

# Advances in Contact Linguistics

In honour of Pieter Muysken

EDITED BY

Norval Smith  
Tonjes Veenstra  
Enoch O. Aboh

JOHN BENJAMINS PUBLISHING COMPANY

Copyright 2020. John Benjamins Publishing Company. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

# Advances in Contact Linguistics

# Contact Language Library (COLL)

ISSN 2542-7059

Contact Language Library (CoLL), formerly *Creole Language Library*, presents descriptive and theoretical studies on contact languages, such as pidgin and creole languages, mixed languages and other languages which have undergone significant contact-induced change. All CoLL publications are anonymously and internationally refereed.

For an overview of all books published in these series, please see [benjamins.com/catalog/cll](http://benjamins.com/catalog/cll) and [benjamins.com/catalog/coll](http://benjamins.com/catalog/coll)

## Editors

Felicity Meakins  
University of Queensland  
Australia

Viveka Velupillai  
Justus Liebig University Giessen  
Germany

## Editorial Advisory Board

Umberto Ansaldo  
Hong Kong SAR, China

Silvia Kouwenberg  
Kingston, Jamaica

Salikoko S. Mufwene  
Chicago, USA

Peter Bakker  
Aarhus, Denmark

Michelle Li  
Hong Kong SAR, China

Shobha Satyanath  
Delhi, India

Marlyse Baptista  
Ann Arbor, USA

Miriam Meyerhoff  
Wellington, New Zealand

Armin Schwegler  
Irvine, USA

Joseph T. Farquharson  
Kingston, Jamaica

Susanne Maria Michaelis  
Leipzig, Germany

Sarah G. Thomason  
Ann Arbor, USA

Diana Guillemin  
Brisbane, Australia

Pieter Muysken  
Nijmegen, The Netherlands

Tonjes Veenstra  
Berlin, Germany

## Volume 57

Advances in Contact Linguistics. In honour of Pieter Muysken  
Edited by Norval Smith, Tonjes Veenstra and Enoch O. Aboh

# Advances in Contact Linguistics

In honour of Pieter Muysken

*Edited by*

Norval Smith

University of Amsterdam

Tonjes Veenstra

ZAS

Enoch O. Aboh

University of Amsterdam

John Benjamins Publishing Company

Amsterdam / Philadelphia



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

DOI 10.1075/coll.57

**Cataloging-in-Publication Data available from Library of Congress:**  
LCCN 2020023535 (PRINT) / 2020023536 (E-BOOK)

ISBN 978 90 272 0756 2 (HB)

ISBN 978 90 272 6073 4 (E-BOOK)

© 2020 – 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 Company · <https://benjamins.com>

# Table of contents

Foreword	VII
Acknowledgements	IX
INTRODUCTION	
Pieter C. Muysken: A brief biography, a language contact bibliography and a Festschrift summary	1
<i>Tonjes Veenstra, Norval S. H. Smith and Enoch Oladé Aboh</i>	
<b>Part 1. Creole languages and creole studies</b>	
Moving into and out of Sranan: Multiple effects of contact	37
<i>James Essegbey and Adrienne Bruyn</i>	
Sociolinguistic characteristics of the English-lexifier contact languages of West Africa	61
<i>Kofi Yakpo</i>	
The quest for non-European creoles: Is Kukama (Brazil, Peru) a creole language?	85
<i>Peter Bakker</i>	
Are creoles a special type of language? Methodological issues in new approaches to an old question	107
<i>Silvia Kouwenberg and John Victor Singler</i>	
<b>Part 2. Linguistic areas</b>	
Separating layers of information: The anatomy of contact zones	161
<i>Rik van Gijn</i>	
Areal diffusion of applicatives in the Amazon	179
<i>Mily Crevels and Hein van der Voort</i>	
Transfer of Swahili 'until' in contact with East African languages	217
<i>Maarten Mous</i>	

**Part 3. Mixed languages and language mixing**

Turkish-German code-switching patterns revisited: What naturalistic data can(not) tell us <i>Jeanine Treffers-Daller</i>	237
Mixing and semantic transparency in the genesis of Yilan Japanese <i>Luis Miguel Rojas-Berscia</i>	261
Pottefers Cant, Groenstraat Bargoens, and the development of “have” and “be” in the wider context of contact <i>Norval Smith and Frans Hinskens</i>	283

**Part 4. Sociolinguistic aspects of language contact**

Sociolinguistic enregisterment through languagecultural practices <i>Leonie Cornips and Vincent de Rooij</i>	341
Snow on the Danish Antilles? Referee design in Virgin Island Dutch Creole <i>Cefas van Rossem</i>	365
Subject index	389
Author index	391
Language index	397

# Foreword

On the 12th of May 2017, Pieter Muysken held his Afscheidsrede “Terug naar Babel” (Farewell lecture “Back to Babel”),<sup>1</sup> on the subject of why there are so many languages in the world. The distribution of these is very puzzling. In Europe, where man has been present for more than 50,000 years, we have only four language-families. In South America there is documentation for about 490 languages, in 109 families. But, humans only crossed into North America 16,000 years ago. Why this extreme linguistic diversity in South America? The answer would seem to lie in an initial sparse and very widely spread population. “Your nearest neighbour might live 300 km away.” But “Terug naar Pieter” (“Back to Pieter”).

Pieter Muysken was born in South America, in Bolivia, and revisited his “roots” in South America on various occasions, including, of course, during his fieldwork on Quechua in Ecuador in the mid-70s. Recently he has revisited his “even deeper roots” with his work on mining languages. His father was a mining engineer.

Other South American events included a Visiting Associate Professorship at San Marcos University, in Lima, Peru in 1981 and a Visiting Professorship at a Summer School at the Federal University of Rio de Janeiro, Brazil in 1989. He has also supervised many dissertations on South American indigenous languages, of which Simon van de Kerke’s dissertation on Bolivian Quechua was just the first.

But here we are not so much concerned with descriptions of languages, as with the effects of language contact. In the article *Pieter C. Muysken: An attempt at a Language Contact overview and bibliography* we divide *Language Contact in General* into nine rough categories, ordered in terms of the number of entries attributed to each. These are as follows:

- Creole studies (two subsections: *defining* and *detailed*)
- Language contact studies
- Bilingualism/Multilingualism
- Code-switching/Code-mixing
- Mixed language studies
- Areal linguistics

---

1. Retired officially 11th July 2017.



- Immigrants (three subsections: *foreign workers*, *ethnolects*, *heritage languages*)
- Regional studies
- Borrowing

In that article can be found a bibliography of Pieter's work on language contact, listing more than 230 scientific articles, scientific encyclopaedia articles and editorial activities.

As all of us know, Pieter has a talent for organization – organizing people, organizing groups, organizing publications, organizing meetings, and last but not least, organizing finance. He has the ability to weave these talents together and produce (as the economists would say) added value. We hope this is reflected in this volume. The topics included and views expressed cover a large part of the range of Pieter's language contact work.

The contents of the book have turned out to be highly apposite, containing as it does a dozen articles on a representative selection of the topics mentioned above, written by colleagues and students of Pieter's, from the very beginning of his career to the present. We hope he will enjoy the contents of this book in honour of him. We know that he will read every word critically.

The editors

# Acknowledgements

We would like to acknowledge the help of everyone who assisted in anyway in the production of this volume.

First of all, we would like to thank a nameless but very important group of people, our 25 anonymous reviewers of the 12 scientific contributions. Then the 17 authors of these 12 contributions (five joint articles), some of whom have had to be very patient.

Then we would like to thank the editors of the retooled Benjamins *Creole Language Library*, henceforth the *Contact Language Library*, in particular Felicity Meakins of the University of Queensland. Felicity has been a great help and encouragement during the whole long process.

We would also like to thank Kees Vaes of of the John Benjamins Publishing Company for his stalwart help with regard to the production process.

We would finally like to think Frans Hinskens, and the staff of the Meertens Institute, for facilitating the presentation of 10 of the 12 papers, in an afternoon symposium on 6th December 2019.

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416591334 – SFB 1412 for Veenstra's involvement in this project is hereby gratefully acknowledged.



## INTRODUCTION

# Pieter C. Muysken

## A brief biography, a language contact bibliography and a Festschrift summary

Tonjes Veenstra, Norval S. H. Smith and Enoch Oladé Aboh  
Leibniz-Centre General Linguistics (ZAS) / University of Amsterdam

### Introduction

How does one sum up someone's career? Pieter Muysken has shown himself to be someone with inspirational ideas in the fields of language contact research. He also has a talent for developing the media required for the expression of these fields, including organising colleagues around relevant research questions and training new talents. The combination is a rare one, and so is the combination of chapters in this volume which shows the breadth of Pieter's theoretical interests and only some of the people he has interacted with in the field.

### Part 1. Brief biography

#### 1.1 Birth

Pieter Cornelis Muysken was born in 1950 in Oruro, Bolivia, where his father was a mining engineer.

#### 1.2 University education

His university education had three phases. First, he graduated B.A. in 1972 at Yale College (the undergraduate division of Yale University) with a major in Latin American Studies and Spanish. Secondly, he graduated Doctorandus (M.A.) in 1974, at the Institute for General Linguistics of the University of Amsterdam, with a thesis entitled *Some syntactic aspects of creolization*. Thirdly, he graduated Doctor in de Letteren (Ph.D.) in 1977, still at the Institute for General Linguistics of the

University of Amsterdam, with a dissertation entitled *Syntactic developments in the verb phrase of Ecuadorian Quechua*.

A compulsory accompaniment to dissertations in Amsterdam at the time was a number of *Stellingen*, or “position statements.” Pieter became famous for his last *stelling* “Mijn werkplek, dat is mijn hoofd,” ‘My workplace, that’s my head.’ The context of this statement was the proposal to install time-clocks in universities.

### 1.3 Career

Pieter Muysken’s career had six stages. First, in 1973/1974 he was a student assistant at the Institute for General Linguistics, University of Amsterdam. Then, from 1976, he was a lecturer at the same institute. In 1989 he became Professor of General Linguistics, still at the University of Amsterdam, with a specialization in *Sociolinguistics and Creole Studies*.

In 1998 he moved to Leiden University, as Professor of Linguistics, with a specialization in *Ibero-American Linguistics*.

Finally, in 2001, he became Professor of General Linguistics at the Radboud University Nijmegen, until 2017, when he retired. He is now an Emeritus Professor at the Radboud University Nijmegen.

## Part 2. A language contact overview with bibliography

Our title indicates an obvious limitation. We have restricted ourselves to studies involving *language contact* in general. The largest number of bibliography entries concerns *creole languages and creolization processes*. These are also the easiest to categorize. The second most prolific area is *language contact* (in particular). Other obvious categories are *second language learning/acquisition by foreign workers* and *mixed languages*, although both of these have links to other topics – such as *bilingualism* and *code-switching*. *Borrowing* is a more discrete topic.

More difficult to distinguish are *code-switching* and *bilingualism*. You can’t switch codes unless you speak (at least) two languages, or varieties thereof, but bilinguals don’t necessarily code-switch. *Language contact studies*, *areal linguistics* and *regional studies* tend to overlap with each other. We have tried to recognise the main thrust of the various articles, in order to bring out the degree of attention devoted to each of these types of study. Occasionally, references are repeated.

## 2.1 Preoccupations

In terms of language contact studies, including fieldwork, Pieter Muysken's main preoccupations can be divided up by decade, roughly as follows.

- |     |        |   |
|-----|--------|---|
| (1) | 70s    | work on Media Lengua, Papiamentu, cross-creole comparisons  |
|     | 80s    | work on Media Lengua, Surinamese languages (with a focus on Saramaccan)                           |
|     | 90s    | work on contact in the Andes, Negerhollands, Papiamentu, cross-creole comparisons, code-switching |
|     | 00/10s | work on code-switching, contact in South America, cross-creole comparisons                        |

He was an excellent publicist for the study of creole languages, founding *Amsterdam Creole Studies* (1977) and editing *Generative Studies on Creole Languages* (1981). The Amsterdam Creole Workshop (April 1985) on the theme "Substrata versus Universals in Creole Genesis" provided, as Glenn Gilbert writes in his first Editor's Foreword in the *Journal of Pidgin and Creole Languages* (1986), the impetus both for that journal and the associated monograph series, the *Creole Language Library* (also 1986), which kicked off with the proceedings of the above-mentioned workshop as its first publication.

## 2.2 Analysis of publications and editorial work

For each of the main categories we have mentioned above these are the following numbers of references, going from large to small.

(2)	<b>Pidgin &amp; Creole studies</b>	75
	<b>Language contact studies</b>	51
	Bi/Multilingualism	24
	Code-switching	22
	Mixed language studies	18
	Areal linguistics	12
	L2 learning (foreign workers)	10
	Regional studies	8
	Borrowing	8
	Ethnolects	6
	Heritage languages	2
	<b>Total number of entries</b>	<b>236</b>

As can be seen mentions of creole studies and studies of language contact are by far the majority of the total of 236 mentions.<sup>1</sup> Another perspective can be found in the publication dates of the first entries in each of these eleven categories.

