to ascertain, Zhuang *ndaej* and Cantonese *dak<sup>1</sup>* do not feature such usages.

In conclusion, all four languages have a lexical verb ACQ expressing 'come to have'. Grammatically, Cantonese differs from the other three languages in that its ACQ, *dak<sup>1</sup>*, does not generally occur by itself. In this grammatical respect, Zhuang patterns with Lao and Vietnamese. With respect to collocational restrictions in language use, however, Zhuang patterns with Cantonese as in these two languages the use of the verb ACQ 'come to have' is much more limited (and limited to similar environments), than in Lao and Vietnamese.

## 3.2 Pre-verbal uses

### 3.2.1 Zhuang, Lao and Vietnamese

ACQ displays a broad range of modal meanings, most of which are loosely translatable with English *can*. Modal ACQ occurs in both pre-verbal and post-verbal position, with different meanings associated with the different positions. As we will see, with respect to the pre-verbal modal use of its ACQ, Zhuang patterns with Lao (and Vietnamese), more so than with Cantonese.

The meanings or interpretations of pre-verbal ACQ we can distinguish are the following: (a) deontic modality; (b) an aspect-like modal meaning, expressing that the event denoted by the verb it precedes can only have come about due to a prior other event ('result of prior event'-reading, as Enfield 2003 calls it); and (c) a benefactive passive reading. I lump these three readings together mostly for convenience, though doing so underlines my impression that they are related and that all have something to do with the core meaning of 'come to have': 'be allowed', 'get to', 'get'. The (a) and (b) readings are not always easily distinguishable, the choice for the one or the other being determined by context (see Enfield 2003 for discussion).

Vietnamese *được* seems to perform all three functions. First, we have a modal, expressing deontic modality. Here are some examples (see also Cadière 1956):

- (8) a. *Tôi được kiểm việc.* (Duff (15a))  
 1s can look.for work  
 'I am/was permitted to look for work.'



- b. *Khách bảo không cho cái gì vào thi mình*  
 guest tell not give CL whichever enter PRT 1p  
*không được cho vào.* (Enf 302 (27))  
 not ACQ give enter  
 ‘(If) guests tell (us) to not put something in, then we are not to put (it) in.’

The ‘result of prior event’-reading can be illustrated with the following sentences.

- (9) a. *Ngày mai tôi sẽ được làm việc.* (Enf 300 (23))  
 tomorrow 1s IRR ACQ do work  
 ‘Tomorrow I will (get to) work.’  
 b. *Ngày mai tôi sẽ làm việc.* (Enf 300 (24))  
 tomorrow 1s IRR do work  
 ‘Tomorrow I will work.’

Enfield explains (p. 300), that the sentence in (9b) states a simple fact: the speaker has a regular job and goes to work every day, and tomorrow is a working day. In (9a), the implication is that the speaker normally does not work or is only on call, and something happened or is to happen with the result that s/he will have work to do the next day. Hence Enfield’s gloss here as ‘result of prior event’. Note that, although a translation into English using ‘get to’ or ‘have a chance to’ comes closer to the meaning of the Vietnamese original (something can happen only as the result of an unspecified prior event), translating *được* using a deontic modal is not entirely unfeasible (and vice versa, cf. (8a)) (see Enfield 2003 for discussion).

Finally, *được* in pre-verbal position can have the so-called benefactive passive reading; here are some examples, which do not need explication. This use can probably also be subsumed under some general pre-verbal usage, related to the core meaning ‘come to have’.

- (10) a. *Dân chúng được chính phủ xây cho một cái cầu.* (Duff (15b))  
 people ACQ government build for one CL bridge  
 ‘The people had this bridge built for them by the government.’  
 b. *Tôi được mời đi ăn cưới ở nhà bạn.* (Enf 303 (32))  
 1s ACQ invite go eat wedding be.at house friend  
 ‘I (fortunately) got invited to attend a wedding at a friend’s house.’  
 c. *Tám được Liên yêu.* (Nguyễn 1997, 113)  
 T’am ACQ Liên love  
 ‘Tám was (had the pleasure of being) loved by Liên.’

The following example shows that the three readings are sometimes hard to tease apart.

- (11) *Khách sạn này được mở bao nhiêu lâu rồi.* (Enf 305 (44))  
 hotel this ACQ open how.much long.time PRT  
 ‘How long has this hotel been allowed to open?’ (i.e., ‘how long since this hotel has got to be open?’)

Turning to Lao, when *daj*<sup>4</sup> occurs pre-verbally adjacent to the verb, it is unstressed and has a zero tone: *daj*<sup>0</sup>. Aside from the zero tone, Enfield mentions a number of other facts which constitute good reasons for assuming that it is not a regular modal verb here. For instance, *daj*<sup>0</sup> is not a possible answer in a yes/no question, which pre-verbal modals generally are.

In Lao, we do not find the benefactive passive reading with *daj*<sup>0</sup>, but both other interpretations (result of prior event; deontic modality) are attested; here too, the precise meaning depends on the context. Here are some examples.

- (12) a. *khan*<sup>2</sup> *haw*<sup>2</sup> *bò*<sup>0</sup> *qòòk*<sup>5</sup> *paj*<sup>3</sup> *juu*<sup>1</sup> *taang*<sup>2</sup>, *haw*<sup>2</sup> *bò*<sup>1</sup> *daj*<sup>0</sup>  
 if 1p not go.out go be.at separate 1p not ACQ  
*kin*<sup>3</sup> *khaw*<sup>5</sup> *děj*<sup>2</sup>. (Enf 141 (222))  
 eat rice PRT  
 'If we don't go and live separately [from your mother], we won't get to eat, you know [because she will continue to gamble our money away].'
- b. *khòò*<sup>3</sup> *haj*<sup>5</sup> *luuk*<sup>4</sup> *nòòj*<sup>4</sup> *khòòj*<sup>5</sup> *ní*<sup>0</sup> *daj*<sup>0</sup> *juu*<sup>1</sup> *pên*<sup>3</sup> *too*<sup>3</sup> *ñaj*<sup>1</sup>  
 request give child small 1s TOP ACQ be.at be body big  
*khùn*<sup>5</sup> *paj*<sup>3</sup>. (Enf 141 (224))  
 ascend go  
 'Please allow this small child of mine to have a chance to grow up into an adult.'
- c. *khòòj*<sup>5</sup> *daj*<sup>0</sup> *kin*<sup>3</sup> *kop*<sup>2</sup>. (Enf 153 (265))  
 1s ACQ eat frog  
 'I got/had to eat frog.'
- d. *phen*<sup>1</sup> *huu*<sup>3</sup> *nuak*<sup>5</sup> – *daj*<sup>0</sup> *vaw*<sup>4</sup> *dang*<sup>3</sup>. (Enf 147 (242))  
 3HON ear deaf ACQ speak loud  
 'He's deaf – (one) has to speak loud.'

Zhuang is very much like Lao. We do not see the benefactive passive reading, but *ndaej* does have the other two. As to the deontic modal, all examples of this usage that I have come across in published sources contain a negation, *mbouj ndaej* 'may not', 'must not', 'should not', 'cannot' (see the examples below), which gives the impression that it cannot be used affirmatively. One of our informants notes that in his own dialect (Héchi), pre-verbal modal *ndaej* is exclusively used with *mbouj* 'not'; for the affirmative, a Mandarin loan, *kéyi* 'can, may', is used.

- (13) a. *Mwngz youq baihrog mbouj ndaej dub vunz ha!*  
 2s at outside not ACQ hit people PRT  
 'Out there you should not hit people!' (Qin 120 (5))
- b. *Giyenz mwngz siengj ok-rog bae ra hong guh, couh*  
 as 2s want outside go seek work do then  
*mbouj ndaej lau dwgrengz.* (Qin 262 (44))  
 not ACQ fear hardship  
 'Now that you want to go and look for work somewhere else, you must not be afraid of hardships.'

- c. *Aen maenz neix gyuk, mbouj ndaej gwn ndip.* (Qin 199 (17))  
 CL yam this dirty not ACQ eat raw  
 'This yam is dirty, you shouldn't eat it raw.'
- d. *Danghnaeuz mbouj coih mieng, couh mbouj ndaej hai raemx*  
 if not build canal then not ACQ open water  
*haeuj naz.* (Z&Q 90)  
 enter field  
 'If we don't build a canal, we will not get to irrigate the fields.'

This last example can be interpreted to illustrate the 'result of prior event'-reading. Here are two more:

- (14) a. *Beix gou ndaej hot geij coenz.* (cf. Enf 327 (29b))  
 brother 1s ACQ speak some word  
 'My elder brother got to say a few words.'
- b. *De ndaej hwnj hag lo.*  
 3s ACQ go.to school PRT  
 'He can go to school now.'

As our informant explains, uttered without any context, (14b) implies that the subject did not go to school on a regular basis and that there was no reason to assume that he would. But something happened such that this all changed (the whole community collected money, for instance).

### 3.2.2 Cantonese

Cantonese is the only language of the four we study here that has no pre-verbal ACQ, at least, on the surface. When we look below the surface, however, we see that Cantonese has the two pre-verbal elements that Lao and Zhuang have, the modal and the 'result of prior event'-marker.

Starting with the former, taken at face value, Cantonese modal *dak*<sup>1</sup> only occurs in post-verbal position, where it can have two meanings: permission (15a) or potential (15b) (see §3.4) (examples from Cheng and Sybesma 2004).

- (15) a. *keoi<sup>5</sup> zaa<sup>1</sup>-dak<sup>1</sup> li<sup>1</sup>-gaa<sup>3</sup> ce<sup>1</sup>.*  
 3s drive-ACQ this-CL car  
 'S/he can (here: is permitted to) drive this car.'
- b. *keoi<sup>5</sup> lo<sup>2</sup>-dak<sup>1</sup>-hei<sup>2</sup> li<sup>1</sup>-seung<sup>1</sup> syu<sup>1</sup>.*  
 3s take-ACQ-up this-box book  
 'S/he can (here: will manage to) lift this box of books.'

Cheng and Sybesma (2004) observe that there is a correlation between the interpretation of *dak*<sup>1</sup> and the structural environment in which it occurs. Their analysis of the different environments leads them to propose that *dak*<sup>1</sup>-permission is actually generated in pre-verbal position where we also find the other modals (*dak*<sup>1</sup>-potential is generated in a position following the verb, as we discuss in §4). Permission *dak*<sup>1</sup> ends

up post-verbally, when the verb, for reasons that have to do with *dak*<sup>1</sup>'s defective nature in certain respects, moves up (for the details, see op. cit.). In other words, according to this analysis, if any modal *dak*<sup>1</sup> occurs pre-verbally in any stage of the derivation, it is *dak*<sup>1</sup>-permission. This fits in with other varieties of Chinese, in which ACQ does occur, and surface, as a pre-verbal modal. Mandarin, for instance, has deontic *děi* 'must'. Wu (2001) reports on the Hunan dialects, whose ACQs appear in pre-verbal position, expressing, loosely, volition and possibility (but not ability) (in most examples Wu provides, pre-verbal ACQ is accompanied by negation).

Cantonese also has ACQ as the pre-verbal 'result of prior event' marker, though again disguised as a post-verbal element. Now, however, it is more literally pre-verbal in the sense that it precedes the main verb; it is post-verbal because it is preceded by *jau*<sup>5</sup> 'have' or *mou*<sup>5</sup> 'not have'. The following sentences illustrate this.

- (16) a. *keoi*<sup>5</sup> *jau*<sup>5</sup>-*dak*<sup>1</sup> *heoi*<sup>3</sup> *Zung*<sup>1</sup>-*gok*<sup>3</sup>.  
 3s have-ACQ go China  
 '[Now the situation is such that] s/he gets to go to China.'
- b. *ngo*<sup>5</sup>-*dei*<sup>6</sup> *gam*<sup>1</sup>-*jat*<sup>6</sup> *jau*<sup>5</sup>-*dak*<sup>1</sup> *sik*<sup>6</sup> *kavai*.  
 1p today have-ACQ eat kaviar  
 'Tonight we will get to eat kaviar (such a special occasion).'
- c. *maa*<sup>1</sup>-*mi*<sup>4</sup> *mou*<sup>5</sup>-*dak*<sup>1</sup> *zou*<sup>6</sup>-*je*<sup>5</sup>.  
 mummy not.have-ACQ work  
 '[This way] mommy can't do any work.'

These sentences have exactly the meaning that Enfield describes for the relevant sentences in Lao: something not disclosed happened such that the event denoted by the predicate followed by ACQ can (or cannot) take place.

### 3.2.3 Conclusion

If our views on Cantonese are right, all languages use their pre-verbal ACQ as a marker of deontic modality or 'result of prior event'. Vietnamese is the only language which has an additional function, as marker of the benefactive passive. Grammatically, Cantonese differs from the other languages in that its pre-verbal ACQ must always be preceded by a verbal element, either the main verb itself or *jau*<sup>5</sup> 'have' or *mou*<sup>5</sup> 'not have' (possibly analysable as inserted dummies). In the other three languages, pre-verbal ACQ also surfaces pre-verbally, though it must be noted that it is phonologically reduced in Lao and that in Zhuang it must be accompanied by negation. In all, taking both language use and grammar into consideration, Zhuang is closest to Lao.

## 3.3 Post-verbal functions

We now move into the post-verbal domain where Zhuang will turn out to share some crucial properties with Cantonese. In this section we look at the three uses of ACQ in post-verbal position, which can be characterized as rather grammatical-functional: the use as (a) post-verbal linker (§3.3.1); (b) *-able* marker (§3.3.2); and (c) aspectual

achievement marker (§3.3.3). The focus will be on the differences in the positioning of ACQ relative to other post-verbal elements, as this will be an important issue in answering the question why post-verbal modal ACQ in Zhuang cannot have an ability reading, while this is no problem for its Lao counterpart.

### 3.3.1 *Post-verbal linker*

In all four languages ACQ is used to link the verb to different expressions following it, expressing notions as duration, manner/degree, and extent. Extent phrases are generally translatable with *so . . . that* and are in some ways result phrases.

In Zhuang, *ndaej* may be followed by all three types of expressions (one at the time), as the following examples show. The meaning given is not necessarily the only meaning these sentences may have; to some we return below.

(17) DURATION

- a. *Duz mou neix gou ngamq ciengx ndaej donh bi.* (Qin 82 (18))  
 CL pig this 1s just keep ACQ half year  
 ‘This pig, I just kept for half a year.’

MANNER/DEGREE

- b. *De ciengq ndaej caen ndei ha!* (Qin 120 (6))  
 3s sing ACQ really good SFP  
 ‘S/he sings really well!’
- c. *Duz mou neix gwn ndaej lai ho!* (Qin 148 (34))  
 CL pig this eat ACQ much PRT  
 ‘That pig ate a lot!’
- d. *De byaet ndaej lai doengz gou lo.* (Qin 228 (5))  
 3s peel ACQ much as 1s LO  
 ‘S/he peeled as much as I did.’

EXTENT

- e. *Aen lwgmanh neix manh ndaej raemxda conh bae.* (Qin 159 (36))  
 CL pepper this hot ACQ tear stream SFP  
 ‘This pepper is so hot that tears come out.’
- f. *Duz ma gyaep ndaej gyoengq bit de lek-byazbyaz.*  
 CL dog chase ACQ CL-group duck that frighten-MODIF  
 ‘The dog chased the ducks such that they panicked.’ (Qin 223 (26))

In these sentences, *ndaej* immediately follows the verb; as far as I have been able to ascertain, adjacency is obligatory in the manner/degree and extent cases and strongly preferred with durational expressions.

In the negative counterparts of these sentences, in the interpretations given, *mbouj* ‘not’ follows *ndaej*. Here is the negative counterpart of (17b).

- (18) *De ciengq ndaej mbouj ndei.*  
 3s sing ACQ not good  
 ‘S/he does not sing well.’

*Ndaej* is not the only element that functions as a linker between the verb and this kind of expression (Qin 1995, §3.15; Zhang and Qin 1993, §3.3.2). Other such elements are *dwk* and *gwj*. The general rule is that *ndaej* is used for duration and manner/degree, but when the manner/degree is expressed using an onomatopoeic expression, *gwj* can also be used. With extent phrases, *ndaej* and *dwk* are mostly interchangeable, though when the expression following the verb is monosyllabic or when it is an adjective, *ndaej* is preferred over *dwk*. Conversely, when the extent expression takes the form of a full sentence, *dwk* is generally preferred over *ndaej*.<sup>11</sup> Here is an example with *dwk* (though it could be replaced by *ndaej*):

- (19) *De nit dwk fwngz saenz lo.* (Qin 217 (2))  
 3s cold DWK hand shake PRT  
 'He was so cold that his hands were shaking.'

In sum, in Zhuang, *ndaej* can be used as a linker between the verb and expressions of duration, manner/degree and extent. For some extent phrases, linker element *dwk* is preferred over *ndaej*. We observe further, that with manner/degree and extent expressions, *ndaej* is invariably adjacent to the verb (as is *dwk*), and for durational phrases adjacency is strongly preferred.

Turning to Lao, we find *daj*<sup>4</sup> with post-verbal constituents expressing duration and manner/degree, but not with extent. Here are some examples; once again, some sentences have more meanings/interpretations than given here.

- (20) a. *man<sup>2</sup> phak<sup>1</sup> juu<sup>1</sup> nii<sup>4</sup> daj<sup>4</sup> sòòng<sup>3</sup> dùan<sup>3</sup>.* (Enf 134 (193))  
 3s stay be.at here ACQ two month  
 'S/he stayed here for two months.'
- b. *com<sup>1</sup> thòòng<sup>1</sup>.khùn<sup>5</sup>.caj<sup>3</sup> daj<sup>4</sup> lian<sup>1</sup>.laj<sup>3</sup> kua<sup>1</sup> muu<sup>1</sup>.*  
 mutter 'off-by-heart' ACQ flowing more.than peer  
 '(I) would mutter (the chants) off by heart much more fluently than the others.' (Enf 139 (214))
- c. *man<sup>2</sup> kin<sup>3</sup> daj<sup>4</sup> laaj<sup>3</sup>.* (Enf 138 (210))  
 3s eat ACQ much  
 'S/he ate a lot.'

Enfield notes that for durational expressions as in (20a), *daj*<sup>4</sup> is optional; this is not true for manner/degree expressions. More importantly, Lao *daj*<sup>4</sup> does not have to be adjacent to the verb, unlike its counterpart in Zhuang. In fact, as Enfield notes, *daj*<sup>4</sup> can at times be preceded by adverbials like *kùap*<sup>5</sup> 'almost' (as in (21a)) or aspect-modality markers such as *ca*<sup>0</sup> for irrealis/future, (21b).

11. One type of extent phrase never has *dwk*. If the subject of the extent expression is also interpreted as the subject or the object of the matrix verb, we exclusively have *ndaej*. For examples: Qin 223 (25), (27).

- (21) a. *man<sup>2</sup> phak<sup>1</sup> juu<sup>1</sup> nii<sup>4</sup> kùap<sup>5</sup> daj<sup>4</sup> sòong<sup>3</sup> dùan<sup>3</sup>.* (Enf 136 (200))  
 3s stay be.at here almost ACQ two month  
 ‘S/he stayed here for almost two months.’
- b. *man<sup>2</sup> phak<sup>1</sup> juu<sup>1</sup> nii<sup>4</sup> ca<sup>0</sup> daj<sup>4</sup> sòong<sup>3</sup> dùan<sup>3</sup>.* (Enf 136 (201))  
 3s stay be.at here IRR ACQ two month  
 ‘S/he will have stayed here for two months.’

In the Zhuang example in (17a), the adverb is in front of the verb; it never appears between the verb and *ndaej*.

Extent expressions taking the form of a sentence are not possible in Lao with *daj<sup>4</sup>*. In such cases Lao uses the element *con<sup>3</sup>* ‘until’, see (22).

- (22) *khaw<sup>3</sup> haj<sup>5</sup> con<sup>3</sup> phaa<sup>5</sup>-sê<sup>1</sup>-mùu<sup>2</sup> piak<sup>5</sup> met<sup>2</sup>.* (Enf 284 (133))  
 3p cry CON cloth-wipe-hand wet all  
 ‘They cried so much that their handkerchiefs got all wet.’

Vietnamese is very much like Lao: like *daj<sup>4</sup>*, *được* can link the verb to expressions of duration and manner/degree:

- (23) a. *Tôi ở Hà Nội được ba năm rồi.* (Enf 261 (41))  
 1s be.at Hanoi ACQ three year PERF  
 ‘I have been in Hanoi for three years already.’
- b. *Anh ấy chạy được nhanh lắm.*  
 man that run ACQ fast very  
 ‘That man runs very fast.’

And, again like Lao *daj<sup>4</sup>*, Vietnamese *được* does not have to be adjacent to the verb; indeed, as Enfield observes, it can at times be modified by an aspectual-modal element, underscoring its independent nature.

- (24) *Tôi ở Hà Nội đã được ba năm (rồi).* (Enf 262 (43))  
 1s be.at Hanoi PERF ACQ three year PERF  
 ‘I have been in Hanoi for three years already.’

Like Lao, Vietnamese doesn’t use its ACQ to introduce extent phrases; instead it uses *đến-nơi* ‘until’ (Enfield 2003, 286):

- (25) *Anh ấy làm tiếng động đến-nơi nhà bên cạnh  
 man that make sound loud until house direction side  
 không ngủ được.* (Enf 286 (145))  
 not sleep ACQ  
 ‘He made noise such that (the people in) the house next door couldn’t sleep.’

Turning, finally, to Cantonese, extent phrases in this language are not marked with *dak<sup>1</sup>*, but rather with *dou<sup>3</sup>* ‘until’ or ‘arrive at’ (see (26a)), and duration phrases are unmarked (as in (26b)). Manner expressions, however, are introduced by *dak<sup>1</sup>*, as

illustrated in (26c) (Lamarre 2001; Yue 2001; Matthews and Yip 1994). We must emphasize that linker *dak*<sup>1</sup> is always adjacent to the verb.

- (26) a. *keoi*<sup>5</sup> *zau*<sup>2</sup> *dou*<sup>3</sup> *gui*<sup>6</sup> *saai*<sup>3</sup>. (Yue 2001, 233)  
 3s run until tired completely  
 'S/he ran to the extent that s/he became exhausted.'
- b. *ngo*<sup>5</sup> *tai*<sup>2</sup> *li*<sup>1</sup>-*bun*<sup>2</sup> *syu*<sup>1</sup> *tai*<sup>2</sup>-*zo*<sup>2</sup> *saam*<sup>1</sup>-*go*<sup>3</sup> *zung*<sup>1</sup>-*tau*<sup>4</sup>.  
 1s read this-CL book read-PRF three-CL hour  
 'I read this book for three hours.'
- c. *keoi*<sup>5</sup> *zau*<sup>2</sup> *dak*<sup>1</sup> *hou*<sup>2</sup> *faai*<sup>3</sup>. (Yue 2001, 233)  
 3s run ACQ very fast  
 'S/he runs very fast.'

Comparing the languages with respect to the use of linker ACQ, we get an interesting picture. Zhuang is the only language that uses ACQ with all three types of expressions. In the other languages, the extent phrase is not marked using ACQ. But even in Zhuang, the extent phrase has a special status in that there are some types that preferably go with another linker. What is so interesting, is that, within the Sinitic world, using a different marker for extent phrases and ACQ for the others is an almost exclusively Cantonese characteristic (Lamarre 2001). Cantonese is much like Tai Lao in this respect, while Zhuang, in using its ACQ in extent phrases as well, is actually quite Sinitic.

With respect to one other feature, Zhuang is also clearly on the Sinitic side of the dividing line, this time together with Cantonese, viz., the fact that ACQ must be adjacent to the verb.<sup>12</sup> This is the case in Sinitic languages, while it is not the case in Lao and Vietnamese.

### 3.3.2 *-able*

In all four languages, ACQ can be added to a transitive verb, yielding an effect similar to the morpheme *-able* in English *doable*. These sentences, exemplified in (27) in Zhuang, have the following structure: [Subj<sub>underlying obj</sub> V ACQ].

- (27) a. *Aen ci neix coih ndaej*. (Qin 205 (1))  
 CL vehicle this repair ACQ  
 'This bike is repairable.'
- b. *Aen ranz neix youq ndaej*. (Qin 205 (2))  
 CL house this live ACQ  
 'This house is livable – one can live in it.'
- c. *Aen bat de lij yungh ndaej*. (Qin 205 (3))  
 CL basin that still use ACQ  
 'This washbasin is still usable.'

12. Except for durational expressions, for which it is strongly preferred. We will not keep mentioning this near-exception, but we have to be aware of it.



- d. *Aen mak neix gwn ndaej.* (Z&Q 174)  
 CL fruit this eat ACQ  
 ‘This piece of fruit is edible.’

The negative counterpart of this type of sentence has *mbouj ndaej* ‘NEG ACQ’ following the verb:

- (28) a. *Go faex de ha, soh seih soh la, hoeng*  
 CL tree that PRT straight be straight PRT but  
 yungh mbouj ndaej. (Qin 125 (36))  
 use not ACQ  
 ‘As to this tree, you can call it straight, but it is unusable.’  
 b. *Aen mak neix gwn mbouj ndaej.* (Z&Q 174)  
 CL fruit this eat not ACQ  
 ‘This piece of fruit is inedible.’