(3) Pidgin & Creole studies	1977
Language contact studies	1980
L2 learning (foreign workers)	1980
Mixed language studies	1981
Regional studies	1981
Bi/Multilingualism	1982
Code-switching	1986
Borrowing	1994
Areal linguistics	2000
Ethnolects	2007
Heritage languages	2013

From this it can be seen that Pieter Muysken has always had a broad field of interest. There are however different patterns to be found in the intensity with which he was involved with certain topics. So, if we take *creole/pidgins*, his most prolific language contact field, we find the following, under (4).

(4) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	14	7	8	23	4	5	4	8	(2)	75

Here we observe that while still a significant concern of Pieter Muysken's, "plain" creole articles have taken a reduced role by comparison with more South American interests. This is reflected to some extent by comparison with (5), which deals with (ordinary) *contact linguistics*. The increased number of mentions in (4) to 8 is due largely to the effect of the *Surviving the Middle Passage*-project.

If we compare (4) with (5) the emphasis shifts between the early to mid 90s, where many of the mentions have to do with chapters of the handbook, *Pidgins and Creoles: An introduction*, and the 2000s, and more so the 2010s, where other language contact phenomena take the overhand.

(5) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	2	2	4	1	6	9	3	12	(12)	51

*Bilingualism and Multilingualism*, on the other hand, shows a very stable picture:

(6) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	–	2	1	3	3	3	4	3	(6)	25

1. Including the very few duplications.

*Code-switching* is different again to the extent that it starts later and shows a slightly clearer growth at the beginning, only to become less significant later on:

(7) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	–	1	3	8	4	–	4	2	(–)	22

*Mixed language studies* show a similar picture quantitatively but also a more consistent picture:

(8) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	2	–	1	3	5	1	1	4	(1)	18

The only other categories that have sufficient entries to be of any (vaguely statistical) significance are probably *Areal linguistics* (9) and *Second language learning (foreign workers)* (10) which show a concentration towards the present and the past respectively.

(9) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	–	–	–	–	1	–	1	9	(1)	12

(10) Years	77–81	82–86	87–91	92–96	97–01	02–06	07–11	12–16	(17–20)	
Entries	6	4	–	–	–	–	–	–	(–)	10

However, something requires to be said about two other recent categories which display a clear relationship to the last-mentioned category. The first is the study of *Ethnolects*, which bears a strong historical relationship to the category of *Second language learning (foreign workers)*. The latter was concerned with features of the L1s of groups of foreign workers, which affected their L2 Dutch. In the Netherlands groups of foreign workers were recruited in the 60s and early 70s to work in industry, initially mostly from Spain, Italy and Portugal, and later also from Turkey and Morocco. Category (10) illustrates Pieter Muysken's involvement with this line of research.

The study of *Ethnolects* involves the modified forms of (mostly) standard languages spoken by minority groups in the population. In the Netherlands some of these concern immigrant populations whose nucleus was formed by former *foreign worker* groups. Another, *Surinamese Dutch*, derives from the large-scale immigration in the early 70s from Surinam, where Dutch is the official language.

The mirror-image of the study of *Ethnolects* is the study of *Heritage languages*. These are the (often somewhat reduced) forms of L1s of immigrants. Pieter Muysken is the co-author of a 2019 book on this topic.

### 2.3 Two examples of early seminal articles

Although it would be an impossible task to strive for completeness in covering the content of Pieter Muysken's language contact work, we would nevertheless like to highlight two of his writings from the eighties, which we feel have been important



and seminal for the field of creole studies, and still are so. They have both become in some sense “classics” and standard references for the topics they cover.

The first one we wish to highlight is his 1981 paper *Halfway between Quechua and Spanish: the case for relexification*, in which he firmly puts the process of relexification on the research agenda, and defines it as follows: “the process of vocabulary substitution in which the only information adopted from the target language in the lexical entry is the phonological representation” (Muysken 1981: 61). This idea then was taken up quickly by Claire Lefebvre and collaborators, defining her research approach for decades to come, and eventually relexify the notion as *relabeling*. Even the more recent *Recombination* approach of Enoch Aboh can be regarded as an off-shoot: whereas the original idea was solely the recombination of the phonological representation of the superstrate with the combined set of semantic and syntactic representations from the substrate, in the latter approach all representations (or features) that make up lexical entries can in principle be combined freely. In that sense, Recombination could be seen as Relexification-gone-wild!

A second paper that has been highly influential is his 1988 paper *Are creoles a special type of language?* in the Cambridge Survey of Linguistics. In this paper he observes there is an implicit assumption in the field that creole languages share some properties that call for an explanatory theory. The three properties he identifies are *aliveness*, *simplicity* and *being mixed*, i.e. creoles are assumed to be more alike/simple/mixed than other languages. These three properties have dominated research on creole languages until this very day (and, even this very book!) The similarity debate started in 1998 with John McWhorter’s claim that creoles should be recognized as a separate typological class and got fired up again in 2011 by Peter Bakker and his collaborators. In between, the simplicity debate really got off the ground in 2001 with the heated discussion on whether creoles have the world’s most simple grammars in *Linguistic Typology*, and the mixed character of creole grammar has only recently begun to be seriously scrutinized within the *Hybridity of Grammar* approach of Enoch Aboh. While some of these debates will probably go on into the future, creolists should not forget the sobering conclusion of this influential paper that “[t]he very notion of a ‘creole’ language from the linguistic point of view tends to disappear if one looks closely; what we have is just a language” (Muysken 1988: 300). Support for this conclusion comes from the important observation that an intricate (interpretive) relation exists between adverbs and TMA elements. This observation has been further systematically analysed in Cinque’s (1999) seminal work on adverb sequencings and their relation to TMA ordering and clause structure which demonstrates that the assumed likeness of creoles actually reflects the likeness of human languages in general.

## 2.4 A provisional list of articles, authored and edited books involving Pieter C. Muysken's *language contact* interests<sup>2</sup>

In this bibliography we do not distinguish references with identical dates of publication in terms of *a, b, c*, etc. We thought this would be pointless and confusing, as we have divided the bibliography into twelve separate sections, which we repeat here for convenience. The first two sections were combined in our preceding analysis, but we have distinguished them in the bibliography.

1. Pidgin & Creole studies (with two subsections: *defining* and *detailed*)
2. Language Contact studies (in the narrower sense of the term)
3. Bilingualism & Multilingualism
4. Code-mixing/Code-switching
5. Mixed Language studies
6. Areal Linguistics
7. The languages of immigrants (three subsections: *L2 learning (foreign workers)*, *ethnolects* and *heritage languages*)
8. Regional studies
9. Borrowing

Of interest too is the number of shared references in this bibliography:

2.4.1.1	4	(10)	Pidgin & Creole studies – defining	40.0%
2.4.1.2	34	(65)	Pidgin & Creole studies – detailed	52.3%
2.4.2	13	(51)	Language Contact studies in the narrow sense	24.5%
2.4.3	12	(24)	Bilingualism & Multilingualism	50.0%
2.4.4	9	(22)	Code-mixing/Code-switching	40.9%
2.4.5	2	(18)	Mixed Language studies	11.1%
2.4.6	8	(12)	Areal Linguistics	66.7%
2.4.7.1	7	(10)	Immigrants – Foreign workers	70.0%
2.4.7.2	4	(6)	Immigrants – Ethnolects	66.7%
2.4.7.3	2	(2)	Immigrants – Heritage languages	100.0%
2.4.8	6	(8)	Regional studies	75.0%
2.4.9	2	(8)	Borrowing	25.0%

This demonstrates the significant role that cooperation played in Pieter Muysken's work.

---

2. We have omitted short reviews, unless they seemed important, have not included articles about L2 teaching, nor about dialectal variation (as this was concerned with intra-language matters), and have ignored any mention of other places of publication than that first-mentioned. For the reader's information, we have translated Dutch titles into English.

### 2.4.1 *Pidgin & Creole studies*

#### 2.4.1.1 *Creole – Defining*

- Muysken, P. C. 1979. Wat is kreolistiek? (What is creolistics?). In *Wetenschap & taal: Een nieuwe reeks benaderingen van het verschijnsel taal*, B. T. Tervoort (ed.), 226–242. Muiderberg: Coutinho.
- Muysken, P. C. & Meijer, G. 1979. Introduction. In *On the Origin and Formation of Creoles: A Miscellany of Articles*, D. C. Hesseling, T. L. Markey & P. T. Roberge (eds), vii–xxi. Ann Arbor MI: Karoma. (ISBN: 0897200055).
- Muysken, P. C. 1981. Introduction. In *Generative Studies on Creole Languages*, P. C. Muysken (ed.), 1–4. Dordrecht: Foris.
- Muysken, P. C. 1984. Do creoles give insight into the human language faculty? Commentary on D. Bickerton, The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7: 203–204. (ISSN: 0140-525X).
- Muysken, P. C. 1988. Are creoles a special type of language? In *Linguistics: The Cambridge Survey*, Vol. 2, F. J. Newmeyer (ed.), 285–307. Cambridge: CUP.
- Muysken, P. C. 1995. Creole linguistics. In *Handbook of Pragmatics: Manual*, J. Verschueren, J.-O. Östman & J. Blommaert (eds), 198–204. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/hop.m.cre1>
- Muysken, P. C. & Law, P. 2002. Creole studies. A theoretical linguist's field guide. *Glott International* 5: 47–57. (ISSN: 1381-3439).
- Muysken, P. C. & Escure, G. 2006. Creole linguistics. (revised and expanded version). In *Handbook of Pragmatics*, J. Verschueren, J.-O. Östman & E. Versluys (eds), 1–24. Amsterdam: John Benjamins. <https://doi.org/10.1075/hop.10.cre1>
- Muysken, P. C. & van den Berg, M. C. 2011. Creole. *Oxford Bibliography Online*. Oxford: OUP.  
<https://doi.org/10.1093/OBO/9780199772810-0011>
- Muysken, P. C. 2016. Creole languages. In *Oxford Research Encyclopedias Online: Linguistics*. Oxford: OUP. <https://doi.org/10.1093/acrefore/9780199384655.013.68>

#### 2.4.1.2 *Pidgin & Creole – Detailed*

- Muysken, P. C. (ed.). 1977. *Amsterdam Creole Studies* 1. (ISSN: 1570-2901).
- Meijer, G. & Muysken, P. C. 1977. On the beginnings of Pidgin and Creole studies: Schuchardt and Hesseling. In *Pidgin and Creole Linguistics*, A. Valdman (ed.), 21–48. Bloomington IN: Indiana University Press.
- Muysken, P. C. 1977. Movement rules in Papiamentu. In *Amsterdam Creole Studies* 1, P. C. Muysken (ed.), 80–102. (ISSN: 1570-2901).
- Muysken, P. C. (ed.). 1978. *Amsterdam Creole Studies* 2. (ISSN: 1570-2901).
- Jansen, B., Koopman, H. & Muysken, P. C. 1978. Serial verbs in the Creole languages. In *Amsterdam Creole Studies* 2, P. C. Muysken (ed.), 125–159. (ISSN: 1570-2901).
- Muysken, P. C. 1978. Three types of fronting constructions in Papiamentu. In *Studies on Fronting: Papers of a Symposium, Held in Leiden, 1977*, 65–79. Lisse: Peter de Ridder Press.
- Muysken, P. C. 1980. Book reviews. A. Bollée, *Le Créole français des Seychelles: Esquisse d'une grammaire – textes – vocabulaire*. (1977); C. Corne, *Elements for Indian Ocean Proto-Creole reconstruction*. (1977); A. Valdman, *Le Créole: Structure, statut et origine*. (1978); I. Vintilă-Rădulescu, *Le Créole français*. (1976). *Lingua* 51: 94–100.
- Muysken, P. C. & Smith, N. S. H. (eds). 1980. *Amsterdam Creole Studies* 3. (ISSN: 1570-2901).