Cantonese (in (29a)), Lao ((29b); Nick Enfield, p.c.) and Vietnamese (in (29c)<sup>13</sup>) are the same:

- (29) a. *go<sup>3</sup> ping<sup>4</sup>-gwo<sup>2</sup> sik<sup>6</sup>(-m<sup>4</sup>-)dak<sup>1</sup>.*  
 CL apple eat-not-ACQ  
 ‘The apple is (in)edible.’  
 b. *maak<sup>5</sup> nii<sup>4</sup> kin<sup>3</sup> (bò<sup>0</sup>) daj<sup>4</sup>.*  
 fruit this eat not ACQ  
 ‘This fruit is (in)edible.’  
 c. *Món này ăn được không?*  
 dish DEM eat ACQ not  
 ‘Is this dish edible?’

### 3.3.3 Aspectual achievement marker

In Zhuang, Vietnamese and Lao, but not in Cantonese, ACQ is used to indicate that the action denoted by the verb has been brought to a successful conclusion. I will call it an “achievement marker” here (for some discussion, see §4.3). In Cantonese one will generally use achievement marker *dou*<sup>2</sup>.

In Zhuang (much less so in Lao and Vietnamese), achievement ACQ is mainly found with verbs of acquisition, like verbs of buying and catching; in Lao, for instance, it is also found with *khaa*<sup>5</sup> ‘kill’. Here are some Zhuang and Lao examples. As before, the interpretation represented in the idiomatic translation may not be the only one these sentences have.

13. Matisoff (1991, 424) notes: “With some preceding verbs, *được* is a neat translation of the English suffix *-able/-ible*.” The example in (29c) is from Thompson (1965, 345).

## ZHUANG

- (30) a. *Ngoenzneix gou cawx-ndaej song duz bit.* (cf. Qin 90 (1))  
 today 1s buy-ACQ two CL duck  
 'Today I bought two ducks.'
- b. *Daegnuengx gaeb-ndaej song lwg gungq.* (Qin 100 (5))  
 l'l.brother catch-ACQ two CL shrimp  
 'Little brother caught two shrimps.'
- c. *Dingqnyi mwngz gauj-ndaej dazyoz, dwg le?* (Qin 122 (15))  
 I.heard 2s test-ACQ university be PRT  
 'I heard that you passed the university entrance exam, is that so?'

## LAO

- (31) a. *man<sup>2</sup> haa<sup>3</sup> kacèè<sup>3</sup> daj<sup>4</sup>.* (Enf 99 (85))  
 3s seek key ACQ  
 'S/he found the keys.'
- b. *laaw<sup>2</sup> khut<sup>2</sup> jaa<sup>3</sup> daj<sup>4</sup>.* (Enf 125 (171))  
 3s dig.for medicine ACQ  
 'S/he dug up the medicine.'

As to word order, Zhuang *ndaej* invariably immediately follows the main verb, and no other order is possible. In the Lao examples in (31), *daj<sup>4</sup>* follows the object: [V O *daj<sup>4</sup>*]. This seems to be the most common order, but we also come across [V *daj<sup>4</sup>* O]. Here is a minimal pair; the word order difference does not seem to correlate with any other differences (see Enfield 2003, 118ff for discussion).

- (32) a. *cùng<sup>1</sup> sùu<sup>4</sup> daj<sup>4</sup> khaw<sup>5</sup> sahakòòn<sup>3</sup>.* (cf. Enf 129 (176), 131 (184))  
 thus buy ACQ rice cooperative  
 'Thus they bought cooperative rice.'
- b. *cùng<sup>1</sup> sùu<sup>4</sup> khaw<sup>5</sup> sahakòòn<sup>3</sup> daj<sup>4</sup>.* (Enf 130 (180))  
 thus buy rice cooperative ACQ  
 'Thus they bought cooperative rice.'

Vietnamese displays this kind of distributional flexibility with respect to the placement of ACQ much more generally (though see for discussion Duffield 2003; Simpson 2001; Matisoff 1991; and below). Here are some examples. Note the minimal pair in (34).

- (33) a. *Tôi bảo vệ tiên si được.* (Enf 239 (188))  
 1s defend Ph.D. ACQ  
 'I passed my Ph.D. defense.'
- b. *Chị ấy thuê được một cái nhà thật rẻ.* (Bisang 1992, 300 (46))  
 3sf DEM rent ACQ one CL house really cheap  
 'She rented a really cheap place.'
- (34) a. *Tôi kiếm việc được.* (both: Duff (14))  
 1s look.for work ACQ

- b. *Tôi kiếm được việc.*  
 1s look.for ACQ work  
 BOTH: 'I found work.'

### 3.3.4 Conclusion

After looking at linker-ACQ, *-able*-ACQ and ACQ-achievement, we conclude that in all cases, Zhuang *ndaej* must be adjacent to the verb. In this respect Zhuang differs from Lao, whose *daj*<sup>4</sup> is generally placed behind the object. The adjacency requirement of verb and ACQ is actually quite Sinitic.

## 3.4 Modal 'can' and lexical verb 'okay'

### 3.4.1 Introduction

All four languages have ACQ in post-verbal position in a meaning translatable into English using 'can'. We saw examples in (2). The use of *can* in the translation disguises the difference between Zhuang and Lao which we singled out as the main problem of this paper: in (1b) ACQ is used to express (acquired) ability, which its Zhuang counterpart could never express. In this respect, Cantonese sides with Zhuang (never ability), while Vietnamese is like Lao (ability possible; see (2d)).

In this section, we investigate the two different 'can'-modalities that are involved: ability and the so-called potential, which will be defined and discussed shortly.

When ACQ follows another verb in the sentence, it is not always a modal element. In some cases it is a lexical verb meaning 'be okay'. Whether we are dealing with the modal or the verb 'okay' is not always immediately clear, as the following Zhuang sentence illustrates.

- (35) *De gangj vah Yeznanz ndaej.*  
 3s speak language Vietnam ACQ  
 'He can speak Vietnamese.' (that is: it's okay if he does)

In order to get a clear picture of modal ACQ, our main focus, it is necessary to tease the 'okay' use apart from the modal use. That is why we look at these two uses in this one section.

Before taking a general view, we have a close look at the individual languages, starting with Zhuang.

### 3.4.2 Zhuang

**3.4.2.1 Modal 'can': potential modality.** We only find Zhuang *ndaej* 'can' in post-verbal position in the context of telic predicates, such as the resultatives in (36) (intransitive (36a, b) as well as transitive (36c, d) and (2a) above).

- (36) a. *Haemhlwenz ha, mbouj miz bouxlawz ninz ndaej*  
 last.night PRT not have/be who fall.asleep ACQ  
*ndaek.* (Qin 127 (47))  
 SUCC  
 'Last night nobody could fall asleep.'

- b. *De saet ndaej gvaqbae bw.* (Qin 142 (24))  
 3s jump ACQ past SFP-SURPRISE  
 'Wow! He can really jump over!'
- c. *Gou sij ndaej liux bonj saw gou.*  
 1s write ACQ done CL book 1s  
 'I can finish my book.'
- d. *De haeb ndaej dai duz-gaeq gou.*  
 3s bite ACQ dead CL-chicken 1s  
 'It can bite my chicken to death.'

In all these sentences we find *ndaej* in between the verb and a result denoting predicate. Other orders are not possible (with the relevant reading).

In the negative counterparts of these sentences, *ndaej* is replaced by *mbouj* 'not':

- (37) a. *Haemhlwenz miz vunz doxceng, gou beanz hwnz ninz*  
 last.night have people fight 1s whole night fall.asleep  
*mbouj ndeak.* (Qin 262 (42))  
 not SUCC  
 'Last night, some people were having a fight, I could not fall asleep the whole night.'
- b. *De siengj benz hwnj bya bae, roen myag lai,*  
 3s want climb up mountain go road slippery very  
*benz mbouj hwnj.* (Qin 258 (24))  
 climb not up  
 'He wanted to climb up that mountain, but the road was very slippery, so he could not climb up.'
- c. *Daxboh dub mbouj dai duzma gou.*  
 father hit not die dog 1s  
 'Father cannot beat my dog to death.'

What exactly is the meaning of these sentences? The sentences in (36) and (37) consist of two predicates, the main predicate and the resultative predicate. The two predicates are separated by *ndaej* or its negative counterpart *mbouj*, which contribute the 'can' modality to the sentence. It is important to realize that the 'can' modality does not have scope over the main predicate. All that is at issue in these sentences is whether or not the subject will manage to reach the resulting state. This is especially clear in the negative sentences: in (37c), for instance, the implication is clearly that *daxboh* 'father' can *dub* 'beat'; what he cannot do is reach the projected resulting state of the dog's death. Similarly, (36b) is not about whether the subject can jump – the assumption is that he can, though nothing is said about it – but whether he can reach the other side, that is, whether he can successfully bring the event to the projected end. We conclude that *ndaej* does not have scope over the matrix predicate. The scope of the modality is restricted to the result denoting part of the sentence: can or can we not reach the designated end point. We follow the Chinese grammatical tradition and call this modality "potential modality".

For the sake of the discussion in the rest of the paper, we need to distinguish the notions “potential construction” from “potential modality”. We define the potential construction as a construction in which a modal ‘can’ occupies a position in between the matrix verb and an element that indicates the end point of the action denoted by the matrix verb (such as a resultative predicate; for details, see §4). Potential modality is defined as concerning the ability to reach the projected end point.

Although there is no potential construction without potential modality, the reverse is not true. The combination of telicity and ‘can’ generally yields potential modality. In English, for instance, when we say *he can wipe the table clean*, we get the potential reading: the focus of the modality is on the result, not on the action denoted by the matrix verb. The question is whether it is the only reading. In other words: telicity is a necessary condition for the potential (by definition), but is it also a sufficient condition? If the English sentence in question has any other reading at all, it is not ability with respect to the activity denoted by the matrix verb. Even if we stress the main verb, *he cannot wipe the table clean*, we still only have the potential interpretation: the sentence says that he cannot get the table clean by wiping, it doesn’t say anything about his ability to wipe. If telicity is indeed a sufficient condition, ability and potentiality are mutually exclusive in any one sentence: they are two shades of the same meaning, the former surfacing in atelic sentences, the latter in telic ones.<sup>14</sup>

In any case, sentences like the English one just discussed underscore the point that one can have potential modality without the potential construction.

Does *ndaej* ever cooccur with atelic predicates? The answer is ‘no’, as the data in (38) show: ACQ in Zhuang does not occur with atelic events or have an (acquired) ability reading (but see §3.4.2.2). To express ability, Zhuang uses the pre-verbal modal *rox* ‘can, be able’. Note that the sentences marked with \* are ungrammatical in the intended reading and that the sentences are ungrammatical wherever *ndaej* is placed: after the verb or at the very end of the sentence.

- (38) a. \**De gangj {ndaej} vah Yeznanz {ndaej}*. (= (1a))  
           3s speak ACQ language Vietnam ACQ  
           INTENDED: ‘S/he is able to speak Vietnamese.’  
       a’ *De rox gangj vah Yeznanz*.  
           3s can speak language Vietnamese  
           ‘S/he is able to speak Vietnamese.’  
       b. \**De hai {ndaej} ci {ndaej}*.  
           3s drive ACQ vehicle ACQ  
           INTENDED: ‘S/he is able to drive a car.’  
       b’ *De rox hai ci*.  
           3s can drive vehicle  
           ‘S/he is able to drive a car.’

14. If *can* in telic sentences has another reading it is the permission reading, thus suggesting that permission is not another shade of the same meaning.

- c. \**De coih {ndaej} gyaq danci gou {ndaej}*.  
 3s repair ACQ CL bicycle 1s ACQ  
 INTENDED: 'S/he can repair my bike.'
- c'. *De rox coih gyaq danci gou*.  
 3s can repair CL bicycle 1s  
 'S/he can repair my bike.'

3.4.2.2 *Apparent counterexamples*. There are two apparent counterexamples to the claim that Zhuang *ndaej* 'can' only occurs with resultatives or telics. Consider first (39):

- (39) *Duz roeg de lij mbin ndaej*. (Qin 206 (8))  
 CL bird that still fly ACQ  
 'That bird can still fly.'

This is the only example of this type I have come across in published sources: *ndaej* follows a simple verb and it seems to express 'can'-ability. On the basis of this example it is impossible to determine whether *ndaej* is a sentence final element that happens to be adjacent to the verb or an element that is more like a verbal affix that happens to find itself in sentence final position. The following example shows however that the latter is the case: with a simple verb and a simple object, *ndaej* appears right in the middle of these two:

- (40) *De lij gangj ndaej vah*.  
 3s still speak NDAEJ speech  
 'S/he can still speak.'

The sentences in (39) and (40) can be used in one kind of context only, namely, in which the subject is expected not to be able to fly or speak. Thus, when someone utters the sentence in (39) out of the blue, the hearer knows that something had happened to the bird – it was wounded, shot at – such that one would expect it to be unable to fly. Similarly, the sentence in (40) can be uttered when someone had a stroke or broke their jaw. In both cases, there is a strong sense of achievement, of being able to reach a certain result. In other words, the events involved in these sentences are actually telic in that a certain projected end point is implied (fly *up*, speak *out*).

The second apparent counterexample to the generalization regarding the distribution of post-verbal modal *ndaej* in Zhuang is constituted by sentences like those in (41): exclamatives with *ndaej* between a verb and a quantity expression.

- (41) a. *Daeg beix gwn ndaej sam gaen haeuxcid bae!* (Qin 157 (18))  
 elder brother eat ACQ three pound rice PRT  
 'Elder brother can eat three pounds of rice!'
- b. *Aen danci neix doz ndaej haj boux vunz bae!* (Qin 157 (20))  
 CL bicycle this fit ACQ five CL people PRT  
 'This bike can fit five people!'

- c. *Duz ma gou bongh ndaej sam cik sang bae!* (Qin 157 (22))  
 CL dog 1s jump ACQ three foot high PRT  
 'This dog can jump three feet high!'
- d. *De fwngz ndeu yaeuj ndaej song doengj raemx bae!* (Qin 157 (23))  
 3s hand one raise ACQ two bucket water PRT  
 'S/he can lift two buckets of water with one hand!'

Although no explicit end point is mentioned, the events in these sentences are nonetheless telic. Like (39) and (40), they have a strong flavor of achievement, of the ability to reach a projected end point, an effect that is strengthened by the exclamative nature of these sentences. Note that the object is a bounded quantity expression, a typical property of telic predicates (Tenny 1987).<sup>15</sup>

In all, none of the sentences in this section constitute a true counterexample to the claim that *ndaej* is only compatible with telic events. The question whether we are also dealing with the potential construction here will be discussed in §4.

**3.4.2.3 'Okay'.** In all the post-verbal uses of Zhuang *ndaej* discussed so far, it immediately follows the verb (or is separated from it by *mbouj*) and if there is an object, it follows *ndaej*. In sentences like the following, however, *ndaej* behaves differently.

- (42) a. *Gou gwn gaemz ndeu ndaej lwi?* (based on Z&Q 27)  
 1s eat bite one ACQ Y/N-PRT  
 'Can I take a bite?'
- b. *De gangj vah Yeznanz hix ndaej.* (cf. (35))  
 3s speak language Vietnam also ACQ  
 'He can also speak Vietnamese.'

In sentences of this type, *ndaej* is to the right of all material in the sentence, including the object, except sentence final particles, if there are any. Analytically, *ndaej* is best viewed as the main predicate of the sentence, meaning 'be okay', the preceding part constituting the sentential subject, as indicated in (43) (cf. Enfield 2003 for Lao, Duffield 2003 for Vietnamese).

- (43) [<sub>S</sub> [<sub>XP</sub> Sentential subject] [<sub>Predicate</sub> ACQ]]  
 [S *gou gwn gaemz ndeu*] [*ndaej*]  
 1s eat bite one ACQ

The translations in (42) using *can* are again somewhat misleading. The sentences would be more literally translated as 'Is me taking a bite okay?' and 'That he speaks Vietnamese is also okay' respectively. In Zhuang, this is the only interpretation possible

15. The function of *bae* in sentence final position is not entirely clear: it is said to be a marker of surprise, but it is also an aspect marker. Without *bae* and without a quantitative object, these sentences are not acceptable.

for sentences in which *ndaej* is not adjacent to the verb and follows the VP as a whole (cf. (38)). I will refer to the analysis in (43) as the “Sentential Subject Analysis” or “SSA”; sentences like those in (42) are “SSA sentences”.

There are a number of arguments for analyzing sentences such as (42) as in (43). The first is placement and scope of negation. If we put *mbouj* ‘not’ right in front of *gwn* ‘eat’ in (42), it only has scope over the part that we identified as the sentential subject. In contrast, *mbouj* ‘not’ right in front of *ndaej* only negates *ndaej*:

- (44) a. *De mbouj gangj vah Yeznanz ndaej lo.*  
 3s not speak language Vietnam ACQ PRT  
 ‘It’s okay if he does not speak Vietnamese.’  
 b. *De gangj vah Yeznanz mbouj ndaej.*  
 3s speak language Vietnam not ACQ  
 ‘It is not okay if he speaks Vietnamese.’  
 c. *De mbouj gangj vah Yeznanz mbouj ndaej.*  
 3s not speak language Vietnam not ACQ  
 ‘It is not okay if he does not speak Vietnamese.’ (i.e., ‘He must . . .’)

This pattern is expected under the SSA, since the negation in the sentential subject is too deeply embedded in the subject-sentence to be able to scope over the main predicate of the sentence, *ndaej* ‘okay’.

Secondly, *ndaej* can be preceded by adverbs whose distribution is limited to the position immediately preceding the main predicate of the sentence, such as *cungj* ‘all’, *cij* ‘only then’, *coh* ‘only’, *lij* ‘still’ and *hix* ‘also’, some of which are illustrated in (45).

- (45) a. *De bae roxnaeux mwngz bae cungj ndaej.* (Z&Q 91)  
 3s go or 2s go all ACQ  
 ‘Whether he goes or you go, it’s all okay.’  
 b. *Bak gaen rap cungj ndaej, . . .* (Qin 265 (55))  
 100 pound carry all ACQ  
 ‘Lifting 100 pounds is already possible, . . .’  
 c. *De gangj vah Yeznanz {hix, lij . . .} ndaej lo.* (cf. (42b))  
 3s speak language Vietnam also, still ACQ PRT  
 ‘It’s {also, still} okay for him/her to speak Vietnamese.’

Next, the part of the sentence preceding *ndaej* can be followed by a short pause (“comma intonation”) or an overt marker, possibly a topic marker, such as *ko*:

- (46) *Mwngz mbouj gwn noh ko, mbouj ndaej.*  
 2s not eat meat TOP, not ACQ  
 ‘It is not okay for you not to eat meat.’ (more literally: ‘As to you not eating meat, it’s not okay.’)

Finally, sentences of the format in (43), that is, SSA-sentences, exist independently of *ndaej*. We also find it with other lexical elements, which do not perform the other functions of *ndaej*.



- (47) a. *Mwngz baenzneix yawj saw mbouj ndei vei.* (Qin 137 (18))  
 2s this.way read book not good PRT  
 ‘You reading this way is not good.’
- b. *Mwngz ndaq vunz mbouj baenz vei.* (Qin 137 (17))  
 2s scold people not okay PRT  
 ‘It is not okay for you to curse people.’
- c. *Bouxboux aeu gaenx guh hong cij baenz ne.* (cf. Z&Q 193 (2))  
 everyone want hard do work only okay PRT  
 ‘Only everyone wanting to work hard is okay.’ (everybody must want to . . .)

In these sentences, *ndei* ‘good’ and *baenz* ‘okay’ can be replaced by *ndaej* without any consequences. Whereas *ndaej*, in view of its many other functions, may invite other analyses, the SSA for (47a) with *ndei* ‘good’ would meet with little resistance. Their interchangeability with *ndaej* constitutes an argument for analyzing sentences like (42) as in (43).

**3.4.2.4 Conclusion.** We conclude, first, that in Zhuang the post-verbal modal ACQ ‘can’ must be adjacent to the verb; there is no modal ACQ in sentence final position. If *ndaej* occurs in sentence final position, it is always the lexical verb ‘be okay’, predicating over a sentential subject (the SSA of (43)). Secondly, we observed that we find modal *ndaej* with telic predicates only and as such we identified the kind of ‘can’ modality it expresses as what the Chinese tradition calls “potential”.

### 3.4.3 Lao

Lao differs in (at least) two respects from Zhuang. First, its post-verbal modal *daj*<sup>4</sup> ‘can’ does not immediately follow the verb; instead, it follows the VP as a whole. Secondly, post-verbal modal ACQ is compatible with atelic events and may have an ability reading. On the other hand, like Zhuang *ndaej*, Lao *daj*<sup>4</sup> in sentence final position can have the ‘okay’ reading too. As a consequence, the following sentence is ambiguous between ‘can’-ability and ‘okay’:

- (48) *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>.* (cf. Enf 102 (95))  
 3s speak language Lao ACQ  
 i. ‘S/he can speak Lao.’  
 ii. ‘It is okay for her to speak Lao.’

The question is whether this sentence is structurally ambiguous as well. There are good reasons to assume that it is. The first is that we get different readings depending on whether we insert the negative element *bò*<sup>0</sup> ‘not’ in front of *daj*<sup>4</sup> or in front of *vaw*<sup>4</sup> ‘speak’.

- (49) a. *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> bò<sup>0</sup> daj<sup>4</sup>.* (Enf 106 (110))  
 3s speak language Lao not ACQ  
 i. ‘S/he cannot speak Lao.’  
 ii. ‘It is not okay for him/her to speak Lao.’

- b. *laaw<sup>2</sup> bò<sup>0</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>.* (Enf 107 (111))  
 3s not speak language Lao ACQ  
 i. 'S/he cannot speak Lao.'  
 ii. 'It is okay for him/her to not speak Lao.'

The ability reading is preserved, wherever we put the negation (although for this reading placing the negation right in front of *daj<sup>4</sup>* is preferred); in other words, in this reading at least, *bò<sup>0</sup>* 'not' has scope over *daj<sup>4</sup>* wherever it is. Note that, when *bò<sup>0</sup>* precedes the verb it *need* not have scope over *daj<sup>4</sup>*, as we see in (49bii), but in that case we are no longer dealing with modal *daj<sup>4</sup>* but rather with *daj<sup>4</sup>* 'okay'. This is the same as when Zhuang *mbouj* 'not' is put in front of the other verb: it does not have scope over *ndaej* 'okay', which we viewed as an argument for the SSA. We conclude that the sentence in (48) is structurally ambiguous. The next question is what the underlying structures are.

That the structure corresponding to the interpretation in (48ii) involves a sentential subject, that is, is an SSA structure, can be argued in a similar way as we did for the corresponding Zhuang sentences. We will not do that here.<sup>16</sup> Discussion of the structure corresponding to the 'can'-ability reading of (48i) is postponed until §4. What will certainly have to be taken into consideration, is the fact that in the 'can'-ability reading, *daj<sup>4</sup>* can be "doubled" by an element in the left periphery of the sentence, namely by the pre-verbal modal element *saa<sup>3</sup>maat<sup>4</sup>* 'can'.

- (50) *laaw<sup>2</sup> saa<sup>3</sup>maat<sup>4</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup> bò<sup>0</sup>?* (Enf 113 (134))  
 3s can speak language Lao ACQ Q-PRT  
 'Can s/he speak Lao?'

I conclude from Enfield's (2003, 113-116) description that *saa<sup>3</sup>maat<sup>4</sup>* 'can' always operates in tandem with *daj<sup>4</sup>*. As argued by Enfield (2003, 115), *saa<sup>3</sup>maat<sup>4</sup>*'s function is that it "signposts the scope of 'can' ". *Saa<sup>3</sup>maat<sup>4</sup> . . . daj<sup>4</sup>* sentences are not ambiguous like (48) and the SSA is not applicable; *saa<sup>3</sup>maat<sup>4</sup>* 'can' and *daj<sup>4</sup>* 'can' are in some sense one whole. In negative sentences, the negation precedes *saa<sup>3</sup>maat<sup>4</sup>* and even then the SSA does not apply, unlike in (49b), where this was one of the options. We return to the doubling phenomenon in §4.

We saw that Lao modal *daj<sup>4</sup>* has the 'can'-ability reading, always in atelic contexts. It differs in this respect from Zhuang *ndaej*, which does not have this reading and is not found with atelic events: it always expresses potential modality. Does Lao *daj<sup>4</sup>* ever induce the potential reading? If, as argued in §3.4.2.2, the potential reading automatically

16. See Enfield (2003, to app.). An argument can be deduced from the distribution of the element *ka<sup>0</sup>* in such sentences as Enf 102 (95); 112 (128), (129), Enfield to app., (249) and (248). For the status of *ka<sup>0</sup>*, see Enfield 2003, 107ff; Enfield to app., sect. 3.4.2 and 4.1.6), esp. the sentences Enf 108 (116b), (117).

arises if we combine ‘can’ with telicity, this question is trivially uninteresting and the answer is ‘yes’:

- (51) a. *man<sup>2</sup> sèt<sup>1</sup> thuaj<sup>5</sup> lòn<sup>4</sup> daj<sup>4</sup>.* (both: Nick Enfield, p.c.)  
 3s wipe bowl clean ACQ  
 ‘S/he can wipe the bowls clean.’  
 b. *man<sup>2</sup> tii<sup>3</sup> pèt<sup>2</sup> too<sup>3</sup> nan<sup>4</sup> taaj<sup>3</sup> daj<sup>4</sup>.*  
 3s hit duck CL DEM.NONPROX dead ACQ  
 ‘S/he can beat that duck to death.’

The next question that comes up is: Does Lao have the potential construction? In the result sentences presented so far, modal *daj<sup>4</sup>* is always in sentence final position. In most cases, putting it anywhere else leads to oddness or ungrammaticality (Nick Enfield, p.c.). In some cases, however, *daj<sup>4</sup>* does appear in front of the result denoting predicate in which case it behaves in all respects like its Zhuang counterpart. Consider (52).

- (52) *paj<sup>3</sup> vit<sup>2</sup>*  
 go escape

Embedded in the proper sentence and context, (52) can mean ‘get/got away’ (among other readings) and (53a) would have to be translated as ‘can/could get away’. However, (53b) has an interpretation which puts more emphasis on the result and the potential of reaching it.

- (53) a. *paj<sup>3</sup> vit<sup>2</sup> daj<sup>4</sup>*  
 go escape ACQ  
 b. *paj<sup>3</sup> daj<sup>4</sup> vit<sup>2</sup>*  
 go ACQ escape

Nick Enfield (p.c.) reports that speakers find this sentence okay, though not very idiomatic. Interestingly, he notes that out of context, (53b) would mean ‘can or could/*managed to get away*’ – emphasise mine. In short, Lao also has the potential construction, though it is not as general as in Zhuang. For further discussion, see §4.<sup>17</sup>

17. For completeness sake, like Zhuang, Lao may combine its ACQ with quantity objects. An additional property is that the object may be split, as in the following example:

- i. *laaw<sup>2</sup> kin<sup>3</sup> khaw<sup>5</sup> daj<sup>4</sup> sòòng<sup>3</sup> thuaj<sup>5</sup> lèw<sup>4</sup>* (Enf 136 (203))  
 3s eat rice acq two bowl perf  
 ‘S/he has (already) eaten two bowls of rice.’

As far as I understand, this must be interpreted as that s/he has eaten rice to the extent of two bowls. Without the PERF-marker in sentence final position, the sentence may, given the right context, also mean that s/he *can* eat that amount. In this respect, the sentence is like simple Lao resultatives which, as we will see, always allow for a “potential” interpretation. *Daj<sup>4</sup>* in sentences like (i) would then have to be analyzed as the result denoting predicate, possibly on a par with

Interestingly, Enfield (2003, 121, fn. 12; and several other places) observes that “a ‘potential’ reading is possible for all V1-V2 resultative constructions”.<sup>18</sup> This means that a verb-result combination will in principle always be ambiguous between a result reading and a potential reading. Here are some examples (Nick Enfield, p.c.; cf. Enfield to app., (223), (224)), cf. (51):

- (54) a. *man*<sup>2</sup> *sət*<sup>1</sup> *thuaj*<sup>5</sup> *lòn*<sup>4</sup>.  
           3s wipe bowl clean  
           i. ‘S/he wiped the bowls clean.’  
           ii. ‘S/he can wipe the bowls clean.’  
       b. *man*<sup>2</sup> *sət*<sup>1</sup> *thuaj*<sup>5</sup> *bò*<sup>0</sup> *lòn*<sup>4</sup>.  
           3s wipe bowl neg clean  
           i. ‘S/he didn’t/doesn’t wipe the bowls clean.’  
           ii. ‘S/he can’t wipe the bowls clean.’  
       c. *man*<sup>2</sup> *tii*<sup>3</sup> *pət*<sup>2</sup> *too*<sup>3</sup> *nan*<sup>4</sup> *taaj*<sup>3</sup>.  
           3s hit duck CL DEM.NONPROX dead  
           i. ‘S/he beat that duck to death.’  
           ii. ‘S/he can beat that duck to death.’

Note that the same ambiguity is found with achievement *daj*<sup>4</sup>, which brings us back to ACQ.

- (55) *man*<sup>2</sup> *haa*<sup>3</sup> *kacèè*<sup>3</sup> *daj*<sup>4</sup>. (cf. (31a))  
           3s seek key ACQ  
           i. ‘S/he found the keys.’  
           ii. ‘S/he can look for the keys.’

The fact in (55) will be discussed at length in §4. Enfield’s observation that V1-V2 result sentences can have an extended ‘can’ reading will turn out to be very important for us; Vietnamese also has this property but Cantonese and Zhuang do not.<sup>19</sup>

In sum, *daj*<sup>4</sup> in the right periphery of the sentence is either a modal element expressing ‘can’ or it serves as the main clause predicate in an SSA sentence, meaning ‘be okay’. Lao has three different ways of expressing potential modality. One is the verb-result combination itself (see (54)), with the potential interpretation as an extension of the primary result reading. The other two (excluding (55)) involve *daj*<sup>4</sup>. First, we have V-*daj*<sup>4</sup>-R, just as in Zhuang, a potential construction. In the second, with *daj*<sup>4</sup> following

---

the achievement marker of §3.3.3; more on this in §4. Sentences like (i) do not have the exclamative reading of their Zhuang counterparts.

18. Enfield’s definition of the term “potential” is not the same as ours, but the differences are not relevant here; see also Enfield (to app.).

19. See Enfield (2003, Ch. 1, sect. 4.) for a pragmatic account of how the “potential” can be an extension of a result reading.

the result, the potential interpretation is the simple result of the combination of ‘can’ with a telic event.

In comparison to Zhuang, Lao ACQ is typically not adjacent to the verb (Zhuang ACQ is) and it may combine with atelic events, expressing ability (impossible in Zhuang).

#### 3.4.4 Vietnamese

Vietnamese provides us with a by now familiar picture: *được* in sentence final position can be a modal ‘can’ and it can also function as the main lexical predicate in SSA-type sentences, meaning ‘be okay’ – but also ‘be possible’ which sometimes makes it hard to distinguish lexical verb *được* from the epistemic use of modal *được*. However, that some sentences with *được* in sentence final position must be analysed according to the SSA (with *được* as the main predicate, the preceding part of the sentence as the sentential subject), can be argued for on similar grounds as for the corresponding cases in Zhuang and Lao above. For reasons of space, we will not spell out the arguments here (see also Duffield 2003).<sup>20</sup>

Turning to post-verbal *được* as a modal expressing ‘can’-ability, we find essentially the same situation as in Lao, except that Vietnamese *được* seems to appear felicitously in more different places in the sentence than was the case for Lao *daj*<sup>4</sup>. Consider the straightforward cases with a ‘can’-ability reading (not necessarily the only reading, as before and after):

- (56) a. *Tôi nói tiếng Lao được.* (all: cf. Enf 212 ff.)  
 1s speak language Lao ACQ  
 ‘I can speak Lao.’  
 b. *Tôi không nói tiếng Lao được.*  
 1s not speak language Lao ACQ  
 ‘I cannot speak Lao.’  
 c. *Tôi nói tiếng Lao không được.*  
 1s speak language Lao not ACQ  
 ‘I cannot speak Lao.’

The sentences in (56a) and (56b) show that, with respect to the placement of negation, Vietnamese displays the same pattern as did Lao: *không* ‘not’ has scope over *được* even when placed before *nói* ‘speak’ (in the reading given).

Just as Lao has its modal *saa<sup>3</sup>maat<sup>4</sup>* as a way to indicate the scope of *daj*<sup>4</sup>, Vietnamese has the modal *có thể* ‘can’ to serve that function for *được*. As Duffield (2003) notes: “Crucially, for every sentence in [Duffield 2003] that contains clause-final *được* – except

20. For instance, what we would call the sentential subject can be marked with a (contrastive) topic marker, *thì*, thus confirming its status as a constituent (as in Duff (21b)). Also, *được* can be preceded by tense and other markers which exclusively immediately precede the main verb of the sentence (Duff (23b); Enf 226 (154)). Finally, the SSA format exists independently of *được*, for instance with *tốt* ‘good’ in place of *được*. In all, there is enough reason to assume that at least some sentences with *được* in sentence final position can be analyzed according to the SSA.

for the sentential subject contexts – there is a completely synonymous sentence with pre-verbal *có thể* added. . . . Whenever *có thể* is present in this position, *được* is . . . wholly redundant” (emphasis added–RS). (A difference between the two languages is that whereas *saa<sup>3</sup>maat<sup>4</sup>* in Lao is hardly ever used without *daj<sup>4</sup>*, *có thể* often occurs without *được*.) To illustrate, if the sentence in (57) has the two readings given, it corresponds to two different sentences with *có thể*, as in (57b) and (57c) (see the discussion in Duffield 2003). The position of *có thể* indicates the scope of the modality.

- (57) a. *Anh ấy kiểm việc được.* (all: cf. Duff; Enf 223)  
 3s DEM look.for work ACQ  
 i. ‘S/he can find/look for a job.’ (ability)  
 ii. ‘It is possible that he will find/look for a job.’ (epistemic)
- b. *Anh ấy có thể kiểm việc (được).*  
 3s DEM can look.for work ACQ  
 i. ‘S/he can find/look for a job.’ (ability; no epistemic reading)
- c. *Có thể anh ấy kiểm việc (được).*  
 can 3s DEM look.for work ACQ  
 ii. ‘It is possible that he will find/look for a job.’ (epistemic; no ability reading)

The interaction with negation gives the same result as Enfield observed for Lao. In sentences with both *có thể* ‘can’ and *được*, if we negate *có thể* (coming out as *không thể*), the sentence can still have the modal ‘can’ interpretation, but when *không* ‘not’ is put right in front of *được*, then this is no longer possible. The following examples show this (examples from Enfield 2003; see also Duffield 2003).

- (58) a. *Tôi không thể nói tiếng Việt được.* (Enf 223 (142))  
 1s not can speak language Vietnamese ACQ  
 ‘I cannot speak Vietnamese.’
- b. *Tôi có thể đi không được.* (Enf 223 (144))  
 1s can go not ACQ  
 ‘It is possible that I can’t go.’  
 NOT: ‘I cannot go.’

In conclusion, just as was the case in Lao, Vietnamese sentences with sentence-final *được* seem to fall into two categories. Some, in which *được* means ‘be okay’ or (possibly, for some) ‘be possible’, we analyze according to the SSA. Others, with modal *được*, have to be analyzed differently – how exactly we will see in §4.

Before moving on to Cantonese, we need to look at one more aspect of *được*. Whereas Lao *daj<sup>4</sup>* ‘can’ seems to have a preference for the phrase-final position, Vietnamese *được* ‘can’ may also appear before the object. The same pattern was displayed by the achievement ACQs in Lao and Vietnamese (cf. (34)). Vietnamese sentences with *được* in between the verb and the object do not all involve the potential construction, as they are not all telic sentences (though some may be). Although in some cases we

seem to deal with free variation as in (59) (within limits; see Simpson 2001), Duffield (2003) observes that with heavy noun phrase objects, the preferred position for *được* is between the verb and the object; this is shown in (60).

- (59) a. *Tôi lái xe được.* (both: cf. Duff (43); Simpson 2001 (79))  
 1s drive car ACQ  
 b. *Tôi lái được xe.*  
 1s drive ACQ car  
 BOTH: 'I can drive.'
- (60) a. ??*Tôi lái xe đắt tiền (và) có máy lạnh được.* (Duff (41a))  
 I drive car expensive (and) have a.c. ACQ  
 b. *Tôi lái được xe đắt tiền (và) có máy lạnh.* (Duff (41b))  
 I drive ACQ car expensive (and) have a.c.  
 BOTH: 'I can drive expensive cars with air conditioning.'  
 c. *Ông ấy nói được mọi tiếng mà người ta nói ở*  
 3s DEM speak ACQ every language REL people speak in  
*bên Nam.* (Duff (42d))  
 side south  
 'He can speak every language that people speak in the south.'

Looking at the examples, I get the impression that what is relevant is not so much that the NP is heavy, but that it denotes something "big", as if in all of them some kind of quantity is involved. In any case, we taste the same achievement flavor as in the sentences in Zhuang with ACQ in between the verb and a quantity expression. Such sentences are also attested in Vietnamese:

- (61) a. *Tôi ăn được bốn bát cơm.* (Bisang 1992, 300 (47))  
 1s eat ACQ four bowl rice  
 'I can eat four bowl of rice.'  
 b. *Một ngày tôi ăn cơm được ba cân.* (Enf 158 (23))  
 one day I eat rice ACQ three kilo  
 'In one day, I can eat three kilos of rice.' (one possible interpretation)

Both Bisang and Enfield stress that despite the modality reading, the sense of attainment is present in these Vietnamese sentence, though there is no exclamative 'so much!' sense as in Zhuang.

### 3.4.5 Cantonese

Cantonese is like Zhuang. Its ACQ, *dak*<sup>1</sup>, can have the 'okay' reading in sentences analyzable according to the SSA, and post-verbal modal *dak*<sup>1</sup> only has the potential interpretation; no ability reading is available. As to 'be okay', examples like the following are common in Cantonese:

- (62) a. *lei<sup>5</sup> m<sup>4</sup> sik<sup>6</sup> m<sup>5</sup> dak<sup>1</sup> a<sup>3</sup>!*  
 2s not eat not ACQ PRT  
 'It's not okay if you don't eat.'

- b. *lei<sup>5</sup> m<sup>4</sup> heoi<sup>3</sup> dak<sup>1</sup>.*  
 2s not go ACQ  
 'It's okay if you don't go.' (NOT: 'You cannot go.')
- c. *lei<sup>5</sup> daa<sup>5</sup> maa<sup>1</sup>-mi<sup>4</sup> zek<sup>3</sup> sau<sup>2</sup> m<sup>4</sup> dak<sup>1</sup> wo<sup>3</sup>!*  
 2s hit mummy CL hand not ACQ PRT  
 'You hitting your mother's hand is not good!'
- d. *ngo<sup>5</sup> bei<sup>2</sup> di<sup>1</sup>-je<sup>5</sup> lei<sup>5</sup> zou<sup>6</sup> dak<sup>1</sup>-m<sup>4</sup>-dak<sup>1</sup>.* (Matth & Yip 1994, 28)  
 1s give CL-things you do ACQ-not-ACQ  
 'Is it okay if I give you something to do?'

From the translations it is clear that negation only has scope over *dak<sup>1</sup>*, if the negative particle immediately precedes it. Like in Zhuang an element meaning 'all', *dou<sup>1</sup>*, relating the subject to the predicate, can precede *dak<sup>1</sup>* 'be okay', as in (63a). In (63b) it is shown that a sentence with lexical *dak<sup>1</sup>* as its predicate can be interrogative. All this goes to show that we are dealing with SSA sentences.

- (63) a. *keoi<sup>5</sup> heoi<sup>3</sup>-m<sup>4</sup>-heoi<sup>3</sup> dou<sup>1</sup> dak<sup>1</sup>.*  
 3s go-not-go all ACQ  
 'It's all okay, whether he goes or not.'
- b. *keoi<sup>5</sup> heoi<sup>3</sup>-m<sup>4</sup>-heoi<sup>3</sup> dou<sup>1</sup> dak<sup>1</sup> me<sup>1</sup>?*  
 3s go-not-go all ACQ PRT  
 'Is it all okay, whether he goes or not?'

As to post-verbal modal *dak<sup>1</sup>*, it is always adjacent to the verb and it only has a potential interpretation; in other words, it occurs in the potential construction. Thus, (64) is a good sentence, the sentences in (65) are not (at least, not under the interpretations given). For discussion, see Cheng and Sybesma (2004).

- (64) *keoi<sup>5</sup> daa<sup>2</sup>-dak<sup>1</sup>-sei<sup>2</sup> zek<sup>3</sup> gat<sup>6</sup>zat<sup>2</sup>.* (Soh 2003 (7b))  
 3s hit-ACQ-dead CL cockroach  
 'He can kill the cockroach.'
- (65) a. *\*keoi<sup>5</sup> gong<sup>2</sup> {dak<sup>1</sup>} Jing<sup>1</sup>-man<sup>2</sup> {dak<sup>1</sup>}.*  
 3s speak ACQ English ACQ  
 INTENDED: 'He is able to speak English.'
- b. *\*keoi<sup>5</sup> zaa<sup>1</sup> {dak<sup>1</sup>} (li<sup>1</sup>-gaa<sup>3</sup>) ce<sup>1</sup> {dak<sup>1</sup>}.*  
 3s drive ACQ this-CL car ACQ  
 INTENDED: 'He is able to drive (this car).'

Just like Zhuang, Cantonese has its set of apparent counterexamples to the claims that *dak<sup>1</sup>* is invariably associated with telics and can only have a potential interpretation. Here is one such case:

- (66) *keoi<sup>5</sup> tiu<sup>3</sup>-dak<sup>1</sup>.*  
 3s jump-ACQ  
 'He can jump.'

This sentence has no resultative phrase and still has a 'can'-reading. We have to realize, however, that *tiu<sup>3</sup>* 'jump' may not have an overt result denoting predicate, but it



does have a typical, projected end point: up in the air; in other words, it is telic. In this respect, a verb like *tiu*<sup>3</sup> ‘jump’ is different from *gong*<sup>2</sup> ‘speak’ and *zaa*<sup>1</sup> ‘drive’ in (65), which have no projected, built-in end points (they are a-telic). As discussed in Cheng and Sybesma (2004), native speakers consulted do seem to concentrate on the end point, “up”, though with a verb like jump, it is hard to see how it could be otherwise. But, importantly, when prompted by (66), they do not get an image of a continuous repetitive jumping action, up and down, up and down. It is one jump. Two more facts are relevant here. First, if we add the unspecific directional *yuk*<sup>1</sup> to (66), we get a sentence which is semantically hardly distinguishable from it.<sup>21</sup>

- (67) *keoi*<sup>5</sup> *tiu*<sup>3</sup>-*dak*<sup>1</sup>-*yuk*<sup>1</sup>.  
 3s jump-ACQ-DIR  
 ‘He can jump (up).’

Secondly, verbs that do not have a typical projected end point cannot be used with post-verbal modal *dak*<sup>1</sup>. We saw examples in (65), but I give two more examples here, with intransitive verbs: *siu*<sup>3</sup> ‘laugh’ and *haam*<sup>3</sup> ‘cry’, which provide a more minimal pair with (66). In the interpretation of ‘can laugh’ and ‘can cry’, the following sentences are out. The minimal difference with *tiu*<sup>3</sup> ‘jump’ is the lack of telicity.

- (68) a. \**keoi*<sup>5</sup> *siu*<sup>3</sup>-*dak*<sup>1</sup>.  
 3s laugh-ACQ  
 b. \**keoi* *haam*<sup>3</sup>-*dak*<sup>1</sup>.  
 3s cry-ACQ

This confirms what we found when looking at Zhuang: in languages like Zhuang and Cantonese an apparent ability reading is derived from the potential; the generalizations with respect to distribution and interpretation of post-verbal modal ACQ stay in tact. The question is, when we say “languages like Zhuang and Cantonese”, what kind of languages are we talking about? We investigate this problem in §4.<sup>22</sup>

21. Adding *hei*<sup>2</sup> ‘up’ is also possible.

22. Cantonese also has sentences in which *dak*<sup>1</sup> is followed by a quantity expression. Their interpretation is very different from the sentences that look similar in the other languages.

- i. a. *keoi*<sup>5</sup> *tai*<sup>2</sup>-*dak*<sup>1</sup> *saam*<sup>1</sup>-*bun*<sup>2</sup> *syu*<sup>1</sup> (Tang 2002 (4))  
 3s read-ACQ three-CL book  
 ‘s/he only read three books’  
 b. *sing*<sup>3</sup>-*daan*<sup>3</sup>-*lou*<sup>5</sup>-*jan*<sup>4</sup> *sung*<sup>3</sup>-*dak*<sup>1</sup> *leung*<sup>5</sup>-*fan*<sup>6</sup> *lai*<sup>5</sup>-*mat*<sup>6</sup> *bei*<sup>2</sup> *ngo*<sup>5</sup>  
 Santa Claus give-ACQ two-CL gift to 1s  
 ‘Santa Claus only gave me two presents’ (Tang 2002 (9a))

While in Zhuang the combination of ACQ and a quantified objects leads to an exclamation (‘so much!’), in Cantonese we get a restrictive interpretation: ‘only this much’. See Tang (2002) for discussion.

### 3.4.6 Conclusion

Summarizing this section on ACQ 'okay' and post-verbal modal ACQ 'can', we observe that in all four languages, some sentences with ACQ in sentence final position are analyzable according to the SSA, involving a sentential subject which is predicated on by lexical verb ACQ 'be okay'. Importantly, for Cantonese and Zhuang, this is the only possible analysis of sentences with ACQ not adjacent to the verb. For Lao and Vietnamese on the other hand, such sentences are structurally ambiguous: besides the SSA 'okay' structure, there is another, with modal ACQ, having a 'can' ability reading.

Although all four languages have ACQ 'can' in post-verbal position, they differ in several respects. First, in Cantonese and Zhuang ACQ is obligatorily adjacent to the verb and never follows the object, which is not the case in Lao and Vietnamese. Secondly, in Cantonese and Zhuang we never find ACQ in atelic contexts, whereas their counterpart in the other two languages is not restricted in that way. As a result, in Cantonese and Zhuang, post-verbal modal ACQ only has a potential interpretation, while this is not true for Lao and Vietnamese.

## 3.5 Zhuang between Lao and Cantonese

The factual overview presented in this section shows how versatile and pervasive the ACQ phenomenon is – and the overview is not even complete for any of the four languages individually. The overview shows the similarities in the use of the ACQ elements across the languages. However, it also shows that there are differences as well.

When we look at the differences, we observe that, while having more in common with Lao than with Cantonese with respect to the pre-verbal use of ACQ, Zhuang is very similar to Cantonese when it comes to the post-verbal use. One property, mentioned several times, that set Cantonese and Zhuang apart from Lao (and Vietnamese) is the placement and distribution of ACQ (other than main verb ACQ 'be okay'): it is obligatorily adjacent to the verb and restricted to telic predicates in the former and not in the latter.

In the following section we will see whether both differences are related, as we analyze the sentence structures involving modal ACQ and try to find out why Zhuang cannot have both 'can'-ability and 'can'-potential.

## 4. Variation in potential and ability: analysis and explanation<sup>23</sup>

In the previous section, we found different ACQs in "post-verbal" position: lexical verb 'be okay', and modal ACQ, with a 'can'-potential or a 'can'-ability reading. We saw that ACQ 'okay', as a non-modal, has to be treated differently from modal ACQ. In particular,

---

23. I am aware of other analyses proposed for sentences of this type, particularly Duffield (1998; 2003) and Simpson (2001). For reasons of space, I cannot discuss these proposals here. They differ in almost all respects from my analysis presented here.

in sentences of this type ACQ was analyzed as the main predicate of the sentence, with the rest of the sentence as the sentential subject ((69), cf. (43)); we will not repeat the arguments for this analysis here.

(69) [<sub>S</sub> [<sub>Sentential Subject</sub> XP ] ACQ ]

For the rest of this paper we concentrate on the sentences with the two modal  $\text{ACQS}$  and pose the following questions, to be answered in §4.1–2 and §4.3 respectively:

- I. How do we analyze sentences with the different post-verbal modal ACQs?
- II. How do we explain the variation among the languages included in this study?

### 4.1 Analysis: the potential construction

ACQ in the potential construction has three properties we need to account for: (i) it occurs in post-verbal position, which is not the place where we find the other modal verbs in the languages under consideration; (ii) it is adjacent to the verb; (iii) it is restricted to telic contexts, especially resultatives; and (iv) its scope is limited to the result.

I adopt the analysis proposed in Cheng and Sybesma (2003; 2004) for the potential construction in Cantonese and Mandarin (but see the discussion in §5). This analysis assumes Hoekstra's view that in resultative sentences, the result denoting part is a small clause, which as a whole is the complement to the verb (Hoekstra 1988; 2005; Sybesma 1999). This means that a resultative sentence consists of two clauses, a main clause and, fully embedded in it, a result denoting small clause. Consider (70) for an analysis of a Cantonese resultative sentence, (70b) representing the base for (70a).

- (70) a. *keoi*<sup>5</sup> *lo*<sup>2</sup>-*hei*<sup>2</sup>-*zo*<sup>2</sup> *li*<sup>1</sup>-*seung*<sup>1</sup> *syu*<sup>1</sup>. (cf. (2c))  
 3s take-up-ZO this-box book  
 'S/he lifted this box of books.'  
 b. [<sub>VP</sub> [<sub>V</sub><sup>0</sup> *lo*<sup>2</sup>] [<sub>SC/AspP</sub> [<sub>Asp</sub><sup>0</sup> *zo*<sup>2</sup>] [<sub>XP</sub> [*li*<sup>1</sup>-*seung*<sup>1</sup> *syu*<sup>1</sup> [<sub>X</sub><sup>0</sup> *hei*<sup>2</sup>]]]]]  
           take                                  ZO          this-box book up

The result small clause is a relatively simple subject-predicate combination; the nominal phrase that is generally regarded as the object is really the subject of the small clause. Essentially, a small clause differs from a full clause in that it lacks an independent tense projection; its temporal reference is determined by the matrix clause. Small clauses do contain other functional projections, though, such as the one labeled “AspP” in (70). Asp<sup>0</sup> is the position where it is indicated whether the result has been realized or not: if it is, the node is occupied by “perfective” *zo*<sup>2</sup> (the Cantonese counterpart of Mandarin verb-*le*), otherwise it is left empty.<sup>24</sup>

24. That is why Sybesma (1999) calls it RealP (for “RealizationP”), but we will simply call it AspP here.

Note that, here, and below, we do not go into derivational details, for instance, how we derive (70a) on the basis of (70b). The bracketed schemas given are simple representations of some of the more basic relations; as such they are not meant to express surface structure or structural variation.

Cheng and Sybesma (2004) propose that a resultative small clause may also contain a position for a modal verb, particularly ACQ-potential. In (71b) I present a structural representation of the relevant parts of a sentence with ACQ-potential, (71a); it is (70b) with an ModP added on top of AspP. Mod<sup>0</sup> is occupied by ACQ-potential. Asp<sup>0</sup> is empty since the result has not been realized.