- Muysken, P. C. (ed.). 1981. *Generative Studies on Creole Languages* [Studies in Generative Grammar 8]. Dordrecht: Foris.
- Muysken, P. C. 1981. Introduction. In *Generative Studies on Creole Languages*, P. C. Muysken (ed.), 1–4. Dordrecht: Foris.
- Muysken, P. C. 1981. Creole tense/mood/aspect systems: The unmarked case? In *Generative Studies on Creole Languages*, P. C. Muysken (ed.), 181–200. Dordrecht: Foris.
- Smith, N. S. H. & Muysken, P. C. (eds). 1983. *Amsterdam Creole Studies* 5. (ISSN: 1570-2901).
- Muysken, P. C. 1983. Review article. D. Bickerton, Roots of Language. *Language* 59: 884–891. (ISSN: 0097-8507).
- Muysken, P. C. & Smith, N. S. H. (eds). 1984. *Amsterdam Creole Studies* 7. (ISSN: 1570-2901).
- Muysken, P. C. & Smith, N. S. H. (eds). 1985. *Amsterdam Creole Studies* 8. (ISSN: 1570-2901).
- Muysken, P. C. & Smith, N. S. H. (eds). 1986. *Substrata Versus Universals in Creole Genesis: Papers from the Amsterdam Creole Workshop, April 1985* [Creole Language Library 1]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.1>
- Muysken, P. C. & Smith, N. S. H. 1986. Introduction: Problems in the identification of substratum features in the creole languages. In *Substrata Versus Universals in Creole Genesis: Papers from the Amsterdam Creole Workshop, April 1985* [Creole Language Library 1], P. C. Muysken & N. S. H. Smith (eds), 1–13. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.1.01muy>
- Muysken, P. C. & Smith, N. S. H. 1987. Question words in creole languages. In *You Me Linguistics Fun!: Nieuwe inzichten in de toegepaste taalwetenschap* (New insights in applied linguistics), R. Appel (ed.), 89–101. Amsterdam: VU Uitgeverij.
- Muysken, P. C. 1987. Prepositions and postpositions in Saramaccan. In *Studies in Saramaccan Language Structure*, M. C. Alleyne (ed.), 89–101. Amsterdam: Universiteit van Amsterdam.
- Muysken, P. C. 1988. Lexical restructuring and creole genesis. In *Beiträge zum 4. Essener Kolloquium über Sprachkontakt, Sprachwandel, Sprachwechsel, Sprachtod<sup>3</sup> vom 9.10.-10.10.1987 an der Universität Essen*, N. Boretzky, H. den Besten & T. Stolz (eds), 193–210. Bochum: Brockmeyer.
- Muysken, P. C. 1988. Parameters for serial verbs. In *Niger-Congo Syntax and Semantics* 1, V. Manfredi (ed.), 65–75. Boston MA: Boston University African Studies Center.
- Muysken, P. C. 1990. Review article. Onderzoek op het gebied van de Surinaamse creolentaalen: Recente resultaten en toekomstperspectief (Research on the Surinam creole languages: Recent results and prospects for the future). *Oso: Tijdschrift voor Surinaamse Taalkunde, Letterkunde en Geschiedenis* 9: 79–94. (ISSN: 0167-4099).
- Muysken, P. C. & Smith, N. S. H. 1990. Question words in pidgin and creole languages. *Linguistics* 28: 883–903. (ISSN 0024-3949).
- Muysken, P. C. & van der Voort, H. 1991. The binding theory and creolization. In *Development and Structures of Creole Languages: Essays in Honor of Derek Bickerton* [Creole Language Library 9], F. Byrne & T. Huebner (eds), 145–158. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.9.17muy>
- den Besten, H. & Muysken, P. C. 1992. De sociolinguïstische geschiedenis van het Negerhollands: Een eerste overzicht (The sociolinguistic history of Negerhollands: An initial overview). In *Thema's en trends in sociolinguïstiek. Toegepaste Taalwetenschap in Artikelen* 42: 67–80.
- Arends, J. T. G. & Muysken, P. C. 1992. Demografische modellen in de creolistiek (Demographic models in creole studies). *Gramma/TTT Tijdschrift voor Taalwetenschap* 1: 41–55. (ISSN: 0927-3255).

- Muysken, P. C. 1993. Reflexes of Ibero-Romance reflexive clitic+verb combinations in Papiamentu. In *Focus and Grammatical Relations in Creole Languages. Papers from the University of Chicago Conference on Focus and Grammatical Relations* [Creole Language Library 12], F. Byrne & D. Winford (eds), 285–302. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/cll.12.21muy>
- Arends, J. T. G., Muysken, P. C. & Smith, N. S. H. (eds). 1994. *Pidgins and Creoles. An Introduction* [Creole Language Library 15]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15>
- Muysken, P. C. & Smith, N. S. H. 1994. The study of pidgin and creole languages. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 3–14. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.05muy>
- den Besten, H., Muysken, P. C. & Smith, N. S. H. 1994. Theories focussing on the European input. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 87–98. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/cll.15.13bes>
- Muysken, P. C. & Veenstra, T. 1994. Universalist approaches. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 121–136. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.16muy>
- Muysken, P. C. & Veenstra, T. 1994. Haitian. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds.), 153–164. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.19muy>
- Kouwenberg, S. & Muysken, P. C. 1994. Papiamentu. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 205–218. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.23kou>
- Muysken, P. C. & Smith, N. S. H. 1994. Reflexives. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 271–288. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.29muy>
- Muysken, P. C. & Veenstra, T. 1994. Serial verbs. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 289–302. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.30muy>
- Arends, J. T. G., Muysken, P. C. & Smith, N. S. H. 1994. Conclusions. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 319–330. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.33are>
- Muysken, P. C. 1994. The search for universals in creole genesis. *Amsterdam Creole Studies* 11: 16–28. (ISSN: 1570-2901).
- Muysken, P. C. & Smith, N. S. H. 1994. Reflexives in the Creole languages: An interim report. In *Creolization and Language Change*, D. Adone & I. Plag (eds), 45–64. Tübingen: Max Niemeyer.
- Muysken, P. C. 1994. Saramaccan and Haitian: A comparison. *Journal of Pidgin and Creole Languages* 9: 305–314. (ISSN: 0920-9034).
- Muysken, P. C. 1994. Column: The first seven years, fat or lean? *Journal of Pidgin and Creole Languages* 9: 103–107. (ISSN: 0920-9034).
- Singh, R. & Muysken, P. C. 1995. Column: Wanted: A debate in pidgin/creole phonology. *Journal of Pidgin and Creole Languages* 10: 157–169. (ISSN: 0920-9034).
- van der Voort, H. & Muysken, P. C. 1995. Eighteenth-Century Negerhollands reflexives revisited. In *The Early Stages of Creolization* [Creole Language Library 13], J. Arends (ed.), 25–42. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.13.04voo>

- Muysken, P. C. 1995. Studying variation in older texts: Negerhollands. *Journal of Pidgin and Creole Languages* 10: 335–347. (ISSN: 0920-9034).
- Muysken, P. C. 1995. Recente publicaties over het Papiamentu (Recent publications on the topic of Papiamentu). *Kruispunt* 161: 111–116. (ISSN: 0166-6355).
- den Besten, H., Muysken, P. C., van Rossem, C., Stein, P. & van der Voort, P. 1996. Introduction. In *Die Creol Tael: 250 Years of Negerhollands Texts*, C. van Rossem & H. van der Voort (eds), 1–46. Amsterdam: Amsterdam University Press.
- Muysken, P. C. 1996. Book review. Africanisms in Afro-American language varieties, S. S. Mufwene (ed.). *Journal of Pidgin and Creole Languages* 11: 167–176.  
<https://doi.org/10.1075/jpcl.11.1.20muy>
- Muysken, P. C., Bruyn, A. & Verrips, M. 1999. Double object constructions in the Creole languages: Development and acquisition. In *Language Creation and Language Change: Creolization, Diachrony, and Development*, M. DeGraff (ed.), 329–373. Cambridge MA: The MIT Press.
- Muysken, P. C. 2001. The creole languages of the Caribbean. In *A History of Literature in the Caribbean, Vol. 2: English- and Dutch-speaking Regions* [A Comparative History of Literatures in European Languages 15], A. J. Arnold (ed.), 399–414. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/chlel.xv.33muy>
- Muysken, P. C. 2001. Creolization. In *Language Typology and Language Universals. An International Handbook 2*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1656–1667. Berlin: Walter de Gruyter.
- Muysken, P. C. 2001. The origin of creole languages: The perspective of second language learning. In *Creolization and Contact* [Creole Language Library 23], N. S. H. Smith & T. Veenstra (eds), 157–173. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.23.07muy>
- Muysken, P. C. 2003. The grammatical elements in Negerhollands: Loss, retention, reconstitution. In *Germania et Alia: A Linguistic Webschrift for Hans den Besten*, J. Koster & H. van Riemsdijk (eds). Groningen: University of Groningen.
- Muysken, P. C. 2004. Pidginization, creolization, and language death. In *Morphology: An International Handbook on Inflection and Word-formation*, G. Booij, C. Lehmann & J. Mugdan (eds), 1653–1661. Berlin: Walter de Gruyter.
- Muysken, P. C. & Veenstra, T. 2006. Serial verbs. In *The Blackwell Companion to Syntax*, Vol 4, M. Everaert & H. van Riemsdijk (eds), 234–270. Oxford: Blackwell.
- Muysken, P. C. 2008. Creole studies and multilingualism research. In *The Handbook of Pidgin and Creole Studies* [Blackwell Handbooks in Linguistics 9], S. Kouwenberg & J. V. Singler (eds), 287–308. Oxford: Wiley-Blackwell.
- Muysken, P. C. 2009. Gradual restructuring in Ecuadorian Quechua. In *Gradual Creolization: Studies Celebrating Jacques Arends* [Creole Language Library, 34] R. Selbach, H. Cardoso & M van den Berg (eds), 77–100. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/cll.34.09muy>
- Kouwenberg, S. & Muysken, P. C. 2011. On the typology of clauses in Papiamentu. In *New Perspectives on the Languages, Literatures and Cultures of the Caribbean*, N. Faraclas, R. Severing & C. Weijer (eds), 13–23. Willemstad, Curaçao: Fundashon pa Planifikashon di Idioma.
- Muysken, P. C. & Smith, N. S. H., with Borges, R. D. (eds). 2015. *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*. Berlin: De Gruyter Mouton.
- van den Berg, M., Muysken, P. C. & Smith, N. S. H. 2015. Introduction: Creole studies and contact linguistics. In *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*, P. C. Muysken & N. S. H. Smith, with R. D. Borges (eds), 1–14. Berlin: De Gruyter Mouton.

- Muysken, P. C. 2015. Relexification, and other language contact scenarios for explaining substrate effects. In *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*, P. C. Muysken & N. S. H. Smith, with R. D. Borges (eds), 107–133. Berlin: De Gruyter Mouton.
- Muysken, P. C. 2015. Conclusion: feature distribution in the West Africa-Surinam Trans-Atlantic Sprachbund. In *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*, P. C. Muysken & N. S. H. Smith, with R. D. Borges (eds), 393–408. Berlin: De Gruyter Mouton.
- Muysken, P. C. 2015. Book review. B. Jacobs. Origins of a Creole: The History of Papiamentu and its African Ties. *New West Indian Guide* 89: 219–221.  
<https://doi.org/10.1163/22134360-08901048>
- Muysken, P. C. 2015. Book review, S. Sessarego. The Afro-Bolivian determiner phrase. Columbus: The Ohio State University Press. *Cincinnati Romance Review* 39: 322–323.
- van Sluijs, R., van den Berg, M. & Muysken, P. C. 2016. Exploring genealogical blends. The Surinamese Creole Cluster and the Virgin Islands Dutch Creole Cluster. *Lingua* 178: 84–103.  
<https://doi.org/10.1016/j.lingua.2015.12.004>
- Muysken, P. C., van Rossem, C. & van Sluijs, R. 2017. Variation in preposition use in the Virgin Islands Dutch Creole (VIDC) Cluster. In *CLARIN in the Low Countries*, J. Odiijk & A. van Hessen (eds), 133–149. London: Ubiquity Press. <https://doi.org/10.5334/bbi.12>
- Veenstra, T. & Muysken, P. C. 2017. Serial verb constructions. In *The Wiley Blackwell Companion to Syntax*, 2nd edn, M. Everaert & H. van Riemsdijk (eds). Oxford: Wiley-Blackwell.  
<https://doi.org/10.1002/9781118358733.wbsyncom075>

#### 2.4.2 Language contact studies

- Muysken, P. C. 1980. Sources for the study of Amerindian contact vernaculars in Ecuador. In *Amsterdam Creole Studies* 3, P. C. Muysken & N. S. H. Smith (eds), 66–82. (ISSN: 1570-2901).
- Muysken, P. C. 1981. Spaans en Quechua in Ecuador (Spanish and Quechua in Ecuador). *TTT: Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 2: 124–138.
- Muysken, P. C. 1983. A sociolinguistic perspective on second language acquisition. In *Amsterdam Creole Studies* 5, N. S. H. Smith & P. C. Muysken (eds), 37–45. (ISSN: 1570-2901).
- Muysken, P. C. 1985. Linguistic dimensions of language contact: The state of the art in interlinguistics. *Revue Québécoise de Linguistique* 14: 49–76. (ISSN: 0710-0167).
- Muysken, P. C. 1987. Taalcontact en grammaticale coherentie: Spaans en Quechua in de ‘wayno’ (Language contact and grammatical coherence). In *Etnische minderheden: Taalverwerving, taalonderwijs, taalbeleid*, G. Extra, R. van Hout & T. Vallen (eds), 1–14. Dordrecht: Foris.
- Muysken, P. C. 1987. Contactos entre castellano y quechua en el Ecuador. In *Antropología del Ecuador: Memorias del Primer Simposio Europeo sobre Antropología del Ecuador*, S. Moreno Yanez (ed.), 377–451. Quito, Ecuador: Ediciones Abya-Yala.
- Muysken, P. C. 1989. A unified theory of local coherence in language contact. In *Urban Language Conflict*, P. H. Nelde (ed.), 177–198. Bonn: Dümmler.
- Muysken, P. C. 1991. Language contact and grammatical coherence: Spanish and Quechua in the wayno of Southern Peru. In *Papers for the Workshop on Constraints, Conditions and Models (ESF2 Network on Code-switching and Language Contact)*, 159–188. Strasbourg: European Science Foundation.
- Muysken, P. C. 1996. Syntax. In *Kontaktlinguistik: Ein internationales Handbuch zeitgenössischer Forschung*, H. Goebel, P. H. Nelde, Z. Starý & W. Wölck (eds), 117–124. Berlin: Walter de Gruyter.