- (71) a. *keoi<sup>5</sup> lo<sup>2</sup>-dak<sup>1</sup>-hei<sup>2</sup> li<sup>1</sup>-seung<sup>1</sup> syu<sup>1</sup>.* (= (2c))  
 3s take-DAK-up this-box book  
 'S/he can lift this box of books.'
- b. [<sub>VP</sub> [<sub>VP</sub> *lo<sup>2</sup>*] [<sub>ModP</sub> [<sub>Mod<sup>0</sup></sub> *dak<sup>1</sup>*] [<sub>SC/AspP</sub> [<sub>Asp<sup>0</sup></sub>  $\emptyset$ ] [<sub>xP</sub> *li<sup>1</sup>-seung<sup>1</sup> syu<sup>1</sup>*  
 take ACQ this-box book  
 [<sub>x<sup>0</sup></sub> *hei<sup>2</sup>*]]]]]  
 up

This analysis accounts for the four properties just mentioned: (i) ACQ occurs in post-verbal position: it is generated there; (ii) it is adjacent to the verb: the position ACQ occupies is the position for modals in the small clause and such a position is necessarily high in the structure of a clause (but see §4.3); (iii) we only find it with resultatives: it occupies a position in a resultative small clause; and (iv) its scope: it has scope over the part of the sentence that it c-commands.

With respect to (iii): there are good reasons for representing all telics as involving a small clause, denoting the end point of the activity expressed by the matrix verb, with a phonologically empty X<sup>0</sup> (Hoekstra 2005). This means that telic sentences such as (39)–(41) would technically have to be analysed as in (70b).

There is one property of ACQ-potential sentences that we have left undiscussed. It has to do with the scope of ACQ in potential sentences, which is more complicated than I have acknowledged so far. It seems without doubt that ACQ's modality only has scope over the result part of the sentence. In other respects, however, particularly the quantificational properties that modals generally have, it induces effects that are the same as modals in the canonical, pre-verbal modal position. This can (among other things) be deduced from the effect *dak<sup>1</sup>* has on the interpretation of the question word *dim<sup>2</sup>* 'how' (Cheng and Sybesma 2003).

The question word *dim<sup>2</sup>* 'how' can be interpreted in two different ways: as 'how, by what means', the method reading, and as a rhetorical 'how could' or 'why (the hell) would', the rhetorical reading; the following sentences illustrate this (sentences (72)–(74) from Cheng and Sybesma 2003).<sup>25</sup>

25. The sentence final particles *le<sup>1</sup>* and *gaa<sup>3</sup>* in (72)–(74) are incompatible with the method and rhetorical reading respectively. It is not the case that the particles are responsible for the

- (72) a. *keoi<sup>5</sup> dim<sup>2</sup> lo<sup>2</sup>-hei<sup>2</sup> li<sup>1</sup> seung<sup>1</sup> syu<sup>1</sup> le<sup>1</sup>/\*gaa<sup>3</sup>?*  
 3s HOW take-up this box books SFP  
 RETHORICAL: 'Why the hell would s/he lift this box of books?'  
 b. *keoi<sup>5</sup> dim<sup>2</sup> lo<sup>2</sup>-hei<sup>2</sup> li<sup>1</sup> seung<sup>1</sup> syu<sup>1</sup> \*le<sup>1</sup>/gaa<sup>3</sup>?*  
 3s HOW take-up this box books SFP  
 METHOD: 'How could s/he lift this box of books?'

Using different adverbials and modal verbs, Tsai (1999) shows for the Mandarin counterpart of *dim<sup>2</sup>*, *zěnmě*, which displays a similar behavior, that the reading depends on the level of adjunction. For Cantonese this means: at or above IP we get the rhetorical reading, under IP, say, at VP, we get the method reading. The following sentences show this.

- (73) a. *keoi<sup>5</sup> dim<sup>2</sup> sei<sup>1</sup>-jiu<sup>3</sup> jat<sup>6</sup>-jat<sup>6</sup> du<sup>1</sup> sai<sup>2</sup> ce<sup>1</sup> le<sup>1</sup>/gaa<sup>3</sup>?*  
 3s HOW must daily all wash car SFP  
 RETHORICAL ONLY: 'Why the hell does he have to do daily car washing?'  
 b. *keoi<sup>5</sup> sei<sup>1</sup>-jiu<sup>3</sup> jat<sup>6</sup>-jat<sup>6</sup> du<sup>1</sup> dim<sup>2</sup> sai<sup>2</sup> ce<sup>1</sup> \*le<sup>1</sup>/gaa<sup>3</sup>?*  
 3s must daily all HOW wash car SFP  
 METHOD: 'How does he have to do daily car washing?'

When we put *dim<sup>2</sup>* 'how' in a *dak<sup>1</sup>*-sentence, only one reading is available, the rhetorical one:

- (74) *keoi<sup>5</sup> dim<sup>2</sup> lo<sup>2</sup>-dak<sup>1</sup>-hei<sup>2</sup> li<sup>1</sup> seung<sup>1</sup> syu<sup>1</sup> {le<sup>1</sup>/\*gaa<sup>3</sup>, \*le<sup>1</sup>/gaa<sup>3</sup>}?*  
 3s HOW take-DAK-up this box book SFP  
 RETHORICAL ONLY: 'Why the hell would he be able to lift this box of books?'  
 (le<sup>1</sup>/\*gaa<sup>3</sup>); NOT: 'How would he manage to lift this box of books?' (\*le<sup>1</sup>/gaa<sup>3</sup>)

In view of the facts in (72), where we had an ambiguous sentence (modulo the sentence final particle), we would expect to find ambiguity here as well, but we do not. The only reading we get is the one that corresponds to a structure in which *dim<sup>2</sup>* has been adjoined to IP or higher. In other words, in a *dak<sup>1</sup>*-sentence, adjunction to VP is impossible. The question is why that should be the case. The reason cannot be semantic or pragmatic as 'how' and 'can' are generally not incompatible. Cheng and Sybesma (2003) propose that we can understand (74) if we assume that in a sentence with potential *dak<sup>1</sup>* we always have a modal element in the canonical, pre-verbal modal position higher up in the sentence and that certain types of constituents, such as question words, may not intervene between them.

In Cantonese, the modal element in the pre-verbal modal position is phonologically empty. In Mandarin, however, ACQ, *de*, is more often than not doubled by pre-verbal *néng* 'can'. Thus, while (75a) is fine, (75b) is more natural (see Cheng and Sybesma 2003).

---

readings we get. This is clear in (73) and (74) when we see that certain readings are impossible, whatever the sentence final particle is.

- (75) a. *tā ná-de-chū-lái ma?*  
 3s take-ACQ-out-come Q-PRT  
 b. *tā néng ná-de-chū-lái ma?*  
 3s can take-ACQ-out-come Q-PRT  
 BOTH: 'Can he (i.e., will he manage to) take it out?'

Mandarin confirms our findings of the intervention effects in Cantonese. Wu (2004) has several sentence pairs showing such effects. The following set is based on data provided there.

- (76) a. *Lǐ Sī néng zhǐ kǎn-dǎo zhè-kè shù.*  
 Lǐ Sī can only chop-down this-CL tree  
 'Li Si can only chop this tree down.'  
 b. \**Lǐ Sī néng zhǐ kǎn-de-dǎo zhè-kè shù.*  
 Lǐ Sī can only chop-ACQ-down this-CL tree  
 c. *Lǐ Sī zhǐ néng kǎn-de-dǎo zhè-kè shù.*  
 Lǐ Sī only can chop-ACQ-down this-CL tree  
 'Li Si can only chop this tree down.'

Whereas the adverbial *zhǐ* 'only' can intervene between *néng* 'can' and the V-R combination, it cannot come between *néng* 'can' and ACQ-potential. As (76c) shows, there is no incompatibility between *zhǐ* 'only' and the potential construction.

Let's conclude that post-verbal elements with a modal interpretation have to be licensed, one way or another, by a modal element in the canonical modal position.<sup>26</sup> Apparently, the latter modal element may be covert, as it is in Cantonese and Zhuang; in neither languages do we have a modal like Mandarin *néng* 'can' that can double ACQ. Cheng and Sybesma (2003) argue that *néng* 'can' in Mandarin is close to obligatory because *de* 'ACQ' lost its modal force and must be supported or licensed by a fully operative modal verb. If this is correct, it amounts to saying that the potential ACQs in Zhuang and Cantonese have full modal force, which leads us to conclude that ACQ-potential is a modal verb.

In sum, if we assume that ACQ-potential occupies a modal position in the resultative small clause but is quantificationally linked to the matrix modal position, we account for most of its properties: all its scopal properties, its occurrence with resultatives, and its post-verbal position. Its adjacency to the verb may also be explained if the position ACQ occupies is the position for modals in the small clause and if such a position is necessarily high in the structure of the small clause. The adjacency issue may be more involved, however, and we return to it in §4.3. Finally, we concluded that ACQ-potential

26. These findings, leading to the conclusion that the post-verbal modals have a partner in the more typical position for modals, fit in with more general theories on sentence structure, which hold that modal verbs occupy a position in the functional domain of the sentence above VP.

is a modal verb. Presumably, this analysis applies to the potential construction in all languages.<sup>27</sup>

## 4.2 Analysis: ‘ability’ and other cases

### 4.2.1 Introduction

In this section we deal with the following three cases of post-verbal ACQ ‘can’, illustrated in (77) using Lao examples:

- Case I. ACQ with atelic events, illustrated in (77a) and (77cii), with an ability reading;
- Case II. ACQ with telic events, with a separate result denoting element, as in (77b); and
- Case III. ACQ with telic events, with no separate result denoting element, as in (77ciii).

- (77) a. *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>.* (cf. Enf 102 (95); cf. (48) above)  
 3s speak language Lao ACQ  
 ‘S/he can speak Lao.’
- b. *man<sup>2</sup> sèt<sup>1</sup> thua<sup>j</sup><sup>5</sup> lòn<sup>4</sup> daj<sup>4</sup>.* (= (51a))  
 3s wipe bowl clean ACQ  
 ‘S/he can wipe the bowls clean.’
- c. *man<sup>2</sup> haa<sup>3</sup> kacèè<sup>3</sup> daj<sup>4</sup>.* (Enf 99 (85); cf. (31a) and (55) above)  
 3s seek key ACQ  
 i. ‘S/he found the keys.’ (a case of ACQ-achievement)  
 ii. ‘S/he can look for the keys.’  
 iii. ‘S/he can find the keys.’

27. This analysis of the different ACQ-sentences presented here fits into the further development of the idea (yet to be articulated in a unified and precise way), that within VP we may find elements responsible for, or, at least, involved in, the expression of such categories as tense and aspect, and, possibly, within certain limits, modality. We find these ideas expressed explicitly in Travis (1991), where it is argued for an “Aspect2” projection within VP (in the context of Tagalog), as well as in some of my own work on resultatives and the so-called aspectual markers in Mandarin (Sybesma 1997; 1999; Sybesma and Vanden Wyngaerd 1997) and tense (Sybesma 2004). The general idea is that not all the temporal-aspectual information is encoded at the CP/IP level of the sentence. Some information, especially in as far as it is related to expressing the realization and completion (or non-completion) of the event, is encoded in layers of the sentence that are embedded in the VP. Tense-related categories in the CP/IP domain of the sentence are mainly concerned with anchoring the temporal reference of the sentence to the outside world (Enç 1987; Guéron and Hoekstra 1989; 1995; Klein 1994). All nodes/elements relevant for the temporal interpretation of the sentence are supposed to be in a configurational relation such that they form a chain, enabling the separate links to license one another consecutively, thus making temporal interpretation of (the different parts of) the sentence possible.

We take into account two additional facts mentioned in §3: (i) the signpost doubling phenomenon; and (ii) the fact that ACQ can be negated by putting the negation element in front of the main verb or in front of ACQ itself.

An observation that is important for our analysis is that the positioning of post-verbal modal ACQ in Lao and Vietnamese displays a striking parallel with the placement of ACQ-achievement and result denoting predicates in these languages. In Lao, all these elements generally follow the object: [V - O - R/ACQ] (but see (32) above plus discussion). The placement of ACQ-achievement and ACQ-ability was illustrated in (77c); Lao resultatives are given in (78), showing the same pattern.

- (78) a. *man<sup>2</sup> sê<sup>1</sup> thua<sup>5</sup> lòn<sup>4</sup>.* (Enfield, p.c.; cf. (54a) above)  
3s wipe bowl clean  
‘S/he wiped the bowls clean.’  
b. *laaw<sup>2</sup> ñing<sup>2</sup> nok<sup>1</sup> taaj<sup>3</sup>.* (Enfield to app. (197))  
3g shoot bird dead  
‘S/he shot a bird dead.’

Vietnamese, as we already saw, shows a greater degree of flexibility: we have seen facts showing that ACQ can be put right behind the verb or in a position following the object. The examples in (79b,c) show that resultatives display the same flexibility:<sup>28</sup>

#### VIETNAMESE: RESULTATIVES

- (79) a. *Xin anh đem cái va-li này đi.* (Thompson 1965, 343)  
please 2s take CL suitcase this away  
‘Please take this suitcase away.’  
b. *Nó giặt quần áo sạch.* (Duffield, p.c.)  
3s wash clothes clean  
‘S/he washed the clothes clean.’  
c. *Nó giặt sạch quần áo.* (Duffield, p.c.)  
3s wash clean clothes  
‘S/he washed the clothes clean.’

#### VIETNAMESE ACQ: ACHIEVEMENT AND ABILITY

- (80) a. *Tôi kiếm việc được.* (both: Duff (14), cf. (34), (59) above)  
1s look.for work ACQ  
b. *Tôi kiếm được việc.*  
1s look.for ACQ work  
BOTH: i. ‘I found work.’ (achievement)  
ii. ‘I can look for work.’ (ability)

28. Nigel Duffield (p.c.) informs me that not every type of result denoting predicate (or secondary predicate more generally) can be put in both positions. This has to be looked into.



In short, in Lao and Vietnamese, the distributional properties of post-verbal modal ACQ are the same as those of ACQ-ability and result denoting predicates.

Let us now turn to an analysis of the Cases in I-III, starting with III.

#### 4.2.2 Case III, ex. (77ciii)

Achievement markers have been argued to be quite similar to result denoting elements (Sybesma 1997). In that light, the parallel distribution of ACQ-achievement and result denoting predicates we observe in Lao and Vietnamese is not surprising. Technically, they would all occupy the same position in the structure,  $X^0$ , the predicate position in the result small clause. This is shown in (81), which is identical to (70b), the proposed base order of result sentences. The lexical items in the slots are from (78a) and (77ci).

- (81) [<sub>VP</sub> [<sub>V0</sub> *sê*<sup>1</sup>] [<sub>SC/AspP</sub> [<sub>Asp0</sub> . . . ] [<sub>XP</sub> *thua*<sup>5</sup> [<sub>X0</sub> *lòn*<sup>4</sup> ]]]]  
 wipe bowl clean  
 [<sub>VP</sub> [<sub>V0</sub> *haa*<sup>3</sup>] [<sub>SC/AspP</sub> [<sub>Asp0</sub> . . . ] [<sub>XP</sub> *kacè*<sup>13</sup> [<sub>X0</sub> *daj*<sup>4</sup> ]]]]  
 look.for key ACQ

What this structure expresses is that there is an event, of wiping or of seeking, and the projected result of the event of the bowl being clean or the key having turned up.

Whether or not the result has been reached (“realized”) depends, in principle, on what is in “Asp<sup>0</sup>”, which (as explained above) is the place to express whether the end point has actually been reached or not. In Sinitic languages, when the end point is reached, Asp<sup>0</sup> must be filled with an aspect marker, such as Cantonese *zo<sup>2</sup>* in (70) (Sybesma and Vanden Wyngaerd 1997); in other words, in Sinitic the realization of the end point must be marked overtly.

Apparently, Lao and Vietnamese lack such requirement. For the sake of consistency, let's say that the position is filled in (78a) and (77ci) since the projected end point was reached, but it is filled covertly. Note that the marker can be overt too, as the following Lao sentences show (cf. Enfield 2003, 122 (163)).

- (82) a. *man*<sup>2</sup> *sêt*<sup>1</sup> *thuaj*<sup>5</sup> *lòn*<sup>4</sup> *lèw*<sup>4</sup>.  
3s wipe bowl clean PRF  
'S/he wiped the bowls clean.'<sup>29</sup>  
b. *man*<sup>2</sup> *haa*<sup>3</sup> *kacè*<sup>3</sup> *daj*<sup>4</sup> *lèw*<sup>4</sup>.  
3s seek key ACQ PRF  
'S/he found the keys.' (plus: see fn. 29)

The schematic structure underlying these sentences would be the same as (81), with Asp<sup>0</sup> filled by the overt perfective marker, *lèèw*<sup>4</sup> (cf. (70b)).

29. Another reading is the one in which *lèw<sup>4</sup>* is paraphrasable as ‘it is now the case that . . .’, the original meaning to be filled in on the dots, whether it is the ‘can’ reading or the attainment reading. See Enfield (2003, 125, 135 among other places).



I propose that ACQ in Lao does the same in atelic sentences such as (77a). The literal paraphrase of the base structure of that sentence, given in (85), may be something like 'her/him speaking Lao has realized'. As in the Case III sentences, the 'can'-interpretation is an extension of this resultative base. The proposed structure of this sentence would be the following; it is just like what we proposed for the Case III sentences, that is, (81), with no AspP or an empty AspP.

- (85) [VP [<sub>V</sub><sup>0</sup> *vaw*<sup>4</sup>] [<sub>SC/AspP</sub> [<sub>Asp</sub><sup>0</sup> . . . ] [<sub>XP</sub> *phaa*<sup>2</sup>*saa*<sup>3</sup>*laaw*<sup>2</sup> [<sub>X<sup>0</sup></sub> *daj*<sup>4</sup> ]]]]  
speak language Lao ACQ

Note that Case I sentences do not have the primary achievement/realization reading. It seems that the achievement/realization structure is utilized, not to express the primary reading, but to express the secondary, the somehow “parasitic” modality reading only! Note however, that such scenario is only possible thanks to the existence of Case III structures, which are due to the conspiracy of two factors: (i) the availability of ACQ as R and (ii) the fact that result structures may be extended to a ‘can’ interpretation.

This means that in Case I too, ACQ is not a modal element. In both Case III and Case I, ACQ is analyzed as R. I will refer to this the “ACQ = R-analysis”.

#### 4.2.4 Case II, ex. (77b)

Case II sentences involve telic events, though with a separate element, not ACQ, functioning as R (such as *lòn'n* 'clean' in (77b)). On the basis of the research done so far, there are two possible analyses, both in the spirit of the proposed analysis of Cases I and III, in the sense that the 'can' reading is seen as an extension of a primary base reading, which itself may or may not be there. The first possible analysis treats ACQ as R, in this case predicating of the entire small clause; in all other respects there is no difference with the analysis of Case III and Case I (represented in (86a)).

The second possible analysis (which I favor) puts ACQ in the Asp<sup>0</sup> position (or a similar position), indicating, in its primary, base reading, that the resulting state, and, as a consequence, the event as a whole has been achieved (“ACQ = Asp<sup>0</sup> analysis”). This is represented in (86b). Note that in all respects this structure is the same as in (83); the only difference lies in the fact that in (83) perfective marker *lèw*<sup>4</sup> is in Asp<sup>0</sup> while in (86b) *daj*<sup>4</sup> is.

- (86) a. [<sub>VP</sub> [<sub>V0</sub> sê<sup>t1</sup>] [<sub>SC/AspP</sub> [<sub>Asp0</sub> . . . ] [<sub>XP</sub> *thuaj*<sup>5</sup> lòn<sup>4</sup> [<sub>X0</sub> *daj*<sup>4</sup> ]]]]  
wipe bowl clean ACQ  
b. [<sub>VP</sub> [<sub>V0</sub> sê<sup>t1</sup>] [<sub>SC/AspP</sub> [<sub>Asp0</sub> *daj*<sup>4</sup> ]] [<sub>XP</sub> *thuaj*<sup>5</sup> [<sub>X0</sub> lòn<sup>4</sup> ]]]]  
wipe ACQ bowl clean

Once again, the primary structure based reading is not there; we only have the secondary reading that is piggybacking on the base structure.

Do we have any evidence that *daj*<sup>4</sup> in Case II sentences is in Asp<sup>0</sup>? When we consider the structures proposed for the potential construction in Zhuang and Cantonese

in (71b), repeated here in (87), why wouldn't it be in  $\text{Mod}^0$ , just like its Zhuang and Cantonese counterparts?

- (87)  $[_{VP} [_{V^0} lo^2] [_{ModP} [_{Mod^0} dak^1] [_{SC/AspP} [_{Asp^0} \emptyset] [_{XP} li^1-seung^1 syu^1$   
take ACQ this-box book  
 $[_{X^0} hei^2]]]]]$   
up

First of all, as we have seen, in Lao, we also have the potential construction, the exact parallel in all respects of the potential constructions in Zhuang and Cantonese. Secondly, the surface distribution of modal ACQ in Case II sentences in Lao has the same distribution as the aspectual particles such as *lèew*<sup>4</sup>: they all follow the other constituents in the sentence. This would, given our assumptions regarding the ModP in the small clause, not be the case for elements in  $\text{Mod}^0$ .<sup>30</sup>

In conclusion, if all this is right, then, as in the other Cases, ACQ in Case II sentences is not a modal verb.

#### 4.2.5 Doubling and negation<sup>31</sup>

The relationship between ACQ-ability and the modal verb in the higher regions of the sentence, *saa<sup>3</sup>maat<sup>4</sup>* 'can' in Lao and *có thể* 'can' in Vietnamese, is similar to what we found in the potential sentences in Cantonese, Mandarin, and, presumably, Zhuang. Modal expressions consist of, at least, two parts, a lexical modal meaning and a quantificational part. Wherever the modal meaning part is, the quantificational part has to be expressed through some element in the higher functional layers of the sentence. As before, this element can be overt or covert. Since ACQ-ability, unlike ACQ-potential, does not occupy a modal position, I would expect doubling to be more frequent than is the case for ACQ-potential, but we do not have enough data to verify this expectation (see Enfield 2003, 225ff. for a report on his attempt to gather some). Also, as noted, in Lao, *saa<sup>3</sup>maat<sup>4</sup>*, not *daj<sup>4</sup>*, is the element that is optional.

Next, in ACQ-ability sentences, negation can be placed in two different positions, with no obvious effect on the interpretation of the sentence: apparently in front of the main verb as in (88a), or right in front of ACQ, see (88b). These are Lao examples, the same applies to Vietnamese.

30. Another question is: how do we know that ACQ in Zhuang and Cantonese is not in  $\text{Asp}^0$ ? After all, in our analysis, in the presence of ACQ,  $\text{Asp}^0$  is always empty. Here too, in view of the fact that the distributional properties of ACQ and those of the aspectual particles are different, they cannot be the same.

31. This section is about Case I and Case III sentences only, as I do not have enough data to say anything worthwhile on Case II. More research on the interaction between Case II sentences and the placement of negation will possibly help us decide whether ACQ in Case II sentences must be analysed as R or as  $\text{Asp}^0$ .

- (88) a. *laaw<sup>2</sup> bò<sup>0</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> daj<sup>4</sup>.* (Enf 107 (111); cf. (49b) above)  
           3s   not speak language Lao   ACQ  
       b. *laaw<sup>2</sup> vaw<sup>4</sup> phaa<sup>2</sup>saa<sup>3</sup> laaw<sup>2</sup> bò<sup>0</sup> daj<sup>4</sup>.* (Enf 106 (110); cf. (49a) above)  
           3s   speak language Lao   not   ACQ  
           BOTH: 'S/he cannot speak Lao.'

As to (88a), I suggest that the negation is not directly in front of the verb, but, rather, in front of the modal element there. This is confirmed if we make the modal element overt: then the negation is in front of *saa<sup>3</sup>maat<sup>4</sup>* (or *(cò) thê<sup>2</sup>* in Vietnamese).

As to the placement of negation in (88b), we see that here too, the pattern is the same as with other secondary predicates such as resultatives, thus confirming our analysis. The negative counterpart of (78a) is in (89a), that of (77c) in (89b).

- (89) a. *man<sup>2</sup> sêt<sup>1</sup> thuaj<sup>5</sup> bò<sup>0</sup> lòn<sup>4</sup>.* (both: Nick Enfield, p.c.)  
           3s   wipe bowl not clean  
           'S/he didn't wipe the bowls clean.'  
       b. *man<sup>2</sup> haa<sup>3</sup> kacèè<sup>3</sup> bò<sup>0</sup> daj<sup>4</sup>.*  
           3s   seek key not ACQ  
           'S/he didn't find the keys.'

In resultative and achievement cases, the paraphrase would amount to saying that one went through the motions towards a certain goal, but with no success. This is even true of the following Lao sentence:

- (90) *khaa<sup>5</sup> bò<sup>0</sup> taaj<sup>3</sup>.* (Enfield to app. (228b))  
           kill not die  
           'I didn't manage to kill it.'