- Muysken, P. C. & Flores Farfán, J. A. 1997. Lenguas en contacto en Iberoamérica: México y los Andes centrales. In *Contactos y transferencias lingüísticas en Hispanoamérica*, 29–70. Buenos Aires: Universidad de Buenos Aires, Facultad de Filosofía y Letras.
- Muysken, P. C. (ed.). 1999. *Sociolingüística: Lenguas en contacto* [Fóro Hispánico 13]. Amsterdam: Rodopi.
- Muysken, P. C. 1999. Contacto lingüístico y coherencia gramatical: Castellano y quechua en los waynos del Perú. In *Sociolingüística: Lenguas en contacto*, P. C. Muysken (ed.), 87–100. Amsterdam: Rodopi.
- Muysken, P. C. 1999. Lexical contact phenomena: Schuchardt and beyond. In *Language Encounters across Time and Space: Studies in Language Contact*, B. Brendemoen, E. Lanza & E. Ryen (eds), 29–36. Oslo: Novus.
- Muysken, P. C. 2000. The genesis of Lowland Ecuadorian Quechua and semantic transparency. In *Creoles, Pidgins, and Sundry Languages. In honour of Pieter Seuren*, J. Arends (ed.). Special issue of *Linguistics* 38: 973–988.
- Muysken, P. C. 2000. Drawn into the Aymara mold? Notes on Uru grammar. In *Essays on Indigenous Languages of Lowland South America: Contributions to the 49th International Congress of Americanists in Quito 1997*, H. van der Voort & S. van de Kerke (eds.), 111–128. Leiden: University of Leiden, Research School of Asian, African, and Amerindian Studies (CNWS).
- Muysken, P. C. 2002. Computation and storage in language contact. In *Studies in Theoretical Psycholinguistics* 30, S. Nootboom, F. Weerman & F. Wijnen (eds), 157–179. Dordrecht: Kluwer.
- Muysken, P. C. 2003. Two linguistic systems in contact: Grammar, phonology, and lexicon. In *The Handbook of Bilingualism and Multilingualism*, W. C. Ritchie & T. K. Bhatia (eds), 193–215. Malden MA: Blackwell. (ISBN: 0631227342).
- Muysken, P. C. 2005. Modeling language contact (otra vez) bilingual optimisation strategies. In *Atti del 4o congresso di studi dell'Associazione Italiana di Linguistica Applicata*, G. Banti, A. Marra & E. Vineis (eds), 11–36. Perugia: Guerra Edizioni.
- Muysken, P. C. 2005. Possessive constructions in language contact settings. In *Language Contact, Hybrids and New Varieties: Emergent Possessive Constructions* [Monash University Linguistics Papers 1–2], S. Musgrave (ed.), 11–20. Melbourne: Monash University.
- Muysken, P. C. 2005. Categories in the syntax and the lexicon: Evidence from language contact research. In *Handbook of Categorization in the Cognitive Sciences*, H. Cohen & C. Lefebvre (eds), 46–71. Amsterdam: Elsevier Science.
- Muysken, P. C. 2005. Psycholinguistic modelling in language contact: Dynamics of language contact. English and immigrant languages (2003) Michael Clyne Cambridge: Cambridge University Press, pp. xv+282 Contact linguistics. Bilingual encounters and grammatical outcomes (2002) Carol Myers-Scotton Oxford: Oxford University Press, pp. xiv+342. *The International Journal of Bilingualism* 9: 511–518. <https://doi.org/10.1177/13670069050090030201>
- Crevels, M. & Muysken, P. C. 2005. La influencia léxica de las lenguas amerindias en el español del Oriente Boliviano y Peruano. In *Sprache in Iberoamerika. Festschrift für Wolf Dietrich zum 65. Geburtstag*, V. Noll & H. Symeonidis (eds), 179–210. Hamburg: Buske.
- Muysken, P. C. 2006. Een model voor taalcontact. *Toegepaste Taalwetenschap in Artikelen* 76: 13–20. <https://doi.org/10.1075/ttwia.76.03muy>
- Muysken, P. C. 2010. Scenarios for language contact. In *Handbook of Language Contact*, R. Hickey (ed.), 265–281. Oxford: Blackwell.



- Muysken, P. C. 2011. Change, contact, and ethnogenesis in Northern Quechua: Structural phylogenetic approaches to clause embedding predicates. In *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archeology, Linguistics, and Ethnohistory*, A. Hornborg & J. D. Hill (eds), 237–258. Boulder CO: University of Colorado Press.
- Muysken, P. C. 2011. Root/affix asymmetries in contact and transfer: Case studies from the Andes. *International Journal of Bilingualism* 16: 22–36. <https://doi.org/10.1177/1367006911403211>
- Muysken, P. C. 2012. Another icon of language contact shattered: Commentary on Poplack, Zentz and Dion. *Bilingualism. Language and Cognition* 15: 237–240. <https://doi.org/10.1017/S1366728911000277>
- Muysken, P. C. 2012. Spanish affixes in the Quechua languages: A multidimensional perspective. *Lingua* 122: 481–493. <https://doi.org/10.1016/j.lingua.2011.10.003>
- Muysken, P. C. 2012. Contacts between indigenous languages in South America. In *The Indigenous Languages of South America*, L. Campbell & V. Grondona (eds), 235–258. Berlin: Walter de Gruyter.
- Muysken, P. C. 2012. Two linguistic systems in contact: Grammar, phonology, and lexicon. In *The Handbook of Bilingualism and Multilingualism*, 2nd edn, T. K. Bhatia & W. C. Ritchie (eds), 193–215. Hoboken NJ: Wiley-Blackwell. [print version 2013] <https://doi.org/10.1002/9781118332382.ch8>
- Muysken, P. C. 2013. A sparkling apéritif or old wine in new bags? *Bilingualism. Language and Cognition* 16: 748–750. <https://doi.org/10.1017/S1366728913000217>
- O'Connor, L. M. & Muysken, P. C. (eds). 2014. *The Native Languages of South America: Origins, Development, Typology*. Cambridge: CUP.
- Muysken, P. C. & O'Connor, L. M. 2014. Introduction: South American indigenous languages; genealogy, typology, contacts. In *The Native Languages of South America: Origins, Development, Typology*, P. C. Muysken & L. M. O'Connor (eds), 1–28. Cambridge: CUP.
- Yakpo, K. & Muysken, P. C. 2014. Language change in a multiple contact setting: The case of Sarnami (Suriname). In *Pidgins and Creoles beyond Africa-Europe Encounters* [Creole Language Library 47], I. Buchstaller, A. Holmberg & M. Almoaily (eds), 101–140. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.47.06yak>
- Muysken, P. C. 2014. Categorical mismatches in the syntax and the lexicon: Evidence from language contact research. In *Grammatical Theory and Bilingual Codeswitching*, J. McSwann (ed.), 119–133. Cambridge MA: The MIT Press.
- Muysken, P. C. 2016. Language contact. Trojan horse or new potential for cross-fertilization? *Linguistic Typology* 20: 537–545. <https://doi.org/10.1515/lingty-2016-0025>
- Muysken, P. C. 2016. Fine tuning cross-linguistic interaction. The nuts and bolts. In *Trends in Applied Linguistics* 17, M. Reif & J. A. Robinson (eds), 207–213. Berlin: De Gruyter Mouton.
- Muysken, P. C. & Muntendam, A. G. 2016. Interfacing interfaces. Quechua and Spanish in the Andes. In *Manual of Grammatical Interfaces in Romance*, S. Fischer & C. Gabriel (eds), 607–634. Berlin: De Gruyter.
- Muysken, P. C. 2017. Using scenarios in language contact studies. Linguistic borrowing into Coptic. In *Lingua Aegyptia Studia Monographica* 17, E. Grossman & P. Dils (eds), 3–16. Hamburg: Widmaier.
- Muysken, P. C. 2017. Linguistic categories in language contact: Modularity and diversity. In *Handbook of Categorization in Cognitive Science*, 2nd edn, H. Cohen & C. Lefebvre (eds), 511–548. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-08-101107-2.00021-X>
- Yakpo, K. & Muysken, P. C. (eds). 2017. *Boundaries and Bridges: Language Contact in Multilingual Ecologies*. Berlin: De Gruyter Mouton.

- Yakpo, K. & Muysken, P. C. 2017. Language contact and change in the multilingual ecologies of the Guianas. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, K. Yakpo & P. C. Muysken (eds), 3–20. Berlin: De Gruyter Mouton.
- Meira, S. & Muysken, P. C. 2017. Cariban in contact: New perspectives on Trio-Ndyuka pidgin. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, K. Yakpo & P. C. Muysken (eds), 197–228. Berlin: De Gruyter Mouton.
- Muysken, P. C. 2017. The transformation of a colonial language: Surinamese Dutch. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, K. Yakpo & P. C. Muysken (eds), 283–308. Berlin: De Gruyter Mouton.
- Borges, R., Muysken, P. C., Villerius, S. & Yakpo, K. 2017. The tense-mood-aspect systems of the languages of Suriname. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, K. Yakpo & P. C. Muysken (eds), 311–362. Berlin: De Gruyter Mouton.
- Yakpo, K. & Muysken, P. C. 2017. Multilingual ecologies in the Guianas: Overview, typology, prospects. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, K. Yakpo & P. C. Muysken (eds), 383–404. Berlin: De Gruyter Mouton.
- Muysken, P. C. 2018. Grammatical archaeology. History and contact in the indigenous languages of South America. In *New Perspectives on the Peopling of the Americas*, K. Harvati, G. Jäger & H. Reyes-Centeno (eds), 61–78. Tübingen: Kerns Verlag. (ISBN: 9783935751285).
- Muysken, P. C. 2018. Three challenges in language contact research. In *Proceedings of The dynamics of language. Plenary and Focus Lectures from the 20th International Congress of Linguists*, R. Mesthrie & D. Bradley (eds), 153–160. Cape Town: University of Cape Town Press. (ISBN: 9781775822363).
- van Sluijs, R., Muysken, P. C. & Los, B. L. J. 2018. Verb particle combinations and word order change in Dutch-lexifier creole languages. In *Word Order Change in Acquisition and Language Contact. Essays in honour of Ans van Kemenade* [Linguistik Aktuell/Linguistics Today 243], B. L. J. Los & P. de Haan (eds), 265–290. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/la.243.12slu>
- Kootstra, G. & Muysken, P. C. 2019. Structural priming, levels of awareness, and agency in contact-induced language change. *Languages* 4(3): 65. <https://doi.org/10.3390/languages4030065>

### 2.4.3 Bilingualism/multilingualism

- Muysken, P. C. 1982. A sociolinguistic theory of second language acquisition. In *Handelingen van het zeven en dertigste Nederlands filologen congres*, R. Stuip & W. Zwanenburg (eds), 119–126. Amsterdam: Vereniging het Nederlandsche Filologen-congres.
- Muysken, P. C. 1984. The Spanish that Quechua speakers learn: L2 learning as norm-governed behaviour. In *Second Languages: A Cross-linguistic Perspective*, R. W. Andersen (ed.), 101–124. Rowley MA: Newbury House.
- Appel, R. & Muysken, P. C. 1987. *Language Contact and Bilingualism*. London: Edward Arnold.
- Muysken, P. C. 1995. Book review. A. M. B. de Groot & C. Barry (eds), *The Multilingual Community: Bilingualism* (European Journal of Cognitive Psychology 4(4)), (1992). *Language in Society* 24: 583–586. <https://doi.org/10.1017/S0047404500019084>
- Muysken, P. C., Vedder, P. & Kook, H. 1996. Language choice and functional differentiation of languages in bilingual parent-child reading. *Applied Psycholinguistics* 17: 461–484. (ISSN 0142-7164).
- Muysken, P. C., Kook, H. & Vedder, P. 1996. Papiamento/Dutch code-switching in bilingual parent-child reading. *Applied Psycholinguistics* 17: 485–505. (ISSN: 0142-7164).