It can be paraphrased as: 'I went through the motions of killing but had no success.' The paraphrase of sentences such as (88b) is different, since we are not dealing with a resultative structure (we would have something like: 'him/her speaking Lao did not materialize'), but the relation between the different parts of the sentence would be the same.<sup>32</sup>

#### 4.2.6 Conclusion

In this section we looked at three different uses of modal ACQ, other than the potential. We concluded that in all cases, ACQ is not itself a modal verb. In two, possibly all three, cases, it is the predicate of the resultative small clause, in one case it may be an aspectual marker. In all cases, the modal reading is really the secondary reading, piggybacking on the primary result reading, a possibility that V-R combinations

32. There is one thing I don't think falls out from the analysis presented here, namely the fact that, at least in Lao, all post-verbal ACQs, including 'it's okay', *daj<sup>4</sup>* is the "head of the construction for yes-answer purposes" as Enfield (2003, 102) phrases it. (See also Simpson 2001.)

in Lao and Vietnamese always have. In Case I, this is straightforward as the primary result reading is also available (in combination with the right AspP). In the other two cases, however, we have the paradoxical situation that the primary result reading is not there; it seems that the structure is abused just to express the secondary reading. The availability of the strategy, it seems, results from the conspiracy of two factors: (i) the availability of ACQ as R and (ii) the fact that result structures may have a secondary 'can' interpretation.

One of the goals of the present paper was to try to understand how the post-verbal modal fits in the grammar of the languages under consideration here; after all, all other modal verbs appear in a pre-verbal position. If our analysis is right, ACQ is in a post-verbal modal position in one case (the potential construction), and is R or Asp<sup>0</sup> in other cases, with the modal reading as a secondary reading only, following general patterns of the languages in question. In all cases, modal ACQ fits in smoothly with the rest of the grammar.

### 4.3 How to explain the variation

After presenting analyses of post-verbal ACQ with modal meanings, we have one task left: explaining the variation. Why are the four languages we investigated in this study so different when it comes to post-verbal modal ACQ? Why don't they all have both ACQ-ability and ACQ-potential in the same way?

First of all, they do not differ with respect to ACQ-potential; they all have it and, presumably, the analysis is the same for all languages. The point is that Cantonese and Zhuang *only* have the ACQ-potential. The point of variation we set off to explain was the difference between (1a) and (1b): in Lao, post-verbal ACQ can have an ability reading, while its Zhuang counterpart cannot. Now, we can rephrase this in structural terms and say that in Zhuang (and Cantonese), post-verbal ACQ can only occur in the potential construction, while that is not the case for Lao. Why is that so? What other difference between Lao and Zhuang can this particular difference be related to? Let us consider some possibilities.

One possibility would be that in Zhuang (and Cantonese), we really only have one ACQ, a modal, and that's it. However, this is clearly not the case: we see non-modal ACQ in both languages (linker ACQ, for instance), and Zhuang also has ACQ-achievement.

Another possibility is that it is related the fact that in Zhuang (and Cantonese), but not in Lao (and Vietnamese), ACQ is obligatorily adjacent to the verb. As mentioned above, this may be just a coincidental effect of the fact that ACQ-potential is generated in a position to the immediate right of the verb, but it seems that more is the matter. First, we observed that in all constructions that involve post-verbal ACQ, ACQ is obligatorily adjacent to the verb in Zhuang while such restriction does not apply in Lao. Wherever Lao could have *daj*<sup>4</sup> in a position separated from the verb, Zhuang either did not have this construction, or it had it, with the restriction that *ndaej* had to be adjacent to the verb.

Note in this context that, whereas in Lao the positioning in the sentence of ACQ is basically the same as that of the result denoting predicate in resultative constructions,

this is not the case in Zhuang. In Zhuang, the result denoting element in resultatives is generally not adjacent to the verb. (In Sinitic, on the other hand, all R-elements are adjacent to the verb.) The behavior of *ndaej* in Zhuang, then, is a particular property of the element itself.

The fact that *ndaej* is distributionally different from result denoting elements could be significant since whatever determines the distributional behavior of R-elements on the one hand and *ndaej* on the other may lead to a clash, making *ndaej* an unlikely R. On the other hand, *ndaej* does appear as the achievement marker, presumably analyzable as R/X<sup>0</sup>, in sentences such as (30), and it does immediately follow the verb. Although more research is necessary, as things stand, it does not seem to be the case that the crucial difference between Zhuang and Lao lies in the distribution of post-verbal ACQ.

The crucial difference between Lao and Zhuang may be the fact that V-R combinations are ambiguous in Lao but not in Zhuang. Aside from the primary result reading, Lao has the secondary ‘can’ reading, while Zhuang does not.

Earlier I said that the secondary reading is only possible when the Asp<sup>0</sup> in the small clause is either entirely empty or simply missing. Since the secondary reading does not arise in Zhuang, we can conclude that its Asp<sup>0</sup> cannot be missing or must be filled, in the relevant cases, that is. In Mandarin, V-R combinations have to be accompanied by perfective *le*, unless they are embedded in generic statements or under a modal verb (Sybesma and Vanden Wyngaerd 1997); this is shown in (91).

- (91) a. *tā yā-biǎn-\*(le) tā-de miànbāo.*  
           3s press-flat-perf 3s-de bread  
           ‘S/he pressed his bread flat.’  
       b. *tā zǒngshì yā-biǎn tā-de miànbāo.*  
           3s always press-flat 3s-de bread  
           ‘S/he always presses his/er bread flat.’  
       c. *tā xiàng yā-biǎn tā-de miànbāo.*  
           3s wants press-flat 3s-de bread  
           ‘S/he wants to press his/er bread flat.’

We can interpret this as that the Asp<sup>0</sup> position must either be filled or governed. On the basis of the limited research I have done so far, I conclude that this is true for Zhuang as well. This means that, even if *ndaej* in Zhuang could be analyzed as R, there would still be no possibility of getting the secondary interpretation for independent (or “bare”) V-R sentences.

In short, the main reason why Zhuang does not have an ability reading for ACQ has to do with the fact that in Zhuang V-R combinations, as opposed to those in Lao, the secondary ‘can’ modality reading never arises.<sup>33</sup> The non-potential modal ACQ results

33. Two other differences may be relevant. First, Lao, and Vietnamese, have ACQ as a lexical verb meaning ‘have procedural knowledge of’ and Zhuang, and Cantonese, do not. It may

from the conspiracy of two factors: (i) the availability of ACQ as R and (ii) the fact that result structures may be extended to a secondary ‘can’ interpretation, made possible by a missing or empty Asp<sup>0</sup>. It seems that at least the latter is missing in Zhuang.

## 5. Conclusion

The problem we set out to solve was why Lao can have a post-verbal ACQ with an ability reading, while genetically related Zhuang cannot. We just concluded that the reason is that Zhuang misses one crucial property: in Zhuang, V-R combinations do not have a secondary ‘can’ reading. The fact that ACQ must be adjacent to the verb may be another factor, as it may make ACQ not generally available as R.

The requirement that *ndaej* be adjacent to the verb is shared by Cantonese. For Cantonese we noted that there are good reasons to view *dak*<sup>1</sup> as having grammaticalized to a large extent, at least with respect to its syntactic behavior. The use as the lexical verb ‘okay’ aside, we saw that it can no longer operate independently, it always needs to be supported (or accompanied) by another element. With some exception here and there, this is a general trend in Sinitic. It seems to be the case that Zhuang follows this Sinitic trend.

The fact that Zhuang’s V-R combinations do not have a secondary ‘can’ reading was analyzed by pointing out that its Asp<sup>0</sup> cannot be absent or empty if ungoverned. This too is a property that Zhuang shares with Cantonese, and Sinitic languages more generally.

The title of this paper is “Zhuang: A Tai language with some Sinitic characteristics”. Zhuang is genetically Tai and as such it is related to Lao, with which it has many traits in common, as we saw. We also saw that it is quite unlike Lao in other respects. In those respects, Zhuang is very much like Cantonese.

If in certain respects Zhuang is like Cantonese and unlike Lao, and Lao is more like Vietnamese in these respects, then there are two possibilities. The first possibility is that Zhuang kept its Tai-ness, and the differences with Lao are explained if we assume that Lao developed in another direction, together with Vietnamese; similarities between Cantonese and Zhuang can be due to the Taiization that Cantonese is supposed to have undergone, as opposed to the northern varieties of Sinitic (true or not). The second

---

be the case that this verb shows up as a secondary predicate here. However, whereas one can have a “procedural knowledge” of languages, stories and songs, as explained in §3.1, it is much harder to envisage such knowledge of keys. Secondly, the absence of ACQ-ability in Zhuang and Cantonese may be related to the fact that they do not feature an overt modal verb in the higher functional layers of the sentence as an overt doubler of post-verbal modal ACQ. I suggested earlier that for ACQ-ability overt support would be more pertinent than for ACQ-potential. If this is right, languages that do not feature such doubler, cannot have ACQ-ability. But, as also mentioned earlier, the evidence is thin.



possibility is that Zhuang lost some of its Tai-ness, and grew away from Lao into the direction of Cantonese. In view of the fact that the two crucial properties that make Zhuang different from Lao in its use of post-verbal modal ACQ are not just Cantonese but more generally Sinitic, I assume that the second scenario just sketched is the correct one. This would make Zhuang a Tai language with some Sinitic characteristics.

## References

- Bauer, R.S. 1988. Written Cantonese of Hong Kong. *Cahiers de Linguistique Asie Orientale* 17: 245–293.
- Bisang, W. 1992. *Das verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer. Vergleichende Grammatik im Rahmen der Verbserialisierung, der Grammatikalisierung und der Attraktorpositionen*. Tübingen: Gunter Narr.
- Cadière, L. 1956. *Syntaxe de la langue Viêtnamienne*. Paris: École Française d'Extrême-Orient.
- Chappell, H. 2001a. Synchrony and diachrony of Sinitic languages: A brief history of Chinese dialects. In *Sinitic grammar. Synchronic and diachronic perspectives*, Hilary Chappell (ed.), 3–28. Oxford: OUP.
- Chappell, H. 2001b. Language contact and areal diffusion in Sinitic languages. In *Areal diffusion and genetic inheritance. Problems in comparative linguistics*, A.Y. Aikhenvald & R.M.W. Dixon (eds.), 328–357. Oxford: OUP.
- Chen, Ping. 1999. *Modern Chinese. History and sociolinguistics*. Cambridge: CUP.
- Cheng, L.L.S. & Sybesma, R. 2003. Forked modality. In *Linguistics in the Netherlands 2003*, L. Cornips & P. Fikkert (eds.), 13–23. Amsterdam: John Benjamins.
- Cheng, L.L.S. & Sybesma, R. 2004. Post-verbal 'can' in Cantonese (and Hakka) and Agree. *Lingua* 114: 419–445. (Special issue on the variation in the syntax of auxiliaries).
- Dài Qìngxià (ed.). 1992. *Hànyǔ yǔ shǎoshùmínzú yǔyán guānxi gǎilùn (General outline of the relation between Chinese and the minority languages)*. Peking: Zhōngyāng Mǐnzhǔ Dàxué.
- Duffield, N. 1998. Auxiliary placement and interpretation in Vietnamese. In *Proceedings of the 34th Regional Meeting of the Chicago Linguistic Society*, M.C. Gruber, D. Higgins, K.E. Olson & T. Wysocki (eds.), 95–109. Chicago IL: CLS.
- Duffield, N. 2003. Final modals, adverbs and antisymmetry in Vietnamese. Ms. McGill University.
- Edmondson, J.A. & Solnit, D.B. 1997. Introduction. In *Comparative Kadai. The Tai branch*, J.A. Edmondson & D.B. Solnit (eds.), 1–26. Arlington TX: SIL/University of Texas.
- Enç, M. 1987. Anchoring conditions for Tense. *Linguistic Inquiry* 18: 633–657.
- Enfield, N. J. 2000. Lao as a national language. In *Laos. Culture and society*, G. Evans (eds.), 258–290. Singapore: Institute for Southeast Asian Studies. (Originally published 1999 by Silkworm Books, Chiang Mai).
- Enfield, N. J. 2001. On genetic and areal linguistics in mainland South-East Asia: Parallel poly-functionality of 'acquire'. In *Areal diffusion and genetic inheritance. Problems in comparative linguistics*, A.Y. Aikhenvald & R.M.W. Dixon (eds.), 255–290. Oxford: OUP.
- Enfield, N.J. 2003. *Linguistic epidemiology. Semantics and grammar in mainland South-East Asia*. London: Routledge.
- Enfield, N.J. 2004. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34: 181–206.

- Enfield, N. J. To appear. Verbs and multi-verb sequences in Lao. In *Survey of the Tai-Kadai language family*, A.V. N. Diller & J. Edmondson (eds.), London: RoutledgeCurzon.
- Ethnologue. www.ethnologue.com.
- Guéron, J. & Hoekstra, T. 1989. T-chains and the constituent structure of auxiliaries. In *Constituent structure. Papers from the 1987 GLOW conference* [Annali Di Ca' Foscari 28:4], A. Cardinaletti, G. Cinque & G. Giusti (eds.), 35–99.
- Guéron, J. & Hoekstra, T. 1995. The temporal interpretation of predication. In *Small clauses* [Syntax and semantics 28], A. Cardinaletti & M.-T. Guasti (eds.), 77–107. New York NY: Academic Press.
- Hoekstra, T. 1988. Small clause results. *Lingua* 74: 101–139.
- Hoekstra, T. 2005. Small clauses everywhere. In *Arguments and structure. Studies in the architecture of the sentence*, R. Sybesma, S. Barbiers, M. den Dikken, J. Doetjes, G. Postma & G. Vanden Wyngaerd (eds.), 319–389. Berlin: Mouton.
- Huáng, Xíng. 2000. *Zhōngguó shǎoshù mínzú yǔyán huó lì yánjiū* (An investigation into the vitality of the minority languages of China). Peking: Zhōngyāng Mínnú Dàxué.
- Huang, Yuanwei. 1997. The interaction between Zhuang and the Yue (Cantonese) dialects. In *Comparative Kadai. The Tai branch*, J.A. Edmondson & D.B. Solnit (eds), 57–76. Arlington TX: SIL/University of Texas.
- Ivarsson, S. 2000. Towards a new Lao: *Lao Nhay* and the campaign for national “reawakening” in Laos, 1941–1945. In *Laos. Culture and society*, G. Evans (ed.), 61–78. Singapore: Institute for Southeast Asian Studies. (Originally published 1999 by Silkworm Books, Chiang Mai).
- Klein, W. 1994. *Time in language*. London: Routledge.
- Lamarre, C. 2001. Verb complement constructions in Chinese dialects: Types and markers. In *Sinitic grammar. Synchronic and diachronic perspectives*, H. Chappell (ed.), 85–120. Oxford: OUP.
- LaPolla, R. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In *Areal diffusion and genetic inheritance. Problems in comparative linguistics*, A.Y. Aikhenvald & R.M.W. Dixon (eds.), 225–254. Oxford: OUP.
- Lǐ Jǐnfāng. 2002. *Dòng-Tái yǔyán yǔ wénhuà* (Dong-Tai language and culture). Peking: Mínnú.
- Luong, Van Ke. 1994. *Der chinesische Einfluss auf die vietnamesische Sprache*. Münster: Lit.
- Matisoff, J.A. 1991. Areal and universal dimensions of grammaticalization in Lahu. In *Approaches to grammaticalization. Vol. II. Focus on types of grammatical markers*, E. Closs Traugott & B. Heine (eds.), 383–453. Amsterdam: John Benjamins.
- Matthews, S. & Yip, V. 1994. *Cantonese. A comprehensive grammar*. London: Routledge.
- Nguyễn, Đình-Hoà. 1997. *Vietnamese*. Amsterdam: John Benjamins.
- Norman, J. 1988. *Chinese*. Cambridge: CUP.
- Norman, J. 1991. The Min dialects in historical perspective. In *Languages and dialects of China*, W.S.-Y. Wang (ed.), 325–360. Berkeley CA: Journal of Chinese Linguistics.
- Qín Xiaoháng. 1995. *Zhuàngyǔ tèshū yǔfǎ xiànxàng yánjiū* (A study of the special grammatical phenomena of Zhuang). Peking: Mínnú.
- Qin, Xiaohang. 2004. *Zhuang lexicology. A study on Zhuang words and an attached Zhuang-English vocabulary*. Peking: The Ethnic Publishing House [mínzú chūbǎnshè].
- Ramsey, S.R. 1989. *The languages of China*. 2nd edn. Princeton NJ: Princeton University Press.
- Simpson, A. 2001. Focus, presupposition and light predicate raising in East and Southeast Asia. *Journal of East Asian Linguistics* 10: 89–128.
- Soh, Hooi Ling. 2003. A V-raising analysis of Cantonese post-verbal modal constructions with *dak* ‘can’. Ms. University of Minnesota.

- Sybesma, R. 1997. Why Chinese verb-*le* is a resultative predicate. *Journal of East Asian Linguistics* 6: 215–261.
- Sybesma, R. 1999. *The Mandarin VP*. Dordrecht: Kluwer.
- Sybesma, R. 2004. Exploring Cantonese tense. In *Linguistics in the Netherlands 2004* [AVT 21], L. Cornips & J. Doetjes (eds.). Amsterdam: John Benjamins.
- Sybesma, R. & Vanden Wyngaerd G. 1997. Realizing end points: The syntax and semantics of Dutch *ge* and Mandarin *le*. In *Linguistics in the Netherlands 1997* [AVT 14], J. Coerts & H. de Hoop (eds.), 207–218. Amsterdam: John Benjamins.
- Tang, Sze-Wing. 2002. Focus and *dak* in Cantonese. *Journal of Chinese Linguistics* 30: 266–309.
- Tenny, C. 1987. Grammaticalizing aspect and affectedness. PhD dissertation, MIT.
- Thompson, Laurence C. 1965. *A Vietnamese grammar*. Seattle WA: University of Washington Press.
- Travis, L. 1991. Derived objects, inner aspect and the structure of VP. Paper presented at NELS 22 [not included in the proceedings]. Ms. McGill.
- Tsai, Wei-tien D. 1999. The hows of *why* and the whys of *how*. In *UCI Working papers in Linguistics* 5, F. del Gobbo & H. Hoshi (eds.), 155–184.
- Wáng Jun. 1962. Zhuàngyǔ zhōng de Hànyǔ jiècí (Chinese loanwords in Zhuang). *Zhōngguó Yǔwén* 1962(6): 251–264.
- Wěi Qīngwén & Qín Guóshēng. 1980. Zhuàngyǔ jiǎnzì (Simple record of the Zhuang language). Peking: Mǐnzú.
- Wu, Ching-Huei T. 2004. On *de/bu* and the syntactic nature of verbal compounding. *Language and Linguistics/Yǔyán jì yǔyánxué* 5: 271–328.
- Wu, Yunji. 2001. The development of [DE+V] and [V+DE] constructions in the Hunan dialects. *Yǔyán Yánjiū*, 45–59.
- Yue-Hashimoto, A. 1991. The Yue dialects. In *Languages and dialects of China*, W.S.-Y. Wang (ed.), 294–324. Berkeley CA: Journal of Chinese Linguistics.
- Yue, A. 2001. The verb complement construction in historical perspective with special reference to Cantonese. In *Sinitic grammar. Synchronic and diachronic perspectives*, H. Chappell (ed.), 232–265. Oxford: OUP.
- Zhāng Yuánshēng. 1987. Zhuàngyǔ [Zhuang]. *Zhōngguó shǎoshù mínzú yǔyán* (China's minority languages), no ed., 268–286. Chéngdu: Sichuan Mǐnzú.
- Zhāng Yuánshēng & Qín Xiǎoháng. 1993. *Xiàndài Zhuàng-Hànyǔ bǐjiào yǔfǎ* (A comparative grammar of modern Zhuang and Chinese). Peking: Zhongyang Mǐnzú Xuéyuan.
- Zhang, Yuansheng & Wei Xingyun. 1997. Regional variants and vernaculars in Zhuang. In *Comparative Kadai. The Tai branch*, J.A. Edmondson & D.B. Solnit (eds.), 77–96. Arlington TX: SIL/University of Texas.

# Language index

## A

Abaza 29, 31, 47, 49, 58–60, 68–74  
 Abkhaz 29–31, 44, 49, 50, 52, 54–58, 60–64, 68–75, 77, 78  
 Abzhywa 45, 47, 63, 74  
 Bzyp 45, 47, 63  
 Proto-Abkhaz 60  
 Sadz 43, 45  
 Samyrzaqan 74  
 South 71  
 Abkhazo-Adyghean 28, 29, 64  
 Abui 110, 136, 146  
 Abun 110, 111, 114–116, 124, 131, 134–136, 146  
 Adang 110, 120, 124, 131, 136, 146  
 Adyghe 29, 30, 49, 53, 55–60, 62, 63, 68, 69, 71, 73, 75, 77  
 Abadzakh 47  
 Bzhadugh 44, 45, 47  
 Shapsygh 44  
 Temirgoy 47  
 African 19  
 Afroasiatic 30, 44, 46, 48, 59  
 Aghul 29, 49, 54, 55, 59, 60, 63, 71–74, 81  
 Burshag 45, 47, 81  
 Fite 46  
 Aikanā 156 (Masaká, Kasupá, Huarí) 159, 160, 165, 167, 169, 171  
 Ainu 47, 48, 60  
 Akhvakh 29, 45, 47, 49, 58–60, 71, 73, 74  
 Northern 51  
 Akkadian 81  
 Akuntsu (Akütsü) 156, 159, 162, 168  
 Akuntsu-Kanoë 160  
 Alamblak 61

Alanian, *see* Old Ossetic  
 Albanian 28, 46, 48, 52, 60, 75, 184, 186, 189, 192, 196, 202–204, 209, 211, 212, 214–217  
 Geg 204, 205  
 Tosk 202, 204, 205, 217  
 Aleut 47, 48  
 Alorese 97, 112, 113, 119, 121, 127–129, 132, 136, 146  
 Alune 112, 133, 134, 146  
 Ambai 111, 121, 122, 127, 129, 132, 136, 146  
 Amerindian 18, 61  
 Andi 29, 44, 45, 47–49, 58, 59, 63, 70–73, 78  
 Andic 58–60  
 Aquitanian 61  
 Arabic 14, 19, 46, 48, 62, 63, 69, 70, 75, 81, 106  
 Araona 158, 161, 163  
 Arawakan 18, 151, 153, 156, 158, 164–166, 172  
 Archi 29, 31, 43, 45, 47, 49, 59, 60, 63, 71–73, 81  
 Arikapú 158, 161, 165, 167, 168  
 Arikém 162  
 Armenian 29–31, 39, 43–45, 48, 49, 51–55, 58–77, 79, 80, 82  
 Artvin 71, 72  
 Classical 43, 45, 53  
 Common 45  
 East 43–46  
 Literary 45  
 Tiflis 71, 72  
 West 43, 46  
 Aromanian 184, 186, 193, 196–198, 200–204, 206, 208, 209, 213–217  
 Farsheriote 215  
 Western 193  
 Aruá 159  
 Asilulu 111

Austroasiatic 60  
 Austronesian 95–139, 146–149  
 Proto-Austronesian 114, 129, 133, 134  
 Avar 29, 31, 45, 47, 49, 51, 53–56, 58, 60, 61, 63, 64, 68, 70–74, 78, 79  
 Antsukh 71, 72  
 Belokan 71  
 Proto-Avar-Andi 34–36  
 South 72  
 Upper Samur 71  
 Zakatala 71, 73  
 Aymara 62, 167, 168, 170–172  
 Aymaran 167  
 Aysor 29  
 Azeri 29–31, 43–46, 48, 49, 52, 58, 60–63, 67, 70, 72–74, 76–78  
 Derbent 67  
 Kuba 67, 71, 78, 79  
 Kutkashen 67, 78  
 Lower-Katruk 78, 79  
 Nukh 67  
 Tabasaran 67, 79  
 Terekemi 78  
 Vartashen 67, 78  
 Zakatala-Kakh 67, 78, 79

## B

Bagvala 29, 46, 55, 58, 60  
 Baham 110, 111  
 Balkan Romani 184, 217  
 Arli 201, 216  
 Balkar, *see* Karachay-Balkar  
 Baltic 28, 46, 48  
 Baluchi 48, 60  
 East 46  
 West 46  
 Bandanese 112, 127, 129, 136, 146  
 Bantu 15  
 Bashkir 46, 48  
 Basque 10, 14, 46, 48, 60, 61