- Muysken, P. C. 1997. Older studies on migrant languages in the United States: A model for European research? In *Festschrift 29 jaar Onderzoekscentrum voor Meertaligheid*, A. Melis & P. Weber (eds), 238–247. Brussel: Vrije Universiteit Brussel.
- Muysken, P. C. 1998. We are all native speakers, but of which language? In *The Native Speaker: Multilingual Perspectives*, R. Singh (ed.), 193–204. New Delhi: Sage.
- Muysken, P. C. 1999. Cómo ser bilingüe en el año 2000? In *Educación bilingüe: Actes del V y VI Alcuentreu “Llingua Minoritaria y Educación”*, 67–71. Uviéu: Academia de la Llingua Asturiana.
- Bennis, H. J., Extra, G., Muysken, P. C. & Nortier, J. W. R. 2002. *Een buurt in beweging: Talen en culturen in het Utrechtse Lombok en Transvaal* (A neighbourhood on the move: Languages and cultures in Lombok and Transvaal districts in Utrecht). Amsterdam: Aksant.
- Olbertz, H. & Muysken, P. C. (eds). 2005. *Encuentros y conflictos: Bilingüismo y contacto de lenguas en el mundo andino* [Lengua y Sociedad en el Mundo Hispánico 14]. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Olbertz, H. & Muysken, P. C. 2005. Introduccion. In *Encuentros y conflictos: Bilingüismo y contacto de lenguas en el mundo andino* [Lengua y Sociedad en el Mundo Hispánico 14], H. Olbertz & P. C. Muysken (eds), 7–10. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Hinskens, F. & Muysken, P. C. 2007. The talk of the town: Languages in Amsterdam. *Zeitschrift für Literaturwissenschaft und Linguistik* 37(4): 7–23.
- Kuiken, F. & Muysken, P. C. 2009. Doe je voordeel met meertaligheid (Take advantage of bilingualism). *Didaktief* 39(6): 36–37.
- Muysken, P. C. 2010. Multilingualism and LADO. In *Language and Origin: The Role of Language in European Asylum Procedures: A Linguistic and Legal Survey*, K. Zwaan, M. Verrips, & P. Muysken (eds), 89–98. Nijmegen: Wolf Legal Publishers.
- Muysken, P. C. 2011. Bridges over troubled waters: Theoretical linguistics and multilingualism research. *Stellenbosch Papers in Linguistics* 40. <https://doi.org/10.5774/40-0-37>
- Muysken, P. C. 2012. Two linguistic systems in contact: Grammar, phonology, and lexicon. In *The Handbook of Bilingualism and Multilingualism*, 2nd edn, T. K. Bhatia & W. C. Ritchie (eds), 193–215. Hoboken NJ: Wiley-Blackwell. [print version 2013] <https://doi.org/10.1002/9781118332382.ch8>
- Muysken, P. C. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism. Language and Cognition* 16: 709–730. <https://doi.org/10.1017/S1366728912000727>
- Muysken, P. C. 2016. Bilingual complexes: The perspective of the Gradient Symbolic Computation framework. *Bilingualism: Language and Cognition* 19: 891–892. <https://doi.org/10.1017/S1366728916000031>
- Kootstra, G. J. & Muysken, P. C. 2017. Cross-linguistic priming in bilinguals: Multidisciplinary perspectives on language processing, acquisition, and change. *Bilingualism: Language and Cognition* 20: 215–218. [online 2016] <https://doi.org/10.1017/S1366728916001127>
- Muysken, P. C. 2017. The global and the local: Making comparisons possible. In *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World* [Issues in Hispanic and Lusophone Linguistics 13], K. Bellamy, M. W. Child, P. González, A. Muntendam & M. C. Parafita Couto (eds), 313–321. Amsterdam: John Benjamins. <https://doi.org/10.1075/ihll.13.13muy>
- Muysken, P. C. 2019. Language and origin: The perspective of multilingualism. In *Language Analysis for the Determination of Origin: Current Perspectives and New Directions* [Language Policy 16], P. Patrick, M. Schmid & K. Zwaan (eds), 119–130. Cham: Springer.

- Aalberse, S. & Muysken, P. C. 2019. Bilingualism and language contact. In *The Cambridge Handbook of Bilingualism*, A. De Houwer & L. Ortega (eds), 524–543. Cambridge: CUP.  
<https://doi.org/10.1017/9781316831922.027>
- Cornips, L. & Muysken, P. C. (eds). 2019. Language in the mines. *International Journal of the Sociology of Language* 258: 1–145. (ISSN: 1613-3668).
- Cornips, L. & Muysken, P. C. 2019. Introduction: Language in the mines. In *Language in the Mines*, L. Cornips & P. C. Muysken (eds). *International Journal of the Sociology of Language* 258: 1–12. <https://doi.org/10.1515/ijsl-2019-2026>
- Muysken, P. C. 2019. Multilingualism and mixed language in the mines of Potosí (Bolivia). In *Language in the Mines*, L. Cornips & P. C. Muysken (eds). *International Journal of the Sociology of Language* 258: 121–142. <https://doi.org/10.1515/ijsl-2019-2031>

#### 2.4.4 Code-switching/code-mixing

- DiSciullo, A. M., Muysken, P. C. & Singh, R. 1986. Government and code-mixing. *Journal of Linguistics* 22: 1–24. <https://doi.org/10.1017/S0022226700010537>
- Muysken, P. C. 1987. Neutrality in code mixing. In *Eigen en vreemd. Identiteit en ontlening in taal, literatuur en beeldende kunst (Handelingen van het 39e Nederlandse Filologencongres 1986)*, 359–373. Amsterdam: VU Boekhandel/Uitgeverij.
- Muysken, P. C. 1987. Kodewisseling: De beperkingen van Appel (Code-switching: The limitations of Appel). *TTT: Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 7: 201–203.
- Muysken, P. C. 1988. Book review. J. Hill & K. Hill, *Speaking Mexicano: Dynamics of Syncretic Language in Central Mexico*, (1986). *Applied Psycholinguistics* 9: 425–427.
- Muysken, P. C. 1993. Constraints on code-mixing: Indic mixed verbal compounds. In *Papers from the ESF Summer School on Code-Switching and Language Contact, Pavia, September 1992*, 195–208. Strasbourg: European Science Foundation.
- Muysken, P. C. 1994. Towards a typology of code-switching. In *Las lenguas en la Europa comunitaria, Vol. 3: Las lenguas de minorías*, F. Sierra Martínez, M. Pujol Berché & H. den Boer (eds.), 11–21. Amsterdam: Rodopi.
- Muysken, P. C. & van Hout, R. W. N. M. 1995. Insertion, alternation, and congruent lexicalization, corpus-based approaches to bilingual speech. In *Summer School on Code-Switching and Language Contact, Ljouwert/Leeuwarden 14–17 September 1994 (Network on Code-switching and Language Contact)*, 302–307. Ljouwert/Leeuwarden: Fryske Akademy.
- Milroy, L. & Muysken, P. C. (eds). 1995. *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*. Cambridge: CUP.
- Milroy, L. & Muysken, P. C. 1995. Introduction: Code-switching and bilingualism research. In *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, L. Milroy & P. C. Muysken (eds), 1–14. Cambridge: CUP.
- Muysken, P. C. 1995. Code-switching and grammatical theory. In *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, L. Milroy & P. C. Muysken (eds), 177–198. Cambridge: CUP.
- Muysken, P. C. & de Rooij, V. 1995. Review article. C. Myers-Scotton, *Social Motivations for Codeswitching: Evidence from Africa* (1993); *Duelling Languages: Grammatical Structure in Codeswitching* (1993). *Linguistics* 33: 1043–1051. (ISSN: 0024-3949).
- Muysken, P. C., Kook, H. & Vedder, P. 1996. Papiamentu/Dutch code-switching in bilingual parent-child reading. *Applied Psycholinguistics* 17: 485–505. (ISSN: 0142-7164).

- Muysken, P. C. 1997. Code-switching processes: Alternation, insertion, congruent lexicalization. In *Language Choices: Conditions, Constraints, and Consequences* [IMPACT: Studies in Language, Culture and Society 1], M. Pütz (ed.), 361–380. Amsterdam: John Benjamins. <https://doi.org/10.1075/impact.1.25muy>
- Muysken, P. C. (ed.). 1999. *Sociolingüística: Lenguas en contacto*. Amsterdam: Rodopi.
- Muysken, P. C. 1999. Patterns of language mixing in minority languages in the Netherlands. In *Sociolingüística: Lenguas en contacto*, P. C. Muysken (ed.), 7–8. Amsterdam: Rodopi.
- Muysken, P. C. 2000. *Bilingual Speech: A Typology of Code-mixing*. Cambridge: CUP.
- Muysken, P. C. 2007. Code-switching and grammatical theory (reprinted from Muysken 1995 with a postscript). In *The Bilingualism Reader*, (2nd edn), L. Wei (ed.), 280–297. Milton Park: Routledge.
- Deuchar, M., Muysken, P. C. & Wang, S-L. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10: 298–340. <https://doi.org/10.2167/beb445.0>
- Gullberg, M., Indefrey, P. & Muysken, P. C. 2009. Research techniques for the study of code-switching. In *The Cambridge Handbook of Linguistic Code-switching*, B. E. Bullock & A. J. Toribio (eds), 21–39. Cambridge: CUP.
- Muysken, P. C. 2011. Codeswitching. In *The Cambridge Handbook of Sociolinguistics*, R. Mesthrie (ed.), 301–314. Cambridge: CUP.
- Khan, A. & Muysken, P. C. 2014. Strategies for incorporating nouns and verbs in code-mixing: The case of Pashto-English bilingual speech. *Lapurdum* 18: 97–137. <https://doi.org/10.4000/lapurdum.2514>
- Auer, P., Moyer, M. & Muysken, P. C. 2014. Connecting approaches to code-switching in West Africa. *International Journal of Bilingualism* 18: 447–454. <https://doi.org/10.1177/1367006913481146>

#### 2.4.5 Mixed language studies

- Muysken, P. C. 1979. La mezcla entre quechua y castellano. *Lexis. Revista de Lingüística y Literatura* 3: 41–56. (ISSN: 0254-9239).
- 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. 1988. Media Lengua and linguistic theory. *Canadian Journal of Linguistics* 33: 409–422. (ISSN: 0008-4131).
- Bakker, P. & Muysken, P. C. 1994. Mixed languages and language intertwining. In *Pidgins and Creoles: An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 41–52. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.08bak>
- Muysken, P. C. 1994. Media Lengua. In *Mixed Languages: 15 case studies in language intertwining*, P. Bakker & M. Mous (eds), 201–205. Amsterdam: IFOTT. ISBN: 90-74698-14-X.
- Muysken, P. C. 1994. Callahuaya. In *Mixed languages: 15 Case Studies in Language Intertwining*, P. Bakker & M. Mous (eds), 207–211. Amsterdam: IFOTT. (ISBN: 90-74698-14-X).
- Muysken, P. C. 1997. Media Lengua in Ecuador. In *Atlas of Languages of Intercultural Communications in the Pacific, Asia, and the Americas*, S. A. Wurm, P. Mühlhäusler & D. T. Tryon (eds.), 1335–1338. Berlin: Mouton de Gruyter.

- Muysken, P. C. 1997. Callahuaya in Bolivia. In *Atlas of Languages of Intercultural Communications in the Pacific, Asia, and the Americas*, S. A. Wurm, P. Mühlhäusler & D. T. Tryon (eds.), 1339–1342. Berlin: Mouton de Gruyter.
- Muysken, P. C. 1997. Media Lengua. In *Contact Languages: A Wider Perspective* [Creole Language Library 17], S. G. Thomason (ed.), 365–426. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.17.13muy>
- Muysken, P. C. 1997. Callahuaya. In *Contact Languages: A Wider Perspective* [Creole Language Library 17], S. G. Thomason (ed.), 427–447. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.17.14muy>
- Muysken, P. C. 2001. Intertwined languages. In *Concise Encyclopedia of Sociolinguistics*, R. Mesthrie (ed.), 481–483. Oxford: Elsevier Science.
- Muysken, P. C. 2006. Mixed codes. In *Handbook of Multilingualism and Multilingual Communication*, P. Auer & L. Wei (eds), 315–339. Berlin: Mouton de Gruyter.
- Muysken, P. C. 2009. Kallawayá. In *Lenguas de Bolivia*, Vol. 2: Amazonia, M. Crevels & P. C. Muysken (eds), 147–167. La Paz: Plural editores.
- van der Wouden, T. & Muysken, P. C. 2012. Introduction. In *Roots of Afrikaans. Selected Writings of Hans den Besten* [Creole Language Library 44], T. van der Wouden (ed.), 1–5. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.44.01int>
- Muysken, P. C. 2013. Media Lengua. In *The Survey of Pidgin & Creole Languages, Vol. 3: Contact Languages Based on Languages from Africa, Asia, Australia, and the Americas*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 143–148. Oxford: OUP.
- Muysken, P. C. 2013. Media Lengua structure dataset. In *The Atlas of Pidgin and Creole Language Structures Online*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Muysken, P. C. 2014. Déjà voodoo or new trails ahead? Re-evaluating the mixing typology model. In *Linguistic Variation: Confronting Fact and Theory*, R. Torres Cacoullós, N. Dion & A. Lapiere (eds), 243–262. New York NY: Francis & Taylor.
- Muysken, P. C. 2017. Multilingüismo y lenguaje mezclado en las minas de Potosí (Bolivia). *Linguística* 33(2): 97–124. <https://doi.org/10.5935/2079-312X.20170019>

#### 2.4.6 Areal linguistics

- Muysken, P. C. 2000. From Linguistic areas to areal linguistics: A research proposal. In *Languages in Contact*, D. Gilbers, J. Nerbonne & J. Schaeken (eds), 263–275. Amsterdam: Rodopi.
- Muysken, P. C. 2008. *From Linguistic Area to Areal Linguistics* [Studies in Language Companion Series 90]. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.90>
- Muysken, P. C. 2012. Modelling the Quechua-Aymara relationship: Structural features, sociolinguistic scenarios and possible archeological evidence. A cross-disciplinary exploration of prehistory. In *Archaeology and Language in the Andes*, P. Heggarty & D. Beresford-Jones (eds). *Proceedings of the British Academy* 173: 85–110.
- O'Connor, L. M. & Muysken, P. C. (eds). 2014. *The Native Languages of South America. Origins, Development, Typology*. Cambridge: CUP.
- Muysken, P. C. & O'Connor, L. M. 2014. Introduction: South American indigenous languages; genealogy, typology, contacts. In *The Native Languages of South America. Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 1–26. Cambridge: CUP.

- van de Kerke, S. & Muysken, P. C. 2014. The Andean matrix. In *The Native Languages of South America. Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 126–151. Cambridge: CUP.
- Muysken, P. C., Hammarström, H., Birchall, J. T. R., Danielsen, S., Erikson, L., Vilacy Galucio, A., van Gijn, R., van de Kerke, S., Kolipakam, V., Krasnoukhova, O. V., Müller, N. J. & O'Connor, L. M. 2014. The languages of South America: Deep families, areal relationships, and language contact. In *The Native Languages of South America. Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 299–322. Cambridge: CUP.
- Muysken, P. C., Hammarström, H., Birchall, J. T. R., van Gijn, R., Krasnoukhova, O. V. & Müller, N. J. 2014. Linguistic areas, bottom up or top down? The case of the Guapore- Mamore region. In *Areal Linguistics in South America*, B. Comrie & L. Golluscio (eds), 205–237. Berlin: De Gruyter Mouton.
- Hannß, K. & Muysken, P. C. 2014. Reduplication in Andean languages. In *Reduplication in Indigenous Languages of South America*, G. Goodwin Gomez & H. van der Voort (eds), 39–76. Leiden: Brill. [https://doi.org/10.1163/9789004272415\\_004](https://doi.org/10.1163/9789004272415_004)
- Muysken, P. C. 2016. From Colombo to Athens. An areal and universalist perspective on bilingual compound verbs. *Languages* 1(2). <https://doi.org/10.3390/languages1010002>
- Muysken, P. C. & van Gijn, R. 2016. Linguistic areas. In *Oxford Bibliographies Online*. Oxford: OUP.
- van Gijn, R., Hammarström, H., van de Kerke, S., Krasnoukhova, O. V. & Muysken, P. C. 2017. Linguistic areas, linguistic convergence, and river systems in South America. In *The Cambridge Handbook of Areal Linguistics, Part II: Case Studies for Areal Linguistics*, R. Hickey (ed.), 964–996. Cambridge: CUP.

#### 2.4.7 *The languages of immigrants*

##### 2.4.7.1 *Second language learning by foreign workers*

- Appel, R., Cruson, C., Muysken, P. C. & de Vries, J. W. (eds). 1980. *Taalproblemen van buitenlandse arbeiders en hun kinderen* (Language problems of foreign workers and their children). Muiderberg: Coutinho.
- Muysken, P. C. & de Vries, J. W. 1980. Een sociolinguistische kijk op het taalgebruik van buitenlandse arbeiders (A sociolinguistic examination of the language use of foreign workers). In *Taalproblemen van buitenlandse arbeiders en hun kinderen*, R. Appel, C. Cruson, P. C. Muysken & J. W. de Vries (eds), 99–111. Muiderberg: Coutinho.
- Muysken, P. C. 1980. Taalkunde en de problemen van buitenlandse arbeiders en hun gezinnen. (Linguistics and the problems of foreign workers and their families). *Wetenschap en Samenleving* 24: 12–17. (ISSN: 0043-4442).
- Snow, C. E., van Eeden, R. & Muysken, P. C. 1981. The interactional origins of foreigner talk: Municipal employees and foreign workers. *International Journal of the Sociology of Language* 28: 81–91. (ISSN: 0165-2516).
- Jansen, B., Lalleman, J. & Muysken, P. C. 1981. The alternation hypothesis: acquisition of Dutch word order by Turkish and Moroccan foreign workers. *Language: A Journal of Applied Linguistics* 31: 315–336. (ISSN: 0023-8333).
- Muysken, P. C. 1981. Developing grammars: The acquisition of German Syntax by foreign workers. *Lingua: An International of General Linguistics* 53: 375–380. (ISSN: 0024-3841).

- Appel, R. & Muysken, P. C. 1982. Inleiding: Taal en minderheden (Introduction: Language and minorities). *TTT: Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 3: 97–101. (ISSN: 0167-4773).
- Appel, R., Muysken, P. C. & Vermeer, A. R. 1983. Taal en minderheden (Language and minorities). *TTT: Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 3: 97–191. (ISSN: 0167-4773).
- Muysken, P. C. 1984. Attitudes and experiences of discrimination: The Netherlandic of Moroccan foreign workers. In *Sociolinguistics in the Low Countries* [Studies in the Sciences of Language Series, 5] K. Deprez (ed.), 333–356. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/ssl5.5.17muy>
- Claahsen, H. & Muysken, P. C. 1986. The accessibility of universal grammar to adult and child learners: A study of the acquisition of German word order. *Second Language Research* 2: 93–119. (ISSN: 0267-6583).

#### 2.4.7.2 *Ethnolects*

- Hinskens, F. & Muysken, P. C. 2007. De nieuwe rijken van het Nederlands (The new rich of the Dutch language). *Taalschrift. Tijdschrift over Taal en Taalbeleid*. Retrieved from: <<http://taalschrift.org/discussie/001582.html>> (22 April, 2020).
- Muysken, P. C. 2010. Ethnolects as a multidimensional phenomenon. In *Language Contact: New Perspectives* [Impact: Studies in Language, Culture and Society 28] M. Norde, B. de Jonge & C. Hasselblatt (eds), 1–25. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/impact.28.02muy>
- Muysken, P. C. & Rott, J. A. 2013. Ethnolect studies in the German and Netherlandic are: An overview. In *Multilingualism and Language Diversity in Urban Areas. Acquisition, Identities, Space, Education* [Hamburg Studies on Linguistic Diversity 1] P. Siemund, I. Gogolin, M. E. Schulz & J. Davydova (eds), 177–206. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/hsl.1.09muy>
- Muysken, P. C. 2013. Ethnolects of Dutch. In *Language and Space: Dutch*, F. Hinskens & J. Taelde-man (eds), 729–761. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110261332.739>
- Hinskens, F., Mourigh, K. & Muysken, P. C. 2020. Urban contact dialects in the Netherlands. To appear in *Urban Contact Dialects and Language Change: Insights from the Global North and South*, P. Kerswill & H. Wiese (eds). London: Routledge.
- Hinskens, F., van Hout, R., Muysken, P. C. & van Wijngaarden, A. 2020. Variation and change in grammatical gender marking: the case of Dutch ethnolects. To appear in *Linguistics*.

#### 2.4.7.3 *Heritage languages*

- Aalberse, S. & Muysken, P. C. 2013. Language contact in heritage languages in the Netherlands. In *Linguistic Superdiversity in Urban Areas: Research Approaches* [Hamburg Studies on Linguistic Diversity 2], J. Duarte & I. Gogolin (eds), 253–274. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/hsl.2.16aal>
- Aalberse, S., Backus, A. & Muysken, P. C. 2019. *Heritage Languages: A Language Contact Approach* [Studies in Bilingualism 58]. Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.58>



### 2.4.8 *Regional studies (sub-continental)*

- Muysken, P. C. 1981. Spaans en Quechua in Ecuador (Spanish and Quechua in Ecuador). *TTT: Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 1: 124–138.
- Charry, E., Koefoed, G., & Muysken, P. C. (eds) 1983. *De talen van Suriname. Achtergronden en ontwikkelingen* (The languages of Surinam. Background and developments). Muiderberg: Coutinho.
- Muysken, P. C. 1983. Inleiding (Introduction). In *De talen van Suriname. Achtergronden en ontwikkelingen*, E. Charry, G. Koefoed & P. C. Muysken (eds), 9–33. Muiderberg: Coutinho.
- Crevels, M. & Muysken, P. C. (eds). 2009–2014. *Lenguas de Bolivia*, Vol. 1–4. La Paz: Plural editores.
- Crevels, M. & Muysken, P. C. 2009. Lenguas de Bolivia: Presentación y antecedentes. In *Lenguas de Bolivia, Vol. 4: Temas nacionales*, M. Crevels & P. C. Muysken (eds), 13–26. La Paz: Plural editores.
- Crevels, M. & Muysken, P. C. 2011. La Amazonía boliviana: Presentación y antecedentes. In *Lenguas de Bolivia, Vol. 2: Amazonía*, M. Crevels & P. C. Muysken (eds), 13–26. La Paz: Plural editores.
- van Gijn, R., Haude, K. & P. C. Muysken. 2011. Subordination in South America: An overview. In *Subordination in Native South American Languages* [Typological Studies in Language 97], R. van Gijn, K. Haude & P. C. Muysken (eds), 1–24. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/tsl.97.01van>
- Crevels, M. & Muysken, P. C. 2014. El oriente: presentación y antecedentes. In *Lenguas de Bolivia, Vol. 3: Oriente*, M. Crevels & P. C. Muysken (eds), 13–17. La Paz: Plural editores.

### 2.4.9 *Borrowing*

- van Hout, R. W. N. M. & Muysken, P. C. 1994. Modelling lexical borrowability. *Language Variation and Change* 6(4): 39–63. (ISSN: 0954-3945).
- Hekking, E. & Muysken, P. C. 1995. Otomí y quechua: Una comparación de los elementos gramaticales prestados del español. In *Lenguas en contacto en Hispanoamérica: Nuevos enfoques*, K. Zimmerman (ed.), 101–118. Frankfurt: Vervuert.
- Muysken, P. C. 1999. Three processes of borrowing: Borrowability revisited. In *Bilingualism and migration*, G. Extra & L. Verhoeven (eds), 229–246. Berlin: Mouton de Gruyter.
- Muysken, P. C. 2001. Spanish grammatical elements in Bolivian Quechua: The Transcripciones Quechuas corpus. In *Lo propio y lo ajeno en las lenguas austronésicas y amerindias: Procesos interculturales en el contacto de lenguas indígenas con el español en el Pacífico e Hispanoamérica*, K. Zimmerman & T. Stolz (eds), 59–82. Madrid: Iberoamericana. (ISBN: 9783865278906).
- Muysken, P. C. 2011. Root-affix asymmetries in language contact. *The International Journal of Bilingualism* 15: 1–15. (ISSN: 1367-0069).
- Muysken, P. C. 2011. Spanish affixes in the Quechua languages: A multidimensional perspective. *Lingua* 122: 481–493. <https://doi.org/10.1016/j.lingua.2011.10.003>
- Muysken, P. C. 2011. Préstamos morfológicos del español en el quechua. In *Préstamos morfológicos del español en el quechua*, W. Adelaar, P. Valenzuela Bismarck & R. Zariquiey Biondi (eds), 425–443. Lima: Fondo Editorial Pontificia Universidad Católica del Perú.
- Muysken, P. C. 2016. Review article. M. Simonovic, Lexicon immigration service. Prolegomena to a theory of loanword integration. *Nederlandse Taalkunde* 21: 178–180.

## 2.5 Final remarks

We can't claim that the bibliography is complete, for one very good reason. The author in question never stops writing! We are satisfied however that it is more accurate than a number of the sources we had at our disposal, in matters such as publication dates, order of authors' names, and even the odd title. Some works will inevitably have been overlooked, and in other cases a lack of access to, or knowledge regarding a particular article, may have caused us to misclassify it, or even omit it.