- Bats 29, 31, 44, 46, 47, 49, 54, 58, 60, 61, 63, 65, 70–72, 77, 81  
 Baure 158, 161, 164, 165, 171  
 Berber 15  
 Bezhta 29, 47, 49, 59, 60, 63, 71, 72, 74  
 Bhojpuri 19  
 Biak 111, 112, 121, 122, 124–126, 129, 134, 136, 137  
     Numfor dialect 146  
 Bimanese 112, 120, 122, 132, 146  
 Blagar 110, 111, 119, 136, 146  
 Border 115  
 Botlikh 29, 45, 49, 58, 60, 63, 71–73  
 Brat, *see* Maybrat  
 Budukh 29, 46–49, 59  
     –61, 72–74, 77, 78  
 Bulgarian 52, 184, 186, 189, 192, 194–196, 200–202, 209, 210, 212, 213, 216  
     Rodopi 189, 204  
     Southern 199  
 Bunak 110, 111, 114–116, 119, 131, 136, 146  
 Buru 111, 112, 116, 121, 127, 128, 134, 136, 146  
 Burushaski 46, 48, 53, 60, 61, 62
- C**  
 Canichana 156, 159, 160  
 Cantonese 2, 221–225, 227–260, 267, 269–272  
     Baihua 226  
     Standard 227  
     Yue 227  
 Carib 18  
 Cariban 162  
 Catalan 14  
 Cavineña 158, 161, 165, 167, 168, 170, 171  
 Cayubaba 156, 160, 161, 164, 165, 171  
 Celtic 10, 28, 47, 48, 60, 61  
 Central-Eastern-Malayo-Polynesian (CEMP) 112, 116, 120, 122, 134, 137  
 Central-Malayo-Polynesian (CMP) 112, 113, 146–149  
 Chácobo 158, 161, 165, 170, 171  
 Chamala 29, 45, 47, 49, 58, 60, 62
- Chapacuran 151, 152, 156, 158, 170  
 Chechen 29, 46, 47, 49, 52–54, 56, 58, 60, 61, 63, 71, 72, 74  
     Kistin 70, 72  
 Chibcha 18  
 Chimane 156, 160, 161  
 Chinese 19, 191, 225–228, 243  
     Classical 228  
 Chiquitano (Bes½ro) 156, 160, 161, 165, 166, 171  
 Chol 61  
 Chukchi-Kamchatkan 47, 48  
 Cinta Larga 154, 159  
 Circassian 29, 31, 44, 49, 51, 54, 68, 69, 74, 75, 77, 78  
     Proto-Circassian 34–36, 60, 63  
 Creole  
     Caribbean 19  
     Portuguese 18  
 Croatian 189
- D**  
 Dacian 189  
 Daghestanian 28, 29, 31, 49–51, 56, 63, 68, 70–72, 75, 76, 78  
 Dani 130  
 Danish 60  
 Dargi 29, 31, 46, 47, 49, 51, 54, 59, 60–64, 67, 71, 72, 74, 79  
     Chirakh 71, 73  
     Khaydag 79  
     Kubachi 29, 71, 73, 81  
     Megeb 71, 73  
     Proto-Dargi 34–36  
 Dari 46  
 Djeoromitxi 158, 161, 167, 168  
 Dobel 112, 127, 132, 136, 147  
 Dravidian 53  
 Dutch 18, 47, 48, 60  
 Dzongkha 60
- E**  
 East Bird's Head (EBH) 110, 115  
 East Caucasian (EC) 64  
 East Timor 110  
 Eastern-Malayo-Polynesian (EMP) 112, 137  
 Enawê-Nawê (Salumã) 158, 166
- English 14, 18, 46, 48, 222, 244  
     Caribbean 18  
     North American 18  
 Ese Ejja 158, 161, 167  
 Eskimo-Aleut 47, 48
- F**  
 Faroese 60  
 Fataluku 109  
 Fijian 137  
 Finnic 47, 48  
 Finnish 47  
 Formosan 134  
 French 14, 19, 47, 48, 60
- G**  
 Gagauz 46  
 Galela 110, 111, 123, 124, 138, 147  
 Gavião 153, 159, 160, 167, 171  
 Georgian 29–31, 33, 35, 45, 46, 47–57, 59–65, 67–77, 79, 81–83  
     Adzharian 75, 76  
     Fereydan 49  
     Gurian 75, 76  
     Imeretian 75, 76  
     Ingilo 49, 72, 73, 76, 212  
     Kakhetian 77  
     Khevsur 78  
     Lechkhum 76  
     Mountain 70, 71, 76, 77  
     Mtiul 77  
     Old 47, 51, 57  
     Pshav 77  
     Standard 47, 49  
     Tush 77, 78  
     Western 71, 75  
 German 46, 48  
     High 14, 19  
     Low 14, 19  
 Germanic 28, 46–48  
 Godoberi 29, 49, 58, 60, 71–73  
 Greek 28, 29, 46–48, 51, 52, 59, 75  
     Ancient 14, 189, 212  
     Byzantine 69, 71  
     Cappadocian 29  
     Koine 14  
     Modern 184, 186, 189, 192, 194, 196, 200, 202, 204–206, 213, 216, 217  
     Northern 29  
     Northern Macedonian 194

Pontic 29, 67, 68, 73, 75, 83  
 Tsalka-Alaverdy 29, 73  
 Guaraní 18

## H

Hamap 136  
 Hamitic 26  
 Hatam 110, 111, 114, 115, 118,  
 122, 125, 130, 133, 147  
 Hindi 46, 48  
 Hinukh 29, 46, 47, 49, 59, 60,  
 63, 71, 72  
 Hittite 81  
 Hungarian 14, 46, 48  
 Hunzib 29, 46, 47, 49, 51,  
 55, 59, 60, 63, 71–73  
 Proto-Hunzib-Bezhta 34–  
 36  
 Hurri 53, 54  
 Hurro-Urartian 53, 76

## I

Ibero-Caucasian 33, 34  
 Icelandic 14, 47, 48  
 Idade 111, 112, 127, 128, 136, 147  
 Ignaciano 158, 161,  
 Iha 109–111  
 Illyrian 189  
 Inanwatan 110, 111, 115, 116,  
 124, 130, 131, 136, 147  
 Indoarian 46  
 Indo-European 26, 27, 29,  
 44–49, 59, 61, 63, 65–67,  
 72, 73, 75, 80–84  
 Proto-Indo-European 34–  
 36, 45, 55  
 Indo-Iranian 54, 60, 61  
 Indonesian (Standard) 106, 123  
 Ingush 29, 46, 47, 49, 56,  
 58, 60, 63, 69, 71, 72, 74,  
 75, 84  
 Iranian 29, 46, 48, 60, 66, 72,  
 75, 83  
 Middle 63  
 North-Eastern 29  
 North-Western 29  
 South-Western 29  
 Irantxe 153, 156, 159, 166  
 Irarutu 111, 112, 132, 136, 147  
 Ishkashim 46  
 Isní 111, 112, 120, 136, 147  
 Italian 14  
 Calabrian 60  
 Sicilian 60

Itelmen 47, 48  
 Itonama 154, 156, 160, 161,  
 163–5, 167, 168, 171

## J

Jabutian 151, 152, 154, 156,  
 158, 161  
 Japanese 53  
 Judeo-Spanish 184, 185  
 Istanbul Judesmo 212

## K

Kabardian 29, 31, 47, 56,  
 58–60, 62–64, 69, 71–73,  
 75, 78, 81, 84  
 Archaic 60  
 Kuban 59  
 Kafoa 110, 136  
 Kaitetu 112, 120, 136, 147  
 Kalmyk 46  
 Kambera 112, 120, 122, 123,  
 125, 128, 129, 132, 147  
 Kanoë (Kapixaná) 156,  
 159–162, 165, 167–171  
 Karachay-Balkar 29–31, 44,  
 46, 48, 49, 58–61, 63, 64,  
 67, 69–72, 74–76  
 Baksano-Chegem 67  
 Upper-Balkar 67  
 Karaim 46, 48  
 Karakalpak 46, 48  
 Karata 29, 49, 58, 60, 71  
 Karelian 46, 48  
 Karipuna 159, 162  
 Karitiana 154, 159, 161, 162,  
 166, 167, 170, 171  
 Karo (Arara) 153, 159, 161,  
 162, 166, 168, 171  
 Kartvelian 1, 25, 28–30, 33–40,  
 43, 47, 49, 51–73, 76–84  
 East 29, 79  
 Proto-Kartvelian 34, 36,  
 43, 45, 47, 60, 80, 82  
 West 29, 79  
 Kati 60  
 Kaw To Yo (Kuyubi) 158  
 Kawahib 159  
 Kaxararí (Kaxarirí) 158, 166  
 Kazakh 47, 48  
 Keiese (Kei) 111, 112, 121, 127,  
 128, 133, 134, 136, 147  
 Kemak 111, 112, 120, 121, 132,  
 136, 147  
 Keo 112, 120, 122, 132, 147

Kerek 47  
 Ket 47, 48  
 Khinalug 29, 35, 36, 46–49,  
 59, 60, 72–74, 77, 78, 81  
 Khvarshi 29, 49, 59  
 Inkhokvari 46, 47, 59  
 Klón 110, 114, 116, 119, 136, 147  
 Kolana 115  
 Komi-Zyrian 46  
 Korean 46, 48  
 Koriak 47, 48  
 Krym-Tatar 46, 48  
 Kryz 29, 46–49, 59, 60, 72–74,  
 77, 78, 81  
 Kui 136  
 Kumyk 29, 30, 44, 46, 48,  
 49, 58–60, 67, 68, 70, 71,  
 74, 76  
 Buynaksk 79  
 Kaytag 67, 79  
 Lower 79  
 Terek 76  
 Kurdish 29, 46, 48, 49, 51, 53,  
 54, 58, 59, 61, 63, 67, 68,  
 70, 83  
 Kwazá (Koaiá) 156, 159, 161,  
 164, 165, 167–171  
 Kyrgyz 46, 48

## L

Lak 29, 31, 35, 36, 45, 47, 49,  
 51, 54, 55, 59, 60, 63, 64,  
 74, 79, 81  
 Lakalei 111, 112, 120, 121, 132,  
 136, 147  
 Lakes Plain 108  
 Lakondé 158, 161, 162, 167,  
 168, 171  
 Lamma 110, 131, 136, 147  
 Lao 2, 221–225, 227–257,  
 262–272  
 Latin 14, 53, 189, 212, 228  
 Latundé 158, 161, 162, 165, 168  
 Latvian 46, 48  
 Laz 29, 31, 47, 49, 50, 54, 55,  
 57, 59, 60, 63, 67–73,  
 75–79, 81–83  
 Leko 156, 160, 165, 171  
 Leti 111, 112, 127, 128, 132,  
 136, 147  
 Lezgi 29, 31, 35, 46, 47, 49,  
 53–57, 59–64, 70–74, 78,  
 79, 81  
 Kimil 49, 73

Lezgic 59, 60  
  Proto-Lezgic 34–36, 60  
Lingua Franca 14  
Lithuanian 46, 48  
Lolein 111, 112, 120, 136, 147

## M

Ma'ya 111, 112, 134, 135, 147  
Macedonian (Standard) 52,  
  184, 186, 189, 192–197,  
  199–206, 208–210,  
  212–217  
  Bitola 213  
  Eastern 195, 209  
  Northern 192, 199, 217  
  Pirin 199  
  Prilep 206  
  Southeastern 194, 195, 204  
  Southern 206  
  Southwestern 195, 196,  
  200, 205, 206, 209, 217  
  Western 195, 204, 208  
Machineri 158  
Macro-Ge 156, 158, 161, 166  
Maimandê 162  
Makasai 110, 114, 115, 136, 147  
Makuráp 159, 168  
Malay 105–107  
  Alor 127  
  Ambonese 146  
  High or Church 106  
  Low 106  
Mambai 111, 112, 127, 132,  
  136, 147  
Manchurian 26  
Mandarin 222, 225–227, 233,  
  258–260, 265, 266, 270  
  Guanhua 226  
  South-Western 226  
Mansim 110, 111, 114, 133, 147  
Mari 47  
Marind 110, 148  
Maropa (Reyesano) 158  
Mashubi 161  
Matbat 134, 135, 148  
Maya 61  
Maybrat 110, 111, 114, 115, 119,  
  124, 130, 134, 136, 148  
Media Lengua 10  
Megeb 29  
Megleno-Romanian 184, 193,  
  197, 198, 200–204, 206,  
  208, 214–216

Megrelian 29–31, 46, 47, 49,  
  51–55, 59, 60, 63–65,  
  68–79, 81–83  
Mekens (Sakiráp) 159, 161,  
  164, 165, 167, 168, 170, 171  
Meyah (Meax) 110, 111, 114,  
  115, 122, 125, 134, 148  
Miguelinho 158  
Moi 104, 110, 111, 114, 125, 148  
Mojo 158  
Moldovan 46  
Mongolian 26, 46, 48  
Mongolic 46, 48  
Mon-Khmer 223, 228  
Mor 111, 112, 132, 134–136, 148  
Moraid 110, 111, 148  
Mordvin 46, 48  
Moré (Itene) 158  
Mosetén 156, 160, 161, 165,  
  167, 168, 170, 171  
Moskona 111  
Movima 156, 160, 161, 165,  
  170, 171  
Mpur 110, 111, 114, 115, 118,  
  134–136, 148  
Muna 112, 120, 122, 148  
Munda 60  
Mýky 159

## N

Nagarotê 162  
Nahuatl 61  
Nakh 29, 50, 51, 58, 60, 63, 68,  
  70, 76, 78  
  Proto-Nakh 34–36  
Nambikwara 158, 160, 162,  
  167  
Nambikwaran 151, 153, 156,  
  158, 166, 168, 169  
Nanay 47  
Nedebang 110, 111  
Negerhollands 18  
Negidal 47, 48  
Nenets 47, 48  
Neo-Aramaic 29, 44, 46, 48,  
  51, 59, 68, 70, 83  
  Eastern 45  
Nganasan 47, 48  
Nivkh 47, 48  
Nogay 29, 30, 46, 48, 49, 59,  
  60, 68–71, 74  
North American indigenous  
  languages 44

North Caucasian (NC) 1, 25,  
  26, 33–40, 51, 62, 64, 65,  
  72, 73, 78–82  
Proto-North-Caucasian 48,  
  80  
North Halmahera 108–110,  
  114, 115, 120, 123  
North-East Caucasian  
  (NEC) 29, 33–40, 44, 45,  
  48, 52, 54–57, 61–63, 67,  
  68, 70, 71, 76–79, 82–84  
North-West Caucasian  
  (NWC) 29, 30, 33–40,  
  44, 47–49, 51–59, 61–64,  
  68–70, 77, 78, 82, 83  
North-West-Coastal (NWC)  
  Papuan 102, 103, 109  
Norwegian 14

## O

Oceanic 61, 112, 117, 118, 122,  
  129, 130, 137, 138  
  Proto Oceanic 116, 122,  
  129, 134, 137, 138  
Oghuz 72, 73  
Old Church Slavonic 212  
Oraita 110, 111  
Oro Towati (Oro Win) 158  
Oroch 47, 48  
Orok 47, 48  
Ossetic 29, 31, 39, 43, 44,  
  46, 48, 51–53, 55, 58–67,  
  70–72, 74–76, 79, 80, 82, 84  
  Digor 48  
  Iron 48, 49, 71  
  Literary 59, 60  
  North 71  
  Old (Alanian) 60, 69, 71,  
  74–76  
  Shepherd's jargon 59  
  South 70, 74

## P

Pacahuara 158, 161  
Pagu 110, 114, 131, 138, 148  
Palmella 162  
Panoan 151, 156, 158, 162, 163,  
  170  
Papiamentu 18  
Papuan 95–139, 146–149  
  East 108  
Pareci 153, 158, 166  
Parya 48

- Pashai 60  
 Permiak 46, 48  
 Persian 46, 48, 52, 62–64, 67, 69, 70, 72  
     Middle (Pahlavi) 63, 64, 70  
 Portuguese 18, 19, 107  
     Brazilian 19  
 Proto-Malayo-Polynesian 137  
 Puruborá 159, 162
- Q**  
 Quechua 9, 10, 18, 62, 170–172
- R**  
 Ramu 108  
 Richa Aghul 46  
 Rikbaktsá 156, 158, 166  
 Romance 46, 47, 48  
     Balkan 186, 191–194, 197, 198, 202, 204, 208, 216, 217  
     Eastern 184  
     Gallo-Romance 61  
     Western 208, 214  
 Romanian 184, 186, 189, 193, 196, 198, 200, 202–205, 208, 214  
 Roshorv 46  
 Rushan 46, 48  
 Russenorsk 14  
 Russian 14, 29–31, 46, 48, 64, 71, 72, 74, 82  
     Literary 64  
 Rutul 29, 46–49, 59–61, 71, 72, 74, 81
- S**  
 Sabané 158, 161, 167  
 Sahu 110, 111, 114, 136, 138, 148  
 Salamāi (Mondé) 159  
 Samoyed 47, 48  
 Sanskrit 53  
 Sararé 158  
 Seget 110, 111, 148  
 Selaru 112, 121, 136, 148  
 Semitic 26, 29, 30, 48, 68, 81, 82  
     West-Central 30  
 Senagi 115  
 Sentani 130, 148  
 Sepik 108  
     Lower 108  
 Serbian 186  
     Southeastern 192, 198  
     Southern 192, 216  
     Standard 198  
     Timok 189  
     Torlak 186, 189  
 Serbo-Croatian 184, 192, 201  
 Sinitic 2, 223, 225, 227, 239, 271, 272  
 Sirionó 159, 161–163, 167  
 Skou 108  
 Slavic 14, 28, 29, 46, 48, 182, 184, 204, 205, 208, 211, 212  
     Balkan 186, 191–193, 200, 205, 209, 211, 212, 216, 217  
     East 29  
     South 192, 203, 204  
 Sougb (Manikion) 110, 111, 114, 115, 122, 134, 149  
 South Bird's Head (SBH) 110, 114, 115, 130, 147  
 South Caucasian (SC), *see* Kartvelian  
 South Halmahera/West New Guinea (SHWNG) 112, 113, 120, 137, 146–149  
 South-East-Interior (SEI) Papuan 102, 103, 109  
 Spanish 9, 10  
     Argentina 96  
     Caribbean 19  
     Latin America 19, 20  
     Peninsula 19  
 Suruí-Paitér 153, 154, 159  
 Svan 29–31, 46, 47, 49, 51, 53, 56, 60–63, 65, 66, 68–78, 81, 82  
     Bal 59, 60  
     Lashkh 59  
     Lentekh 59, 60  
     Upper Bal 59
- T**  
 Taba 104, 111, 112, 116, 121, 127–129, 132, 133, 136, 137, 149  
 Tabasaran 29, 31, 45, 47, 49, 54, 59, 60, 71, 72, 74, 79, 81  
     Dübek 48  
     North 72  
     South 71  
 Tacana 158, 161, 163  
 Tacanan 151, 156, 158, 170  
 Tai 2, 223, 226, 228, 271, 272  
 Tajik 48, 212  
 Talysh 29, 46, 48, 54, 58, 59, 67, 68, 72  
 Tamang 60  
 Tat 29, 32, 46, 48, 55, 58–60, 70, 72–74  
     Northern 32  
     Southern/Armenian 32  
 Tawandê 162  
 Tehit 110, 111, 114, 125, 149  
 Teiwa 110, 111, 114, 119, 124, 130, 136  
 Tetun 107, 111  
     Dili 112, 126, 132, 136, 149  
     Fehan 112, 126, 132, 136, 149  
 Thai 228  
 Thracian 189  
 Tibeto-Burman 60, 223  
 Tidore-Ternate 110, 111, 114, 119, 124, 125, 131, 136, 138, 149  
 Timor-Alor-Pantar 110  
     West 110  
 Tindi 29, 46, 47, 49, 58, 71  
 Tobelo 110, 120  
 Tokodede 111, 112, 132, 149  
 Toromona 163  
 Torricelli 108, 130  
 Trans New Guinea (TNG) family 107, 108, 110, 130, 146–149  
     Proto TNG 110  
 Trans-Fly 108  
 Trinitario 158, 161  
 Trukhmen 67  
 Tsakhur 29, 45, 47–49, 55, 59, 67, 71–74, 78, 79, 81  
 Tsez 29, 46, 47, 49, 59, 60, 63, 71–73, 78  
     Proto-Tsez 34–36  
     Proto-Tsez-Khvarshi 34–36  
 Tsezic 58–60  
 Tskhur Sabinchin 73  
 Tugun 112, 127, 149  
 Tukang Besi 112, 120, 122, 149  
 Tunguso-Manchu 47, 48  
 Tuparí 159, 167, 168  
 Tupian 151, 152, 156, 158, 159, 164, 168, 169  
 Tupí-Arikém 159, 170  
 Tupí-Guaraní 156, 158, 159, 162–164  
 Tupí-Mondé 153, 154, 159



- Tupí-Ramarama 153, 159, 168  
Tupí-Tuparí 159, 164, 170  
Turkic 26, 29, 30, 44, 46–49,  
51, 60, 65, 67, 68, 72–75,  
77, 79, 212  
North-Western  
(Kypchak) 30, 60  
South-Western (Oghuz) 30  
Turkish 29, 31, 43, 44, 46, 48,  
51, 52, 58–64, 67, 69, 70,  
72, 73, 75, 184, 186, 209,  
212, 216  
Anatolian 67, 68, 76, 83  
Balkan 184  
Meskhetian 71, 76  
Ottoman 69–71  
Rize-Trabson 67  
Turkmen 67
- U  
Ubykh 29–31, 42, 43, 45, 47,  
49, 52, 54, 56–58, 60, 68,  
69, 71, 73–75  
Udi 29, 46, 47, 49, 52, 59, 60,  
63, 64, 67, 70–74, 76–78  
Nidzh 49, 73  
Vartashen 49  
Udmurt 46, 48  
Ugric 48  
Uralic 26, 46–48, 53  
Urartu (Urartian) 53, 79  
Uru 167, 168, 170–172
- Uru-Chipayan 167  
Uru-eu-wau-wau (Amondawa,  
Jupa'ú) 159
- V  
Veps 46  
Vietnamese 2  
Vietnamese 221–225, 227–257,  
263, 265, 267–271  
Vod 48
- W  
Waimaha 112, 120, 136, 149  
Wakhi 46, 48  
Wandamen-Windesi 127, 132,  
134, 149  
Wari' (Pakaanova) 158, 160,  
165, 170, 171  
Waropen 111, 112, 121, 122, 127,  
129, 132, 133, 136, 137, 149  
Napan 133  
Wayuru (Ajurú) 159, 168  
Welsh 19, 47, 48  
Argentinian 19  
West Bird's Head (WBH) 110,  
114, 115, 146–149  
West Bomberai 110  
West Makian 104, 110, 111,  
114, 136, 138, 149  
West Tangaran 111  
West Caucasian 60, 77, 78,  
81, 83
- Proto-West-Caucasian 47  
Western Melanesian 137  
Western-Malayo-Polynesian  
(WMP) 146  
Wetar 136
- Y  
Yaghnobi 60  
Yakut 47, 48  
Yaminahua 158  
Yanaigua 163  
Yawa 104, 110, 111, 114, 118,  
124, 130, 149  
Yazguliarmi 60  
Yéli Dnye (Yele) 108  
Yenissey 48  
Yiddish 19  
Yoruba 61  
Yuki 158, 161, 164, 165, 171  
Yurakaré 156, 159, 161,  
165, 171
- Z  
Zhuang 2, 221–257, 267,  
269–272  
Buyi 226  
Héchi 233  
Northern 226  
Southern 226  
Standard 226  
Wumin 226  
Zoró 154, 159

# Author index

## A

Abaev, V.I. 53, 55, 62, 65,  
66, 77  
Abdokov, A.I. 34, 36  
Abdullaev, I.X. 70  
Adelaar, W.F.H. 156, 161, 163  
Aikhenvald, A.Y. 4, 6, 12, 27,  
52, 164, 168, 169  
Akan, E.B. 212  
Akekseev, A. 34, 80  
Aleksseev, M.E., 54, 60, 71  
Alves, P.M. 159  
Ammann, A. 6, 14  
Andaya, L.Y. 105, 106  
Angenot de Lima, G. 158  
Araujo, G.A. de 158, 161

## B

Baarda, M.J. van 138  
Bacelar, L. 159, 161, 168  
Baird, L. 114, 117, 132, 147  
Bakker, D. 172  
Bakker, P. 10, 14, 186  
Ballard, J.A. 15  
Baptista Morales, J. 156  
Barz, R.K. 19  
Baudouin de Courtenay,  
I.A. 26, 36, 37  
Bauer, R.S. 227  
Becker-Donner, E. 160, 162,  
168  
Belič, A. 90  
Bellwood, P. 100, 104, 105  
Bereznak, C. 6  
Berg, R. van den 120, 148  
Berry, C. 110, 114, 117, 124, 131,  
134, 146  
Berry, K. 110, 114, 117, 124, 131,  
134, 146  
Birdsell, J.B. 100  
Birnbaum, H. 186, 189  
Bisang, W. 254  
Blažek, V. 61, 81  
Bloomfield, L. 37

Blust, R.A. 112, 113, 116, 122,  
137, 146–149  
Boas, F. 5  
Bokarev, E.A. 59  
Borella, C. 158  
Boretzky, N. 186  
Bork, F. 81  
Boswood, J. 158  
Bowden, J. 116, 121, 127, 128,  
132, 149  
Braga, A. de Oliveira 159  
Brandes, J.L.A. 122  
Britain, D. 4  
Broch, I. 14  
Brotherson, A. 114, 115, 147

## C

Cadière, L. 231  
Cagov, M. 59  
Callaghan, C.A. 6  
Camp, E.L. 161  
Campbell, L. 5, 7, 96, 164  
Capell, A. 109  
Capelli, C. 102  
Carter, F.W. 181  
Caspar, F. 153, 161, 169  
Catford, J.C. 38, 40, 42, 43,  
47–50  
Čeerčiev, M.Č. 73  
Cereteli, G.V. 37, 45  
Chappell, H. 116, 226–228  
Charachidzé, G. 40, 80  
Chen, Ping 227  
Cheng, L.L.S. 224, 234, 256,  
258–261  
Chirikba, V.A. 1, 3, 6, 34, 36,  
54, 77  
Čikobava, A.S. 33, 40, 54,  
76, 77  
Clark, E.V. 125  
Clark, R. 113  
Colarusso, J. 81  
Collins, J.T. 120, 123, 127, 129,  
146, 147

Constenla, A. 6  
Cowan, H.J.K. 123, 127, 132, 149  
Creissels, D. 15  
Créqui-Montfort, G. de 160  
Crevels, M. 2, 152, 155, 156,  
160, 164, 167, 169, 170  
Crowley, T. 129, 134

## D

Dahl, Ö. 8  
Dài Qingxià 227  
Danielsen, S. 158, 161  
Deeters, G. 34, 37, 40, 77, 80,  
81, 84  
Denevan, W.M. 152, 153, 172  
Derbyshire, D.C. 6, 164  
Desnickaja, A. 188  
Dibirov, I.A. 72, 73  
Dimitriu, C. 214  
Dix Grimes, B. 105, 106  
Dixon, R.M.W. 12, 27, 96,  
164, 168, 169  
Dol, Ph. 114, 119, 124, 130, 148  
Domosileckaja, M.V. 190  
Donohue, M. 120, 134, 149  
Duffield, N. 222, 224, 230,  
241, 252–254, 257, 263  
Dumézil, G. 34  
Duran, I. Rodrigues 158  
Duridanov, I. 188  
Džavaxišvili, I. 76  
Džidalaev, N.S. 74, 78, 79

## E

Èdel'man, Dž.I. 63  
Edmondson, J.A. 226  
Emeneau, M.B. 6, 37  
Emkow, C. 158  
Enç, M. 262  
Enfield, N.J. 6, 222–225,  
228–234, 237, 238, 241,  
248–254, 262–265, 267, 270  
Engelenhoven, A. van 127,  
132, 147

Erickson, C. 153  
 Everett, D.L. 158, 160

## F

Fabre, A. 163  
 Faron, L.C. 153  
 Faust, N. 158  
 Fedorov, J. 30, 79  
 Feuillet, J. 190  
 Fine, J. 182  
 Flassy, Don A.L. 114, 149  
 Florey, M. 133, 146  
 Foley, W.A. 6, 104, 107–109,  
 113, 114, 148  
 Fonseca, J.S. da 162  
 Fortescue, M. 170  
 Foster, G.M. 18  
 Fraassen, C.F. van 105, 146  
 França, M.C. Victorino  
 de 158  
 Friedberg, C. 114, 115, 117,  
 130, 146  
 Friedman, V. 190, 212, 213  
 Fähnrich, H. 78, 79

## G

Gabas Jr., N. 159, 161, 162  
 Gabunia, Z. 40, 83  
 Gadžieva, N.Z. 67, 73  
 Galeote Tormo, J. 160  
 Gallis, A. 189  
 Galucio, A. Vilacy 159, 161,  
 162  
 Galvão, E. 152, 154  
 Gamkrelidze, T.V. 45, 80–82  
 Gamrekeli, V. 77  
 Gerhardt, L. 15  
 Geurtjens, H. 127, 128, 133, 147  
 Gijn, R van 160, 161  
 Gill, W. 160  
 Girard, V. 163  
 Golab, Z. 187, 188, 208  
 Goodman, M. 18  
 Gravelle, G. 114, 122, 125,  
 134, 148  
 Greenberg, J. 161  
 Greenberg, J. 7, 14  
 Greppin, J.A.C. 65  
 Grimes, C.E. 113, 116, 121, 123,  
 127, 128, 133, 138, 146  
 Grjunberg, A.L. 32, 55  
 Gudjedjani, Ch. 82  
 Guéron, J. 262  
 Guerra, M. de Lacerda 159

Guillaume, A. 158, 161, 168  
 Gumperz, J.J. 4, 10  
 Guzman-Tirado, R. 40, 83  
 Güldemann, T. 15

## H

Haan, J.W. 120, 124, 131, 146  
 Haas, M. 5, 6  
 Harris, A.C. 54, 80  
 Haspelmath, M. 6, 12, 13, 17  
 Hasselt, F.J.F. van 121, 126  
 Haude, K. 160, 161, 168  
 Heine, B. 116, 125  
 Held, G.J. 122, 127, 133, 149  
 Hemmauer, R. 159  
 Henderson, J.E. 108  
 Hesseling, D.C. 18  
 Heuvel, W. van den 126, 146  
 Hewitt, B.G. 54, 77, 79, 80  
 Himmelmann, N.P. 97, 98,  
 123, 138, 139  
 Hindrichs, U. 181  
 Hinton, B. 127, 149  
 Hinton, L. 6  
 Hoekstra, T. 258, 259, 262  
 Hoerder, D. 11  
 Holm, J. 18  
 Holton, G. 120  
 Hornborg, A. 172  
 Huáng, Xíng 226  
 Huang, Yuanwei 227  
 Hudson, R. 10  
 Hughes, J. 127, 132, 147  
 Hull, G. 120, 121, 132, 147, 149

## I

Ibragimov, M.-R.A. 72–74  
 Iggesen, O. 158  
 Ivanov, V.V. 34, 45, 80–82  
 Ivarsson, S. 228

## J

Jacobs, H.Th.Th. 105  
 Jahr, E.H. 14  
 Jakobson, R. 27, 40  
 Jones, L.K. 114, 118, 124, 130, 149  
 Joseph, B. 189  
 Judy, J. 161  
 Judy, R.A. 161

## K

Kaartinen, T. 127, 129, 146  
 Kakabadze, S.N. 76

Kaufman, T. 6, 163  
 Kayser, M. 102  
 Keraševa, Z.I. 59  
 Kerke, S.C. van de 160  
 Kern, B. 158, 160  
 Key, H. 160, 161  
 Key, M.R. 163  
 Kiazimov, I.B. 72  
 Klammer, M. 1, 6, 97, 113, 114,  
 121, 124, 127, 132, 146,  
 147, 149  
 Klein, H.E.M. 19  
 Klein, W. 262  
 Klimov, G.A. 28, 36–40, 47,  
 52, 54–56, 63–66, 72, 76,  
 77, 79–81  
 Klinken, C. van 126, 132, 149  
 Kodzasov, S. 34, 43, 50  
 Kopitar, J. 181, 189  
 Koptjevskaja-Tamm, M. 6,  
 11, 14  
 Kortlandt, F. 45, 81  
 Kratochvíl, F. 114, 146  
 Kroeber, A.L. 5  
 Kroecker, M. 158, 160  
 Kuipers, A.H. 34  
 Kumaxov, M.A. 34, 60

## L

Lamarre, C. 239  
 Lan Qingyuan 226  
 LaPolla, R. 227  
 Laycock, D. 132, 135, 148  
 Leer, J. 6  
 Leonel, M. 172  
 Leslau, W. 6  
 Lévi-Strauss, C. 152, 154  
 Lewy, E. 37, 40  
 Lǐ Jǐnfāng 226  
 Liccadi, M.R. 161  
 Lichtenberk, F. 118, 120, 122,  
 129  
 Lindstedt, J. 186–189, 217  
 Lipski, J. 18  
 Lomtadze, K. 33  
 Loos, E.E. 158, 163  
 Lowe, I. 162  
 Luong, Van Ke 228  
 Lynch, J. 116, 129, 134, 137, 138

## M

Mačavariani, G.I. 38, 45, 80  
 Madieva, B. 31

- Magometov, A.A. 59, 73  
Maldi, D. 152–154  
Manoliu-Manea, I.M. 208  
Marr, N.Y. 77  
Martinet, A. 27  
Martirosyan, H. 55  
Masica, C. 6, 26  
Massey, D. 4  
Mastumura, M. 132, 147  
Matisoff, J.A. 224, 231, 240  
Matras, Y. 12, 186  
Matsumura, T. 132, 147  
Matthews, S. 239, 255  
McGregor, W. 116  
Meer, T. van der 159  
Megreldze, I.V. 75, 76, 79  
Meillet, A. 54  
Melikišvili I.G. 43, 50  
Menick, R.H. 114, 125  
Métraux, A. 172  
Michaelis, S. 17  
Migge, B. 17  
Mikailov, Sh.I. 44, 79  
Miklosich, F. 181, 189  
Minde, D. van 114, 146  
Mladenov, M. 190  
Monserrat, R.M. Fonini 159  
Moore, D. 152, 159, 160  
Mufwene, S.S. 19  
Musaev, M.M. 72  
Muysken, P.C. 6, 10, 17, 152, 163, 167, 170
- N**  
Needham, R. 10  
Nguyễn, inh-Hoà 225, 228, 233  
Nichols, J. 4, 28, 30, 32, 102–104, 107, 109, 116, 138, 167, 172  
Nikolayev, S.L. 34, 36, 81  
Nimuendajú, C. 161, 162, 167  
Nitbani, S.H. 114, 130, 147  
Nordenskiöld, E. 160, 162  
Norman, J. 228  
Nurmagomedov, M.M. 73
- O**  
Odé, C. 114, 118, 134, 135, 148  
Olza Zubiri, J. 158  
Orbigny, A. d' 155
- Owens, M. 132, 146
- P**  
Palmitis, M.L. 82  
Pasveer, J. 100  
Pawley, A.K. 110  
Payne, D.L. 6, 154, 168  
Pires, N. Nascimento 158  
Pisani, V. 37  
Pisowicz, A. 43, 45, 79  
Plotnikova, A.  
Polák, V. 37  
Price, P.D. 160, 169  
Provasi, E. 62
- Q**  
Qín Guósheng 226  
Qín Xiàoháng 222, 224, 227, 229, 233, 234, 236, 237, 239–243, 245–248  
Queixalós, F. 167
- R**  
Ramat, P. 14  
Ramsey, S.R. 226, 227  
Rankin, R.L. 102  
Reesink, G.P. 1, 6, 110, 114, 115, 118, 122, 125, 130–134, 147, 149  
Reid, A. 105  
Reiter, N. 181  
Remijsen, B.C. 134, 135, 147, 148  
Renault-Lescure, O. 167  
Ribeiro, E. 161  
Ricardo, C.A. 162  
Rivero, M.L. 202  
Rivet, P. 160  
Rogava, G.V. 33, 57, 80  
Rose, F. 158  
Ross, M.D. 104, 109, 110, 113, 115, 116, 122, 129, 134, 138
- S**  
Šagirov, A. 34  
Sakel, J. 160, 161, 168  
Salvatierra, C. 158  
Sampaio, W.B. de Araujo 159  
Sandfeld, K. 186–189  
Santos, M. dos 162, 163  
Šaradzenidze, T. 26  
Schaller, H.W. 186–188  
Schmidt, G. 76, 80
- Schmidt, K.-H. 34, 79  
Schuchardt, H. 77  
Schulze, W. 52  
Sherzer, J. 5, 6  
Shimomiya, T. 40, 51, 66, 69, 80  
Siegel, J. 19  
Siewierska, A. 172  
Silva, L. de Jesus 158  
Silzer, P.J. 121, 127, 129, 132  
Simpson, A. 225, 254, 257, 270  
Smith, N.S.H. 6, 17  
Sobolev, A. 189, 190  
Soh, Hooi Ling 255  
Solnit, D.B. 226  
Solta, G.R. 186–189  
Sorensen, A.P. 164, 169  
Squartini, M. 214  
Staden, M. van 1, 6, 106, 119, 120, 124, 131, 138, 149  
Starostin, S.A. 34, 36, 81  
Steinhauer, H. 114, 119, 126, 146  
Steinke, K. 190  
Steward, J.H. 153  
Stokhof, W.A.L. 114, 115, 124, 149  
Stolz, Th. 2, 14  
Storto, L. 154, 159, 161  
Svorou, S. 117  
Sybesma, R. 2, 6, 224, 234, 256, 258–262, 264, 265
- T**  
Tang, Sze-Wing 256  
Tekin, T. 212  
Telles, S. 158, 161, 162  
Tenny, C. 246  
Testelec, J.G. 34, 55  
Thomason, S.G. 2–4, 97  
Thompson, L.C. 240, 263  
Tikkanen, B. 53, 62  
Tomic, O.M. 2, 191, 198, 204, 209  
Tosco, M. 6  
Travis, L. 262  
Trombetti, A. 81  
Trubetzkoy, N.S. 3, 6, 26, 33, 34, 36, 40, 42, 50, 55, 181  
Tryon, D.T. 113  
Tsai, Wei-tien D. 260  
Tuite, K. 26, 37, 38, 40, 47, 52, 80  
Turčaninov, G. 59

U

Uturgaidze, T. 77, 79

V

Valentyn 106

Vanden Wyngaerd., G. 262,  
264

Vasconcelos, I.P. 159

Veen, H. van der 123, 131

Veth, P. 100

Villafañe, L. 158

Visser, L.E. 138

Vogt, H. 39, 40, 45, 65, 66, 72,  
76, 77, 79–81

Voorhoeve, C.L. 104, 109, 110,  
114, 115, 138, 147–9

Voort, H. van der 2, 154, 158,  
159, 161, 164, 168, 169

Vries, L. de. 110, 114, 115, 117,  
130, 147

Vuillermet, M. 158

W

Wáng Jūn 226, 227

Wěi Qìngwěnn 226

Wei Xingyun 226

Wälchli, B. 6, 11, 14

Weigand, G. 186, 189

Weinreich, U. 181

Wetzels, L. 152

Williams-van Klinken,  
C. 126, 132, 149

Wilson, R. 4, 10

Wimbish, S.G. 114, 131, 138, 148

Winford, D. 9

Wise, M.R. 168

Wolf, H.E. 16

Wolff, E. 15

Wu, Yunji 235, 261

Wurm, S.A. 107, 108, 110

X

Xajdakov, S.M. 59

Y

Yip, V. 239, 255

Yue, A. 239

Yue-Hashimoto, A. 226, 227

Z

Zanten, E. van 134

Zhāng Yuánshēng 222,  
226, 229, 234, 237, 240,  
246–248

# Subject index

## A

ablaut 41  
 ACQ 221–272  
 active-stative 103  
 adjacency 236  
 adjectival verbs 15  
 adjectives shape 61  
 adpositions  
   postpositions 41, 103, 113, 171  
   prepositions 113, 187, 188,  
     192–195  
 adverb 230, 238  
 agentivity  
   agentive 230  
   non-agentive 229  
 agglutination 41, 51, 53  
 agreement strict 13  
 agriculture 32, 152  
   irrigation 153  
   raised field 153  
   rice 226  
   slash and burn 152, 153  
 Agta 100  
 Alexandrian period 14  
 alignment  
   ergative 38, 41, 54, 103, 109  
   nominative-accusative 171  
 allogetic relationship 37  
 animacy  
   animate 57, 125  
   human 196–198  
   inanimate 57, 125  
 anthropology 30  
 anti-causative 202  
   prominence 13  
 applicative 167–169  
 archeology 8, 30, 100, 102,  
   154, 166, 173  
 Arctic 13  
 area  
   bottom-up approach 7  
   circumstantialist 4  
   contact superposition zone,  
     *see* multi-tier alliance

contiguous language 68, 69  
 core languages 3, 65, 188  
 historicist approach 4  
 lower-level alliance 68, 69  
 macro 3, 5  
 marginal 67  
 meso 3, 5  
 micro 3, 5  
 multi-tier alliance 11, 68  
 peripheral 3, 65, 66  
 sub-area, *see* sub-regional  
   alliance  
 sub-network, *see* sub-  
   regional alliance  
 sub-regional alliance 12,  
   69, 70, 164, 167, 168  
 top-down approach 7  
 articles  
   definite 13  
   indefinite 13  
 aspect 20, 223, 231  
   achievement 224, 229,  
     236, 251  
   durative 55, 236–38  
   preverbal 19  
   repetitive 206  
   resultative 208, 222, 234, 235  
   system 52  
 Austro-Hungarian 182  
 Austronesians 104, 105  
 auxiliary  
   ‘be’ 206  
   ‘have’ 214

## B

Balkan language first, second,  
   third degree 186  
 Balkanism 186, 188–190  
   primary 187  
   secondary 187, 188  
 Balkanologist 187  
 Bible translation 106  
 bilingualism, *see* multilin-  
   gualism

bisyllabic morphemes 113  
 body parts 116  
 borrowing 9, 12, 16, 25, 39, 62,  
   63, 112  
   hierarchies 5  
 bronze techniques 226

## C

*cacique* 163  
 calques 12, 14, 25, 39  
 case 51  
   accusative 103, 192, 195  
   dative 192–195, 204  
   genitive 120, 123, 128,  
     192, 194  
   instrumental 192  
   locative 167  
   loss 192  
   merger dative/genitive 187,  
     188  
   merger location/  
     direction 187, 188  
   nominal 2, 191  
   nominative 192, 193  
   oblique 191, 192  
   prepositional 187, 188,  
     192–195  
   pronominal 2  
   vocative 187, 188, 192  
 categories  
   nominal 38  
   verbal 38  
 Caucasology 33, 34, 37  
 causative 103  
   morphological 41, 52  
*chicha* 143  
 Christian 32, 106  
   Catholicism 106, 182  
   Orthodox 32, 182  
   Protestantism 106  
 classifier  
 classifiers 2, 167–169, 172  
   numeral 103, 104,  
     108

- possessive 104
- verbal 171
- clause chaining 108
- clause linkage 38
- clauses relative 13, 124, 125
  - subordinate 187, 188
- clicks 14
- clitics 196
  - agreement 202, 216
  - dative 191, 216
  - doubling 2, 191, 194, 196–200
  - doubling objects 187, 188
  - genitive 127
  - impersonal 2
  - possession marker 2, 187, 188, 191, 200–202, 216
  - proclitic 199
  - reflexive 191
- code-switching 10
- colonial 12, 16
- common semantic patterns 41
- comparative grammar 33
- comparative method 107
- comparative 15
  - particles 13
  - analytic 187, 188
- comparison, *see* comparative
- comparison, *see* comparative
- complementizers 113
- compound
  - echo-compounding 55, 56
  - bahuvrihi* 55
- concordances, *see* parameters
- connected narrative 210
- consonant 42–49, 65
  - inventory 41, 43–49
  - consonant-type
    - languages 43
  - cons affricates 41
  - back 42
  - back palatalized 42
  - dental 42
  - dental labialized 42
  - fricative interdental 42
  - harmonic clusters 41
  - interdental fricatives 42
  - labial 42
  - labial pharyngealized 42
  - laryngeal 41
  - laryngeal features ternary
    - contrast 44
  - laryngeal labialized 42
  - lateral 42
  - middle 42
  - middle labialized 42
  - obstruent 43
  - palatal 42
  - palatalized 27
  - pharyngeal 41
  - postvelar 41
  - postvelar 47, 48
  - resonant glottalized 42
  - resonants 48, 49
  - retroflex 27
  - sibilant 41
  - sibilant 45–47
  - voiced aspirated 42
  - voiced aspirated
    - consonants 42
- contact scenarios 9, 97
- contract labor 17, 19
- convergence 25, 26, 38, 39, 189, 190, 217, 228
  - gradual 12
  - grammatical 9, 14
  - unilateral 28, 136
- coordination markers on
  - each conjunct 41
- creole 11, 17–19
- cross-reference 171
- cultural complex 152, 153
  - Guaporé 153, 167
  - Ji-Paraná 154
  - Marico 152–154
  - Moxos 167
  - Tapajós-Madeira 154, 166, 167
- culture area, *see* cultural complex
- D**
  - definite 125, 196, 198, 200
    - article postpositive 187
  - degree 237–239
  - deictic system 41
    - binary 41
    - ternary 41, 57
  - demography 154, 156
    - restructuring 16
  - desiderative 203
  - dialect 8, 19, 162, 189–191, 198, 200, 205, 217
  - diasystem 191
  - diffusion 39, 112, 113, 115, 163–6, 169
  - diglossia 227
  - diphthongs 41, 113
    - lack of 41
  - directionals 2, 167, 170, 171
  - displacement 156
  - dolmen 30
  - DP 190, 194, 215
  - drift 189
  - E**
    - education 106, 107, 228
    - embeddedness paradox 6
    - enclitic 188
    - Enlightenment 13
    - epidemics 160, 162
    - equative, relative-based 13
    - ethnography 160
    - etymology 34
    - evidentials 2, 41, 52, 167, 170, 171, 187, 188, 191, 204, 205, 209–217
      - admirative 210–212, 214, 215
      - analytic 211
      - circumstantial
        - deduction 210, 214
      - reported 210, 214
      - synthetic 211
    - exclamative 245
    - experience 208
    - extent 237–239
  - F**
    - focus 196, 197, 199
    - FUNAI 162, 163
    - fusion 51
    - future analytic 187, 188
  - G**
    - gender 109, 116, 117, 126, 164, 170
      - feminine 193
      - masculine 193, 196, 215
      - merger 191, 195, 196
      - neuter 119, 196
    - genealogy 123, 127
    - genetic taxon 28, 36
    - genocide 163
    - geolinguistics 190
    - glide 113
    - globalization 11
    - glottalization 27, 36, 38, 41, 44, 45
    - glottochronology 82
    - grammaticalization 171, 196, 197, 212, 217

**H**

Habsburg Empire 14  
 Hanseatic League 14  
 harmonic clusters 41, 50  
 Hindu 105  
 history 96, 98, 100, 104, 138,  
     154, 182  
     ethnohistory 152  
     social 11  
 human genetics 102  
     Y chromosome 102  
 hunting 32  
 hunting and gathering 152  
     whale 13

**I**

Ibero-Caucasian 33  
 implosive 14  
 indefinite 125  
 Indonesian Republic 107  
 infinitive 187, 188  
 innovation linguistic 128  
 intensifier 13  
 intermarriage 169  
 intertwining 10  
 intonation 27  
 invasion 16  
 inversive  
     construction 41  
     verbs 41, 56  
 Islam, *see* Muslim  
 isogloss 8, 26, 27, 39, 40  
 cultural 39  
 isolates 95, 114, 115, 151, 152,  
     160, 161, 164, 167, 169

**J**

jaguar cult 153

**K**

kinship 61, 116, 127, 191,  
     200–202  
 koine 19, 71

**L**

language  
     death 96, 155, 161, 162  
     elite 16  
     endangered 161  
     family 25, 26, 37, 107  
     heritage 16  
     lexifier 17  
     mixing 107

national 16, 106, 107  
 replacement 39, 104  
 second 213  
 shift 10, 97  
 standard 191  
 phonological union 40  
 leveling 14, 16, 17, 19  
 lexical correspondence 34,  
     163  
 lexical semantics 61, 62  
 lexicon, *see* vocabulary  
 lingua franca 12, 15, 16, 30, 31,  
     71, 72, 74, 105, 106  
 linguistic community  
     secondary 26–38  
     primary 26  
 linker 224, 231–239  
 liquid 108, 113  
 Lower Palaeolithic 182

**M**

macro features 170  
 majority language 186  
 Malayo-Muslim trading  
     network 105  
 manner 237–239  
 markedness paradox 8  
 marking  
     dependent 103  
     head 103, 108, 164, 170, 171  
 Maykop culture 30  
 media 228  
 Mennonite 19  
 Middle Ages 13, 14, 184  
 migration 12, 100, 104, 109,  
     112, 136, 162  
     labor 17  
 minority language 186  
 mission 106  
     Jesuit 152, 154–156, 162  
     New Tribes 160, 161  
     Summer Institute of  
         Linguistics (SIL) 160, 161  
 modal 2, 221, 223, 231, 233, 234  
     ability 222, 235, 242–246  
     deontic 223, 231–233, 235  
     irrealis/future 237  
     permission ('okay') 234, 247  
     possibility (-able) 239–242  
     potential 234, 243  
 monosyllabic morphemes 223  
 mood  
     optative 187, 188  
     potential 41

potential 41, 52  
 subjunctive 187, 188  
 morphology  
     agglutination 41  
     analytic 20, 187  
     asymmetrical 170, 171  
     bound 110  
     isolating 223  
     limited 20  
     nominal complexity 108  
     prefix 2, 103, 167, 170, 171  
     prefixal conjugation 41  
     suffix 170, 171, 229, 230  
 Mountain Jews 32  
 multilingualism 30, 39, 95,  
     105, 169, 189  
     active 30  
     community-level 8  
     degree of 7  
     stable 9  
     symmetrical 31, 217  
     unilateral 31  
 Museu Goeldi 162  
 Muslim 15, 32, 105, 106  
     Shia 32  
     Sunni 32

**N**

nasal 113  
     clusters 14  
     harmony 167, 171  
     uvular 42  
     velar 42  
 nautical language 14  
 negation 113, 130–134, 236,  
     248, 249  
     adverb 130, 131  
     clause final 1, 95, 98, 116,  
         131–133, 135, 137, 138  
     clause initial 132  
     double 131  
     particle 130  
     post-predicate 114,  
         130–134, 137, 138  
     pre-predicate 95, 98, 131,  
         132  
     pre-verbal, *see* pre-predicate  
     verbal 13, 138  
 network 4, 14  
 nominalization  
     masdar 41, 57  
 nominative experiencers 13  
 Norman 12  
 noun class 103, 108, 109