## Part 3. The Festschrift

This volume includes twelve chapters organized around four sub-disciplines:

- i. Creole languages and creole studies;
- ii. Linguistic areas;
- iii. Mixed Languages and language mixing;
- iv. Sociolinguistic aspects of language contact.

### 3.1 Creole languages and creole studies

#### 3.1.1 *Moving into and out of Sranan: Multiple effects of contact* (Essegbey & Bruyn)

In this chapter, *Essegbey and Bruyn* reflect on the question of language change, and how it may proceed in a contact situation, such as in colonial Suriname, which saw the emergence of Sranan and other Suriname creole varieties in the early 18th century. Earlier sources are lacking. Building on Leonard Talmy's influential work (Talmy 2000) on complex motion events, and how they map onto event structure, *Essegbey and Bruyn* show that Sranan expresses motions in and out of enclosure differently from its lexifier language, English. This diachronic study indicates that Sranan complex motion events share semantic and structural properties with the Gbe languages, which together with Kikongo, represent the Niger-Congo source languages of this Suriname creole. The chapter further shows that these complex motion events were modelled on a Gbe pattern in early Sranan. For instance, Sranan is a serialising language and resorts to lexical verbs to encode Path, Manner, and Motion, notions which are expressed through verb-adposition complexes in English. In this regard, Sranan and the Gbe languages can be regarded as equipollently-framed languages, while English is commonly analysed as a satellite-framed language. Despite their similarities, Sranan and the Gbe

languages are not isomorphic, however: The creole has developed a new system that is not present in its source languages. *Essegbey and Bruyn* conclude that this contact-induced change is an instance of pattern replication “where the Gbe languages of West Africa, [...] serve as the external source”. Since most creoles emerge in a situation of multilingualism, it is important to understand how such a change can evolve. The history of Sranan appears interesting in this respect because the language has been in intensive contact with Dutch since the late 1660s. Dutch, however, displays different strategies from both English and the Gbe languages when it comes to complex motion events. It appears from the discussion that prolonged contact with Dutch has given rise to new developments in which Dutch patterns are being integrated in Sranan, thus making it look less like the Gbe languages. This chapter therefore shows that a diachronic perspective on contact languages is necessary to understand how changes occur in the first place, but crucially how a change can evolve given different contact settings or different speaker profiles.

### 3.1.2 *Sociolinguistic characteristics of the English-lexifier contact languages of West Africa (Yakpo)*

In this chapter, *Kofi Yakpo* discusses the socio-cultural and political dynamics of five major Afro-Caribbean English-lexifier contact languages, AEC (i.e., Nigerian Pidgin, Cameroon Pidgin, Ghanaian Pidgin English, Pichi (Equatorial Guinea), and Krio (Sierra Leone)). Together, these languages involve the largest population of English-lexifier contact languages in the world. As the African population grows, with intense multilingualism in the urban centres, these vernaculars gain in communicative functions, expand to new domains such as the media, art, literature, movie and music industries, and therefore become very attractive to new speakers, both as urban means of communication and as identity markers. While this characterization applies to each AEC studied to some degree, *Yakpo's* comparative approach offers a precise evaluation of each AEC within its national community and within the region. This comparison indicates that while each AEC seems to thrive within its speech community, they all “struggle with a low sociolinguistic prestige” and the virtual lack of a state policy toward their inclusion in the educational and formal contexts. From the perspective of some African government officials or policy makers, these ‘low-level’ or ‘corrupt’ languages cannot represent the future of modern Africa. There is still a strong belief that only ‘civilized languages’ (which happen to be the language of former colonial powers) can allow a sustainable education and lead to prosperity. *Yakpo* shows, however, that such a view might be fundamentally wrong given the vitality of these AECs, their adaptability to new contexts, and their shared function as common denominators within highly multi-ethnic and multi-lingual communities. Indeed, one cannot blame these AECs for being the instrument of power of any specific ethnic or socio-economic group.

The chapter also shows that the absence of state policies for the development of AECs is comparable to the fate of other native African languages which are subject to “exoglossic language policies”, which uniquely promote the use of colonial languages (e.g., English, Spanish (in Equatorial Guinea) in formal domains (e.g., education, administration).

### 3.1.3 *The quest for non-European creoles: Is Kukama (Brazil, Peru) a creole language? (Bakker)*

In a quest for ‘creoleness’, *Peter Bakker’s* chapter focusses on the question of whether a somewhat archaic creole could exist, outside the classical cases discussed in the literature, which involve creoles in the former colonial areas around the Atlantic Ocean, the Indian Ocean, the Pacific, and other former colonial territories in Africa and Asia. From a typological perspective, this would make the claim that creoles form a typological class more convincing since one could show that such language types do not only arise in typical colonial historical settings such as the ones which gave birth to say, Caribbean creoles. In terms of Bakker’s creole typology, such a discovery would be the missing link between ‘modern colonial creoles’ which arose in the last four centuries, and much older languages, which have gradually lost their creoleness. *Peter Bakker* identifies one possible remote creole in Kukama, a language that has been classified variously by descriptivist linguists, for instance as a member of the Tupi-Guarani family, as a contact language, or as a creole. After submitting Kukama to a ‘creole test’ based on work by various scholars advocating a creole prototype, *Peter Bakker* concludes that “Kukama language does not quite fit to proposals for creole prototypes.” Nevertheless, he speculates that the impossibility of clearly classifying Kukama as a prototypical creole might be due to the fact that Kukama creolization happened in a remote past, maybe some “1500 years ago.”

### 3.1.4 *Are creoles a special type of language?: Methodological issues in new approaches to an old question (Kouwenberg & Singler)*

Studies on contact languages in general and on creoles in particular revolve around the question of whether these varieties are ‘normal’ languages, or whether they exhibit some specific characteristics that could set them apart from common human linguistic creations. The question seems a priori relevant to the study of human cognition. Indeed, under the assumption that languages are commonly transmitted from adults to children, that is, from adult L1 speakers to children L1 acquirers (i.e. so-called normal transmission), creoles could be seen as instantiating ‘abnormal transmission’ (or breaks in transmission) given that they are assumed by some authors to have developed in historical contexts in which the majority of their speakers were L2 speakers of mutually unintelligible languages. Creole languages would therefore represent a scenario in which transmission happens from a degenerated

L2 (referred to as a pidgin) to L1 speakers. L2 speakers have been shown in the literature to engage in learning strategies quite different from those of L1 learners. As a result, one could imagine that creoles, due to their unfortunate origin, would display a coherent set of such L2 strategies which would make them an exceptional typological class. For more than thirty years, this question has generated a lot of discussion within creolistics. In his 1988 article entitled ‘Are creoles a special type of language?’ Pieter Muysken addressed this question very systematically based on a structural analysis of several so-called creole features, and concluded: “[t]he very notion of a ‘creole’ language from the linguistic point of view tends to disappear if one looks closely; what we have is just a language” (300). In recent years, some scholars within creolistics have challenged this view and propose different tools (e.g., morphologically based metrics, phylogenetic trees) that can be used to show that creoles systematically cluster together as a language-type. Rather than adopting the ‘close scrutiny’ suggested by *Pieter Muysken* in his conclusion, these new studies mainly rely on quantity: an aggregate of disparate features somewhat arbitrarily associated with creoles and creoles only. *Kouwenberg and Singler’s* conclusion that such methods have been wrongly applied does not mean that computational methods are useless in creolistics and beyond. As the authors state, “computational methods and networks have the potential to reveal the classification of creoles based on their ancestry (Blasi, Michaelis & Haspelmath 2017), or can reveal the distribution of typological patterns, as for instance in the carefully conducted study of Danielsen, Dunn & Muysken (2011).” This study therefore proposes a shift of perspective from what makes creoles special, to rather how to account for their humanness, despite the horrific socio-historical conditions of their emergence.

### 3.2 Linguistic areas

#### 3.2.1 *Separating layers of information: The anatomy of contact zones (van Gijn)*

*Rik van Gijn’s* chapter proposes a new perspective on linguistic areas or *Sprachbünde*, and how the study of contact phenomena in such areas informs us on the communicative practices of their inhabitants, their ecology, and their history. Citing *Thomason’s* (2001: 99) definition of a linguistic area as “a geographical region containing a group of three or more languages that share some structural features as a result of contact,” *van Gijn* highlights the drawbacks of this notion, drawing from discussions in the literature. Indeed, the debate about the relevance of linguistic areas to the study of language and language variation over the past decade clearly shows that linguistic areas are hard to “define in a consistent and meaningful way”. The concept of linguistic area is problematic because (i) it implies a space (i.e.,

a geographic area) that often cannot be clearly identified; (ii) contact between a number of languages which cannot be established in any principled manner, (iii) a cluster of structural linguistic features for which one does not know how and when they emerged and spread across the relevant languages. These, and other complications, have led to different proposals ranging from a pure rejection of the notion of linguistic area *per se* to a redefinition of the objectives of this sub-field, as areal linguistics, that is, “the study of the distribution of features independently of geographical, cultural, or communicative contexts.” Building on this, *van Gijn* proposes a new perspective: *the anatomy of contact zones*. In a plea for the relevance of the notion of areal linguistics, he proposes “a conceptual change, moving away from the idea of a linguistic area as a well-circumscribed geographical area towards an idea of contact zones as windows on the past.” In this multidisciplinary perspective, involving geographical, cultural-historical, communicative, and (structural-) linguistic aspects, he develops a deterministic model in which linguistic areas reveal “the interaction between geography, human behaviour, and language”. In this model, “one layer (e.g. geography) fully explains another (e.g. cultural contact), which in turn fully explains communicative behaviour, and finally the distribution of shared features.” A crucial advantage of this model is that while it cannot explain all the relevant facts, it allows us to identify which aspects of contact and which features resist explanation and call for alternative approaches. As such, *van Gijn’s* chapter illustrates very well an essential aspect of *Pieter Muysken’s* work: continuity, parsimony, and a multifaceted approach to the study of contact phenomena and their distribution cross-linguistically.

### 3.2.2 *Areal diffusion of applicatives in the Amazon* (*Crevels & van der Voort*)

As discussed in *van Gijn’s* chapter, areal linguistics does not focus on identifying specific isoglosses with clearly identified boundaries, but rather focuses on identifying how certain linguistic traits spread and are distributed across genetically (and sometimes typologically) unrelated languages of a certain region. By tracking such linguistic expansion, not only does areal linguistics inform us on the history of speech communities, but it also unravels how some peculiar linguistic traits can be adapted from one language to another. Understanding such adaptations opens a window into how grammars emerge in the mind of language learners and speakers, and how they adapt to their communicative ecologies. *Mily Crevels and Hein van der Voort’s* chapter discusses a possible areal feature across North(-Western) and Southern Amazonia: the applicative morpheme *-tA* and its potential cognates. The authors show that many languages in this relatively large area exhibit an applicative-like morpheme often expressed as *-tA* (or some similar form). The morpheme tends to be a suffix, and displays a wide variety of argument structure

changing functions across these languages, ranging from the typical valency increase commonly associated with applicative morphemes to valency decrease as in Shawi and Mapudungun. *-tA* has also been identified in the function of a causative morpheme or verbalizer in some of the languages of this area. It therefore appears from this chapter that while this morpheme may have started out as an applicative morpheme, its expansion across the languages of the area leads to it acquiring new and sometimes diverging functions. What remains constant, however, is the capacity of this morpheme to impact argument structure in all these languages. In this regard, the authors conclude: “In view of the genetic diversity of the languages involved, and considering also other independent shared traits in classifier systems, we regard the ubiquity of the *-tA* applicative across the (North)western and Southern Amazon region as an areal trait”.

### 3.2.3 *Transfer of Swahili ‘until’ in contact with East African languages (Mous)*

A crucial aspect of areal linguistics is to understand how certain structural properties may spread from one language to other neighbouring languages. The question is not trivial since a particular structure may emerge accidentally in different languages or spread from one language to another. The situation in areal linguistics is therefore similar to the discussion in evolution biology over homology versus homoplasy. In the case of linguistics, the process that is usually assumed to mediate such changes is grammaticalization: the development of an open class item (usually a content word) into a closed class item: a grammatical marker. *Mous’s* chapter discusses the grammaticalization of the noun *mpaka* which means ‘boundary’ or ‘border’ into a grammatical item meaning ‘until’, and how this grammatical element has been borrowed into various East African languages. The chapter shows that the source of the grammaticalization path of *mpaka* is an N-N compound lacking the associative marker typically found in such Bantu expressions, and which ‘delimits’ the event expressed by the verb phrase temporally. As *Mous* concludes, “the main function is in the time interpretation of ‘until’. Though one could in principle hypothesize that *mpaka* emerged spontaneously in the relevant languages, this chapter argues that this is not the case. Instead, a survey of some thirty Bantu languages shows that many of them “borrowed” *mpaka* from Swahili in the function of *until*.”