- Noun Phrase 231  
 number 119, 126  
   dual 103  
   group inflection 41, 55  
   merger 191, 195, 196  
   nominal 2, 108, 167, 171  
   plural 8, 51, 103, 196, 215  
   singular 103, 196, 202  
   trialis 8  
   verbal 2, 164, 167, 170–172  
 numeral system 9, 98  
   decimal 58–61  
   vigesimal 41, 58–61  
   distributive 61
- O**  
 object  
   direct 194, 196–198, 203, 241  
   indirect 193–196, 199  
 one way contact, *see* unilateral convergence  
*orang asli* 100  
 order 9, 241  
   Attr-N 41  
   free 56  
   head final 108, 114  
   head initial 113  
   head-modifier 171  
   noun phrase 98  
   possessor possessed 41, 56–57  
   postposition 41, 53  
   postverbal modal 221  
   preposition 53  
   preverbal modal 224, 233  
   SOV 41, 98, 108, 113, 114, 124, 125, 129, 130, 133, 135, 138, 139  
   SVO 1, 19, 95, 97, 113, 114, 124, 130, 133, 139, 223  
   verb initial  
   V-initial 98, 104, 138, 164  
   V-second 98  
 orthography 106  
 Ottoman 182
- P**  
 parameters 26, 186  
 part-whole relations 116, 127  
 passive participle 206, 208  
 perfect 54  
   analytic 187, 188  
   periphrastic, *see* analytic
- person  
   first 103, 201  
   inclusive/exclusive 1, 2, 95, 97, 102, 108, 113–115, 135, 138, 139, 167, 170, 171  
   marking prefixes 98  
   marking proclitics 98  
   second 104, 201  
   third 201  
 phoneme inventory 108  
 phonetic correspondence 36  
 phonostatistics 50  
 phonotactics 50, 51  
 phraseology 27, 41, 64  
 phylum 107  
 pidgin 11, 12, 156, 160  
 pitch-accent 134  
 Pleistocene 100  
 polysynthetic 41, 52, 171  
 possession  
   adnominal, *see* attributive  
   affixes 128–130  
   alienable/inalienable 1, 95, 98, 115–130, 135, 164, 170, 171  
   attributive 127, 135  
   classifier 118, 137  
   dative external possessive 13  
   direct 118–122, 137  
   indirect 118–122, 137  
   ligature 118, 121, 123, 125, 128–130  
   marked on possessum 170, 171  
   marker prefixed 113  
   marker suffixed 113  
   marking 115–128  
   predicative 127  
   pronoun 119, 120, 123, 216  
   reversed genitive 123, 126, 127  
   word order 1, 41, 95, 97, 113, 115, 118–120, 123–129, 135, 137  
 prefixal conjugation 41, 54  
 prestige 8, 216  
 preverbs  
   directional 41  
   orientational 41  
 pro-drop 19  
 pronoun 110  
   interrogative 57  
   free 115  
   negative 13
- relative 13  
 proto-language 36
- Q**  
 quantifiers 98
- R**  
 railway 162  
 reconstruction 107  
 reduplication 55, 113  
   plurals 103  
 reflexive 13  
 Reformation 14  
 regularity paradox 7  
 relator 124  
 relexification 10  
 Renaissance 13, 14  
 restructuring 20  
 rich consonantism 41  
 rich postvelar (uvular, pharyngeal and laryngeal) systems 41  
 rich sibilant systems 41  
 Roman 12, 13, 182  
   post-Roman 12  
 rubber boom 155, 163  
 rural transhumancy 213
- S**  
 scope 252  
 second position 198  
 semantic patterns 41  
 semantically empty noun  
   formative 167–169  
 shamanism 152  
 simplification 11, 16, 19  
 slavery 15–20, 105, 106, 136, 154  
 sociolinguistics 191  
 sociology 162  
 sound correspondences 181  
 spatial relations 116  
 specificity 196, 197, 199  
 spices 105  
 stops 41, 103, 113  
   stops and affricates ternary contrast 41, 43  
   labial-velar 14  
 stress 134  
   dynamic 52  
 subject  
   sentential 249  
   marking 110  
   prefix 108

- suffix 108  
 subordination 20, 171  
 subordination  
   analytic 20  
   nominalization 171  
 substrate 25, 39, 75–79, 113,  
   139, 189  
 superstrate 9, 10, 17, 75–79,  
   189  
 suppletion  
   adjectives 61  
   verbs for singular and  
     plural arguments 41  
 switch reference 108, 171  
 syllable 134  
   structure 51
- T**
- TAM 108
- telicity  
   atelic 244, 248–250  
   resultative 242, 245  
   telic 242, 245, 246
- tense  
   ‘be’ perfect 204–206, 208,  
   209  
   ‘have’ perfect 13, 187, 188,  
   191, 204–206, 208, 209  
   aorist 51, 204  
   compound perfect 187,  
   188  
   future in the past 187, 188  
   future perfect 187, 188  
   imperfect 187, 188, 204  
   inverted perfect 212, 214  
   perfect 204, 209, 216  
   present 51  
   preverbal 19  
   prior event 231, 232  
   system 52  
   time depth 8  
   Tobler-Mussafia rule 199,  
   200  
   tone 15, 103, 104, 108, 109, 116,  
   134, 135, 223, 228, 233  
   contour 134  
   level 134  
   lexical 134  
   zero 233  
   topicalization 196, 198, 199  
   trade 12, 32, 104–106, 72, 213  
   traditional religion 32  
   transfer 17  
   transitivity  
   ambitransitive (labile)  
     verb 41  
   derived intransitivity 103  
   intransitive verbs 51, 208,  
     209  
   transitive verbs 51,  
     208–209  
   transitive with indirect  
     object 56  
   typology 13, 36–41, 97, 102,  
   107, 113, 123, 127, 139, 189,  
   191, 192, 209, 215, 223
- U**
- unclassified language, *see*  
   isolate
- Uniformitarian principle 9
- universal development 11, 42
- urban contact vernacular 16
- urbanization 16, 17, 19, 153
- uvular 41, 42  
   labialized 42  
   nasal 42  
   palatalized 42  
   pharyngealized 42
- V**
- variation 229
- velar 42  
   labialized 42  
   nasal 42  
   palatalized 42
- verb complex morphology 167  
   dynamic 41, 57  
   inversive 41  
   object pronominal  
   prefix 108  
   perfect 2  
   prefixes 115  
   stative 41, 57  
   verbal noun 41  
   suppletive  
   vertical dependency 31  
   vocabulary 5, 27, 34, 36, 110,  
   227  
   basic 34, 113  
   collocations 230, 231  
   common cultural terms 41,  
   63, 64  
   comparison 107  
   core 164  
   loanword 17, 227, 228  
   shared 167  
   spread 12  
   voice alternations  
   asymmetrical 98  
   symmetrical 97, 98  
   voice  
   benefactive passive 231–233  
   impersonal 202–204  
   inversive 41  
   medio-passive 202  
   participial passive 13  
   passive 103, 202–204  
   vowel 8, 113, 228  
   ablaut 41  
   clusters 41  
   harmony 170, 171  
   inventories 49, 50, 106, 162  
   lack of vocalic clusters 41  
   lower high 15  
   schwa 41, 49, 50
- W**
- Wallace line 100
- Z**
- zone  
   residual 32, 167, 172  
   spread 167



## Place index

### A

Abk Gal 69  
 Abk Samyrzaqan 69  
 Abkhazia 32  
 Adriatic Sea 182  
 Adzharia 31, 32, 75, 76  
 Africa 12, 14–16  
     Southern 16  
     West 16, 18  
 Alania 75, 76  
 Albania 182, 184, 212, 215  
     Alba Hereti 76  
     Gardabani 76  
     Kakheti 76  
 Alor 97, 99, 102, 105, 109–111,  
     114, 115, 117, 119–122, 124,  
     127, 128, 132–37, 139  
 Amazon River, *see* Amazonia  
 Amazonas 154, 162  
 Amazonia 4, 6, 151, 163, 164,  
     166, 170  
 Ambon 105, 106, 111  
 Andaman Islands 100  
 Andes 172  
 Angola 16  
 Antilles 6  
 Argentina 19, 96  
 Armenia 32, 56, 79  
 Aru Islands 100, 105, 106, 109, 111  
 Asia 17  
     Minor 75  
     South 17  
     South East 1, 2, 96, 100, 223  
 Assam 225  
 Atlantic 11–20  
     North 12  
 Australia 100, 102–104, 107  
 Azerbaijan 32, 56, 78  
     Baku 32  
     Lenkoran 32

### B

Bacan 105  
 Bali 100, 105

Balkan 1, 4, 6, 14, 26, 28, 52,  
     56, 181–219  
 Baltic 6, 8, 12  
 Banda 105, 106  
 Banggai 105  
 Baures River 162  
 Belém 162  
 Belize 18  
 Beni 151, 154  
 Beni River 151  
 Benin 6  
 Biak 104, 111, 121  
 Bima 132  
 Bird's Head 99, 100, 102, 105,  
     106, 109–111, 114, 115,  
     122–125, 128–139  
 Bismarck archipelago 100  
 Bitola 195, 196, 209, 213  
 Bolivia 1, 19  
 Bomberai 109–111, 132, 139  
 Bosnia 184, 185  
 Branco River 153  
 Brazil 17, 18, 151–173  
 Buenos Aires 19  
 Bulgaria 184, 185, 199, 212  
 Buru 100, 104, 105, 111

### C

Caribbean 12, 16–20  
     Anglo-Caribbean 18  
 Carpathian region 25  
 Caucasus 1, 6, 26–84  
     Western Transcaucasus 29  
 Cayenne 18  
 Cenderawasih 110, 111, 114,  
     118, 121, 122, 124, 129, 130,  
     132, 134–137  
 Central America 5, 6  
 Chad 15  
 Chechnya 32  
 China 19, 105, 223, 225–28  
     Southern 2, 6, 221  
 Chipinguaia 163  
 Circumpontic 52

Clear Lake 6  
 Colombia 6  
 Colorado River 162  
 Constantinople (Istanbul) 182  
 Corumbiara River 153  
 Croatia 182, 184  
 Cuba 18

### D

Daghestan 31, 32, 70–74, 78  
     Charodin 31  
     Derbent 32  
     Samur 72  
 Dantzig 14  
 Danube 181, 182, 190  
 Dutch Antilles 18  
     Aruba 18  
     Curaçao 18

### E

East Nusantara 1  
 Ecuador 6  
 Elbasan 212  
 Ethiopia 4, 6, 15  
 Eurasia 62  
 Europe 6, 12–14  
     Western 12

### F

Fak-Fak 106  
 Flores 99, 132  
 French Cameroon 16

### G

Georgia 30, 32, 70, 72, 75–78  
     Artaani 67  
     Artanudzhi 67  
     Artvin 67  
     Dvaleti 77  
     Kars 67  
     Lechkhumi 76  
     Shavsheti 67  
     Tiflis 32

Gorna-Belica 215  
 Great Basin 6  
 Greece 184, 185, 195  
 Guajar  -Mir  m 151  
 Guangdong 223, 225–27  
 Guangxi 223, 226, 227  
 Guangzhou (Canton) 227  
 Guapor   1, 2, 6, 151–173  
 Guarayos Forest Reserve 163  
 Guayaramer  m, *see*  
     Guajar  -Mir  m  
 Guianas 18, 162  
 Guinea 15  
 Guizhou 226  
 Guria 76  
 Guyana, *see* Guianas

## H

Haiti 18  
 Halmah Jailolo 105  
 Halmahera 97, 99, 100, 102,  
     104, 105, 109–111, 114, 123,  
     128, 132–138  
 Hercegovina 184, 185  
 Hindu-Kush 60  
 Hong Kong 222, 227  
 Hopa 67, 68  
 Hunan 226  
 Hungary 184

## I

Imeretia 76  
 India 4, 6, 27, 56, 60, 96, 105,  
     225  
     North 19  
 Indo-China 2, 6, 221  
 Indonesia 1, 6  
     Eastern 95  
     Java 19  
 Ingushetia 32  
 Iran 56  
 Irian Jaya, *see* Papua  
 Italy Southern 61  
 Ivory Coast 16

## J

Japan 60  
 Java 105  
 Juruna 166

## K

Kai, *see* Kei  
 Kalahari 15

Kalimantan 100  
 Karachay-Cherkes Republic 31  
 Kei Islands 106, 109  
 Kisar 111  
 Kolkhian lowlands 82  
 Kolkhida lowlands, *see*  
     Kolkhian lowlands  
 Kosovo 184, 185  
 Kru  evo 209  
 Kuban valley 32  
 Kumyk 44

## L

La Paz 163  
 La Paz Iturralde Province 163  
 Laos 223, 225, 228  
 Lerin (Florina) 195, 196  
 Lesser Sunda Islands 100  
 Leti 100, 111  
 Lianshan 225  
 Lombok 105  
 Lower Chubut Valley 19  
 Lowland, *see* Amazonia  
 L  beck 14

## M

MacCluer Gulf 106  
 Macedonia 184, 185, 199, 215  
     Pirin 199  
 Madeira River 151, 166  
 Magdalena 163  
 Makian 104, 105, 111, 132  
 Malaysia 100  
 Malkar 44  
 Maluku 134  
 Mamor   1, 2, 6, 151–173  
 Manam 104  
 Manaus 151  
 Manus 104  
 Mato Grosso 154  
 Mediterranean 6, 12, 14, 26  
 Megrelia 30  
 Mekens River 153  
 Melanesia 102  
 Mesoamerica 5, 6  
 Metohia 184, 185  
 Mindanao 105  
 Misool 100  
 Moluccas 97, 99, 100, 102,  
     104–106, 110, 116, 121,  
     123, 124, 127, 128, 133, 135,  
     136, 139  
 Montenegro 182, 184  
 Montreal 19

Mosul 45  
 Moti 105  
 Moxos 153, 154, 164  
 Mozambique 16

## N

New Guinea 96–139  
 New World 17, 20  
 North America 12, 16, 18  
     Northern Northwest  
     Coast 6  
     Northwest Coast 6  
 North Sea 14  
 Nusa Tenggara 102  
 Nusantara (East) 95–139

## O

Obi 100  
 Oceania 116  
 Ohrid 200, 205, 212  
 Ossetia 32

## P

Pacific Hinterland 103, 104,  
     109  
 Pacific Rim 103, 104, 109  
 Pamir 60  
 Pantar 97, 99, 102, 105,  
     109–111, 114, 115, 119–122,  
     124, 127, 128, 130, 132, 133,  
     135–137, 139  
 Papua 1, 6, 61, 97, 106, 107, 109  
 Papua New Guinea, *see* Papua  
 Par   154  
 Paraguay 19  
 Peking (Beijing) 227  
 Phasis, *see* Rioni  
 Philippines 100  
 Pimenta Bueno River 153, 154  
 Plains 6  
 Plateau 6  
 Preslav (T  rnovo) 212  
 Prilep 206  
 Pueblo 6  
 Puerto Alvoroa 151  
 Puerto Araona 163

## Q

Quebec 19

## R

Raja Ampat Islands 100,  
     111, 135  
 R  o Grande 163

- Rioni 77  
 Romania 184  
 Rome 182  
 Rondonia 1, 151–173  
 Roosevelt River 154  
 Rossel Island 108  
 Roti 100
- S**  
 Sahara 15  
 Sahul 100–102  
 San Miguel River 162, 163  
 Santa Cruz 151, 154  
 São Miguel River 162  
 Sava River 182  
 Scadar 182  
 Scandinavia 14  
 Sepik River basin 6  
 Seram 100, 104–106  
 Seram-Laut Islands 105  
 Serbia 182, 184  
 Shanghai 227  
 Sicily 61  
 Skopje 199  
 Slovenia 182  
 Solomon Islands 100, 109  
 South Africa 19  
     Cape 16  
 South America 5, 12, 16,  
     18, 62  
 Southern Cone 6  
 Sremska-Mitrovica 182  
 St. Eustatius 18  
 St. Petersburg 190
- Struga 205  
 Sudan  
     Kingdom 15  
     Macro-Sudan 15  
 Sula 100, 105  
 Sulawesi 99, 100, 105  
 Sumba 99, 132  
 Sumbawa 99  
 Sunda 100, 101  
 Surinam 6, 18, 19, 96  
 Surpresa 151  
 Svatenia 30
- T**  
 Taiwan 104  
 Tanimbar 100, 105, 109  
 Tapajós 166  
 Ternate 105, 106  
 Tidore 105, 106, 111, 114,  
     119  
 Timor 97, 99, 100, 102, 104,  
     109–111, 114, 115, 117,  
     119–122, 126–132, 136,  
     137, 139  
     East 107, 109, 110  
 Trebizond, *see* Trabzon  
 Turkey 31, 56, 75, 184  
     Anatolia 52, 75  
     Rize 67, 68  
     Trabzon 67, 68, 77
- U**  
 Una River 182  
 United States  
     New York 19  
     Northeast 6  
     Northern California 6  
     South Coast Range 6  
     Southeast 6  
     Southern California 6  
 Urmia 45
- V**  
 Van 45  
 Vaupés 4, 6, 164  
 Venezuela 6  
 Vietnam 223, 228  
 Virgin Islands 18  
     St. Thomas 18  
 Voden (Edesa) 195, 196  
 Vojvodina 184
- W**  
 Waigeo 100  
 Wallace line 100, 101  
 Wallacea 102  
 Wenshan 226  
 Western Arizona 6  
 Wetar 127  
 Wuming 226
- Y**  
 Yapen 104, 111, 121, 139  
 Yawa 109, 110  
 Yugoslavia 184  
 Yunnan 223, 226



## *Studies in Language Companion Series*

A complete list of titles in this series can be found on the publishers' website, [www.benjamins.com](http://www.benjamins.com)

- 100 AMEKA, Felix K. and Mary Esther KROPP DAKUBU (eds.): Aspect and Modality in Kwa Languages. vii, 344 pp. *Expected March 2008*
- 99 HØEG MÜLLER, Henrik and Alex KLINGE (eds.): Essays on Nominal Determination. From morphology to discourse management. xviii, 366 pp. + index. *Expected April 2008*
- 98 FABRICIUS-HANSEN, Cathrine and Wiebke RAMM (eds.): 'Subordination' versus 'Coordination' in Sentence and Text. A cross-linguistic perspective. vi, 351 pp. + index. *Expected April 2008*
- 97 DOLLINGER, Stefan: New-Dialect Formation in Canada. Evidence from the English modal auxiliaries. 2008. xxii, 355 pp.
- 96 ROMEO, Nicoletta: Aspect in Burmese. Meaning and function. 2008. xv, 289 pp.
- 95 O'CONNOR, Loretta: Motion, Transfer and Transformation. The grammar of change in Lowland Chontal. 2007. xiv, 251 pp.
- 94 MISTAMO, Matti, Kaius SINNEMÄKI and Fred KARLSSON (eds.): Language Complexity. Typology, contact, change. xiv, 356 pp. *Expected February 2008*
- 93 SCHALLEY, Andrea C. and Drew KHELENTZOS (eds.): Mental States. Volume 2: Language and cognitive structure. 2007. x, 362 pp.
- 92 SCHALLEY, Andrea C. and Drew KHELENTZOS (eds.): Mental States. Volume 1: Evolution, function, nature. 2007. xii, 304 pp.
- 91 FILIPOVIĆ, Luna: Talking about Motion. A crosslinguistic investigation of lexicalization patterns. 2007. x, 182 pp.
- 90 MUYSKEN, Pieter (ed.): From Linguistic Areas to Areal Linguistics. 2008. vii, 293 pp.
- 89 STARK, Elisabeth, Elisabeth LEISS and Werner ABRAHAM (eds.): Nominal Determination. Typology, context constraints, and historical emergence. 2007. viii, 370 pp.
- 88 RAMAT, Paolo and Elisa ROMA (eds.): Europe and the Mediterranean as Linguistic Areas. Convergencies from a historical and typological perspective. 2007. xxvi, 364 pp.
- 87 VERHOEVEN, Elisabeth: Experiential Constructions in Yucatec Maya. A typologically based analysis of a functional domain in a Mayan language. 2007. xiv, 380 pp.
- 86 SCHWARZ-FRIESEL, Monika, Manfred CONSTEN and Mareile KNEES (eds.): Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference. 2007. xvi, 282 pp.
- 85 BUTLER, Christopher S., Raquel HIDALGO DOWNING and Julia LAVID (eds.): Functional Perspectives on Grammar and Discourse. In honour of Angela Downing. 2007. xxx, 481 pp.
- 84 WANNER, Leo (ed.): Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In honour of Igor Mel'čuk. 2007. xviii, 380 pp.
- 83 HANNAY, Mike and Gerard J. STEEN (eds.): Structural-Functional Studies in English Grammar. In honour of Lachlan Mackenzie. 2007. vi, 393 pp.
- 82 ZIEGELER, Debra: Interfaces with English Aspect. Diachronic and empirical studies. 2006. xvi, 325 pp.
- 81 PEETERS, Bert (ed.): Semantic Primes and Universal Grammar. Empirical evidence from the Romance languages. 2006. xvi, 374 pp.
- 80 BIRNER, Betty J. and Gregory WARD (eds.): Drawing the Boundaries of Meaning. Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn. 2006. xii, 350 pp.
- 79 LAFFUT, An: Three-Participant Constructions in English. A functional-cognitive approach to caused relations. 2006. ix, 268 pp.
- 78 YAMAMOTO, Mutsumi: Agency and Impersonality. Their Linguistic and Cultural Manifestations. 2006. x, 152 pp.
- 77 KULIKOV, Leonid, Andrej MALCHUKOV and Peter de SWART (eds.): Case, Valency and Transitivity. 2006. xx, 503 pp.
- 76 NEVALAINEN, Terttu, Juhani KLEMOLA and Mikko LAITINEN (eds.): Types of Variation. Diachronic, dialectal and typological interfaces. 2006. viii, 378 pp.
- 75 HOLE, Daniel, André MEINUNGER and Werner ABRAHAM (eds.): Datives and Other Cases. Between argument structure and event structure. 2006. viii, 385 pp.
- 74 PIETRANDREA, Paola: Epistemic Modality. Functional properties and the Italian system. 2005. xii, 232 pp.
- 73 XIAO, Richard and Tony MCENERY: Aspect in Mandarin Chinese. A corpus-based study. 2004. x, 305 pp.



- 72 **FRAJZYNGIER, Zygmunt, Adam HODGES and David S. ROOD (eds.):** *Linguistic Diversity and Language Theories*. 2005. xii, 432 pp.
- 71 **DAHL, Östen:** *The Growth and Maintenance of Linguistic Complexity*. 2004. x, 336 pp.
- 70 **LEFEBVRE, Claire:** *Issues in the Study of Pidgin and Creole Languages*. 2004. xvi, 358 pp.
- 69 **TANAKA, Lidia:** *Gender, Language and Culture. A study of Japanese television interview discourse*. 2004. xvii, 233 pp.
- 68 **MODER, Carol Lynn and Aida MARTINOVIC-ZIC (eds.):** *Discourse Across Languages and Cultures*. 2004. vi, 366 pp.
- 67 **LURAGHI, Silvia:** *On the Meaning of Prepositions and Cases. The expression of semantic roles in Ancient Greek*. 2003. xii, 366 pp.
- 66 **NARIYAMA, Shigeko:** *Ellipsis and Reference Tracking in Japanese*. 2003. xvi, 400 pp.
- 65 **MATSUMOTO, Kazuko:** *Intonation Units in Japanese Conversation. Syntactic, informational and functional structures*. 2003. xviii, 215 pp.
- 64 **BUTLER, Christopher S.:** *Structure and Function – A Guide to Three Major Structural-Functional Theories. Part 2: From clause to discourse and beyond*. 2003. xiv, 579 pp.
- 63 **BUTLER, Christopher S.:** *Structure and Function – A Guide to Three Major Structural-Functional Theories. Part 1: Approaches to the simplex clause*. 2003. xx, 573 pp.
- 62 **FIELD, Fredric:** *Linguistic Borrowing in Bilingual Contexts. With a foreword by Bernard Comrie*. 2002. xviii, 255 pp.
- 61 **GODDARD, Cliff and Anna WIERZBICKA (eds.):** *Meaning and Universal Grammar. Theory and empirical findings. Volume 2*. 2002. xvi, 337 pp.
- 60 **GODDARD, Cliff and Anna WIERZBICKA (eds.):** *Meaning and Universal Grammar. Theory and empirical findings. Volume 1*. 2002. xvi, 337 pp.
- 59 **SHI, Yuzhi:** *The Establishment of Modern Chinese Grammar. The formation of the resultative construction and its effects*. 2002. xiv, 262 pp.
- 58 **MAYLOR, B. Roger:** *Lexical Template Morphology. Change of state and the verbal prefixes in German*. 2002. x, 273 pp.
- 57 **MELČUK, Igor A.:** *Communicative Organization in Natural Language. The semantic-communicative structure of sentences*. 2001. xii, 393 pp.
- 56 **FAARLUND, Jan Terje (ed.):** *Grammatical Relations in Change*. 2001. viii, 326 pp.
- 55 **DAHL, Östen and Maria KOPTJEVSKAJA-TAMM (eds.):** *Circum-Baltic Languages. Volume 2: Grammar and Typology*. 2001. xx, 423 pp.
- 54 **DAHL, Östen and Maria KOPTJEVSKAJA-TAMM (eds.):** *Circum-Baltic Languages. Volume 1: Past and Present*. 2001. xx, 382 pp.
- 53 **FISCHER, Olga, Anette ROSENBACH and Dieter STEIN (eds.):** *Pathways of Change. Grammaticalization in English*. 2000. x, 391 pp.
- 52 **TORRES CACOULOS, Rena:** *Grammaticization, Synchronic Variation, and Language Contact. A study of Spanish progressive -ndo constructions*. 2000. xvi, 255 pp.
- 51 **ZIEGELER, Debra:** *Hypothetical Modality. Grammaticalisation in an L2 dialect*. 2000. xx, 290 pp.
- 50 **ABRAHAM, Werner and Leonid KULIKOV (eds.):** *Tense-Aspect, Transitivity and Causativity. Essays in honour of Vladimir Nedjalkov*. 1999. xxxiv, 359 pp.
- 49 **BHAT, D.N.S.:** *The Prominence of Tense, Aspect and Mood*. 1999. xii, 198 pp.
- 48 **MANNEY, Linda Joyce:** *Middle Voice in Modern Greek. Meaning and function of an inflectional category*. 2000. xiii, 262 pp.
- 47 **BRINTON, Laurel J. and Minoji AKIMOTO (eds.):** *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. 1999. xiv, 283 pp.
- 46 **YAMAMOTO, Mutsumi:** *Animacy and Reference. A cognitive approach to corpus linguistics*. 1999. xviii, 278 pp.
- 45 **COLLINS, Peter C. and David LEE (eds.):** *The Clause in English. In honour of Rodney Huddleston*. 1999. xv, 342 pp.
- 44 **HANNAY, Mike and A. Machtelt BOLKESTEIN (eds.):** *Functional Grammar and Verbal Interaction*. 1998. xii, 304 pp.
- 43 **OLBERTZ, Hella, Kees HENGVELD and Jesús SÁNCHEZ GARCÍA (eds.):** *The Structure of the Lexicon in Functional Grammar*. 1998. xii, 312 pp.