### 3.3 Mixed languages and language mixing

#### 3.3.1 *Turkish-German code-switching patterns revisited: What naturalistic data can(not) tell us (Treffers-Daller)*

A recurrent question that arises in studies of language contact, language creation, and language change is that of the profile of the multilingual speaker, and what we can learn from understanding how such speakers use their different languages, sometimes mixing them in the same sentences or utterances. Pieter Muysken is one of the pioneers who proposed formalisms for code-switching/mixing, and brought such phenomena to the core of studies of speakers' competence and performance. More recently, bilingualism and its possible cognitive advantages have led to an increasing number of experimental studies on code-switching/mixing. As *Jeanine Treffers-Daller* shows clearly in her chapter, however, some experimental studies are conducted in total ignorance of previous discussions on language mixing and in particular with no reference to *Pieter Muysken's* seminal work on the typology of code-mixing. *Jeanine Treffers-Daller's* chapter bridges the gap between experimental studies and typological and formal studies on code-mixing, by offering a state-of-the-art review of patterns of mixing observed in naturalistic contexts, how such patterns instantiate *Muysken's* typology, and how they can be investigated in experimental settings. This chapter is therefore an interface between descriptive and formal studies of code-mixing and experimental studies on language mixing. One prominent aspect of the discussion in this chapter relates to the often described variation within and across speakers when it comes to language mixing. *Jeanine Treffers-Daller* argues that while speakers may display certain tendencies in their mixed outputs, many internal and external factors, such as the "language proficiency of speakers (and hearers), the frequency with which they use the languages, the practice they have in switching, the bilingual mode of the conversation, the attitudes towards mixing in the speech community" are at play as well. Wherever possible, these factors must be taken into account in experimental studies. As the author concludes, understanding what factors trigger language mixing and under which conditions, requires us to design ecologically valid experiments that can shed light on cognitive processes underlying this phenomenon.

#### 3.3.2 *Mixing and Semantic Transparency in the genesis of Yilan Japanese (Rojas Berscia)*

The quest for 'non-European creoles' is also at the core of *Luis Miguel Rojas Berscia's* chapter, though he does not subscribe to the structural definition of creoles. It is common practice in creolistics to define creoles on socio-historical grounds: the varieties which emerged as a consequence of the European colonial expansions starting in the 16th century. Such a definition, however, raises the question of whether



similar contact varieties emerged in colonial contexts not involving European languages (namely, Portuguese, Spanish, Dutch, French, English, German). Related to this question is whether such ‘non-European colonial contact languages’ share any structural properties with classic creoles of the Atlantic or Indian Oceans. Answering these questions can teach us a lot on how socio-historical, cultural, and political factors can affect the structure of language. In this regard, *Luis Miguel Rojas Berscia* discusses the case of a contact variety spoken in Yilan county (Taiwan), which emerged in the late nineteenth century and the first half of the twentieth century, when Taiwan was invaded and colonised by Japan. While this variety has been labelled as ‘Yilan Creole’ in the literature, this chapter shows that the language can be best described “as a case of Atayal+Japanese language mixing [...], whereby a large amount of lexicon, e.g. nouns, verbs, and particles, have been retained, maintaining not only their original phonological form, but operating in the original Atayal way, within an all-encompassing Japanese grammar skeleton.” The chapter further shows that the label ‘mixed language’ should be used with caution since Yilan Japanese, as the author labels it, shows different degrees of mixing strategies in which different grammatical patterns in the language point to different developmental paths arguably indicating different stages in the genesis of the language. *Luis Miguel Rojas Berscia* concludes “only after a careful description of the language is conducted, taking into account cross-generational variation, will we be able to classify this vernacular and claim a rightful spot for it in modern typologies of mixed languages”.

### 3.3.3 *Pottefers Cant, Groenstraat Bargoens, and the development of “have” and “be” in the wider context of contact (Smith & Hinskens)*

Studies on language contact commonly focus on socio-historical situations in which learners and speakers of different varieties are in contact, interact regularly, and learn aspects of their mutual languages. Questions that are often raised in such contexts are which learning hypotheses such learners and speakers employ, what type of mental grammars do they develop, and how do the output of such grammars affect change. Yet, a fundamental question underlying this research is to what extent learning strategies adopted by speakers remain constant across situations and generations of learners. In this regard, one context of language creation and language change that has remained virtually unexplored is that of secret languages. Do learners and speakers deploy the same learning strategies in creating secret languages? What happens when speakers of related secret languages meet? Do we observe similar linguistic patterns, or spread of features, as in common contact situations? The chapter by *Norval Smith and Frans Hinskens* addresses some of these questions. This chapter discusses two Dutch cryptolects: *Pottefers Cant* (‘Pot-repairers’

language’) and Groenstraat Bargoens. These two cryptolects were spoken within the “geographical space of the Brabant and Limburg dialect areas of Dutch dialects,” respectively. Their comparative study of these varieties involves the lexicon, morphology, syntax, and most importantly, a striking similarity between these languages: “the fact that various uses of “have” and “be” verbs, including “main verb”, auxiliary and copular usages are expressed by a single form.” Taking this similarity as resulting from contact, the authors conclude “we ascribe this shared structural pattern to early contact between Dutch and Belgian “travellers” and Romani, despite the fact that the languages of the first group are inflection-poor, whereas Romani languages are inflection-rich”. This chapter therefore shows that cognitive processes underlying contact phenomena should not be envisaged as isolated processes tied to specific language behaviour, but rather they should be placed on a continuum of cognitive processes that learners deploy in different communicative settings. This supports *Pieter Muysken’s* (2013) broad view on language contact and change.

### 3.4 Sociolinguistic aspects of language contact

#### 3.4.1 *Sociolinguistic enregisterment through languagecultural practices: (Cornips & de Rooij)*

Many chapters in this volume focus on the emergence of linguistic patterns in a contact situation and how such patterns can inform us on language speciation, linguistic variation, language and areal typology, and ultimately the human mind. *Leonie Cornips’ and Vincent de Rooij’s* chapter adds yet another dimension by focusing on language use and appropriation of specific linguistic features by speakers as a part of identity construction. This chapter discusses contemporary cultural and linguistic effects of language contact in Heerlen, the centre of the booming mining industry in the Dutch province of Limburg in the early 20th century, and as in the case of plantation societies in the colonial Caribbean, the attraction of waves of migrants and speakers of different ethnicities from within the Netherlands and Europe (e.g., Italy, Poland, Germany, Belgium, Slovenia, Austria, and Hungary). Contact between these different groups led to the emergence of a new variety, sometimes referred to as Heerlen Dutch: a variety seen as “neither dialect nor Dutch but something in between”. As this variety crystalized and could be heard, spoken, or felt as belonging to specific people in a specific location, i.e., a process of enregisterment, it appears interesting to investigate how new generations in a digital era relate to Heerlen Dutch, and which function they assign to it. To this end, this chapter investigates “the effects of sociolinguistic enregisterment of Heerlen Dutch in the carnivalesque summer song *Naar Talië/Naar Talia* ‘To (I)taly’, performed by the band the *Getske Boys* and circulated on the internet. “By selecting

a particular set of linguistic forms such as dialect, Dutch, in-between, Italian and English” this band uses these features, seen as belonging to “Heerlen-Noord and the speech of the coal miners who once lived there.” A central question in this chapter is “how do actors engage in processes of place-making in a specific locality in the province of Limburg in the Netherlands, i.e., Heerlen, through language practices on social media, creatively drawing on features from what are considered local and regional language varieties?” The authors conclude that this “case study has shown how particular linguistic forms are being re-enregistered as local Heerlen-Noord speech. The *Getske Boys*’ selection of these specific co-occurring forms is based on perceived past patterns of co-occurrences: an experiential knowledge, accumulated over the years, of the indexical ties between linguistic forms, specific (groups of) people and a specific place.”

#### 3.4.2 *Snow on the Danish Antilles?: Referee design in Virgin Island Dutch Creole (van Rossem)*

A major challenge in the study of contact phenomena and the emergence of creoles in some 18th century colonial territories is the availability of historical data. The scarcity of historical data in the creole world is sometimes used as an excuse for exceptionalist views on creole formation that claim to rely on ‘big numbers’ of features cross-linguistically. Thanks to *Pieter Muysken*’s team however, there have been some valuable diachronic studies such as *Adrienne Bruyn*’s and *Margot van den Berg*’s work on Early Sranan. Likewise, the work on Virgin Islands Dutch Creole conducted by *Hein van der Voort*, *Cefas van Rossem* and *Robbert van Sluijs* under *Pieter Muysken*’s supervision represents yet another valuable contribution to the study of the emergence of creoles and how the study of the competition between different varieties within a speech community can inform us on the history of that community, the profiles of its speakers, and to some extent the social dynamics regulating their interactions. In his chapter, *Cefas van Rossem* revisits the Clarin-NEHOL-corpus of Virgin Islands Dutch Creole within the framework of Allan Bell’s *Audience Design model* (Bell 1984) in order to find out to what extent the apparently artificial character and variability of the contents of these eighteenth-century missionary texts might hide some meaningful regularities that shed light on the communication situations in the colony. Indeed, the author concludes that even though eighteenth-century Virgin Islands Dutch Creole manuscripts often appear artificial, close-scrutiny reveals “some aspects of the natural vernacular.”

## Acknowledgements

Veenstra and Smith were primarily responsible for the bibliographical section, while Aboh was primarily responsible for the overview of the Festschrift. The order of author's names is basically irrelevant. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416591334 – SFB 1412 for Veenstra's involvement in this project is hereby greatly acknowledged.

## References

- Bell, A. 1984. Language style as audience design. *Language in Society* 13: 145–204.  
<https://doi.org/10.1017/S004740450001037X>
- Blasí, D. E., Michaelis, S. M. & Haspelmath, M. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1: 723–729.  
<https://doi.org/10.1038/s41562-017-0192-4>
- Cinque, G. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: OUP.
- Daniëlsen, S., Dunn, M. J., Muysken, P. 2011. The spread of the Arawakan languages: A view from structural phylogenetics. In *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archeology, Linguistics, and Ethnohistory*, A. Hornborg & J. D. Hill (eds), 173–196. Boulder CO: University of Colorado 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. 1988. Are creoles a special type of language? In *Linguistics: The Cambridge Survey*, Vol. 2, F. J. Newmeyer (ed.), 285–307. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511621055.017>
- Muysken, P. C. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism. Language and Cognition* 16: 709–730.  
<https://doi.org/10.1017/S1366728912000727>
- Talmy, L. 2000. *Toward a Cognitive Semantics*, 2 Vols. Cambridge MA: The MIT Press.
- Thomason, S. G. 2001. *Language Contact: An Introduction*. Edinburgh: EUP.



PART 1

# Creole languages and creole studies



# Moving into and out of Sranan

## Multiple effects of contact

James Essegbey and Adrienne Bruyn

University of Florida / Meertens Institute

In Sranan, a creole language of Suriname, motion in and out of locations is expressed differently from English, its primary lexifier language. Talmy (2000), among others, has shown that the expression of motion in English involves a verb that indicates the Manner of movement, e.g., *walk*, and, where relevant, a satellite that expresses Path, e.g., *into* or *out of*. In Sranan, by contrast, both Manner and Path are expressed by the verb, which may be part of a serial verb construction, such as *go* and *kmopo* in (ia) and (ib), respectively. A locative complement after the Path verb usually contains a specifying locative item, such as *ini*:

- (i) a. John waka go na a oso ini.  
J walk go LOC DEF house in(side)  
'John went into the house.'
- b. John waka kmopo na a oso ini.  
J walk come\_out LOC DEF house in(side)  
'John came out of the house.'

We argue that influence from the West-African Gbe languages that were part of the substrate in the early history of Sranan may account for the typological make-up of Sranan regarding the expression of movement. Recently, speakers have begun to use the Dutch preposition *uit* 'out' in Sranan to express moving out. This is related to a more general trend whereby simple locative prepositions have become possible in Sranan, whereas from around 1700 to the mid-20th century, the only simple locative preposition was *(n)a*. The recent developments imply that in this respect Sranan is moving away from its original state in which it is typologically closer to the Gbe languages, to become more like Dutch and English.

**Keywords:** expression of motion, substrate influence, contact-induced change, pattern replication, Sranan, Gbe languages, Ewe, English, Dutch



## 1. Introduction

In Sranan, a creole language of Suriname, motion in and out of locations is expressed differently from English, its primary lexifier language. In this paper, we shall analyse the differences in Talmy's (2000) framework and provide an account in terms of processes of language contact. The strategies prevalent in Sranan can be attributed to substrate influence from the West-African Gbe languages in the earlier stages of the history of the language, leading Sranan to resemble these substrate languages rather than English in this respect. Recent developments, however, tend to reduce the original similarities to the substrate languages to some extent.

According to Matras and Sakel (2007: 829–830) contact-induced change can lead to what they call replication of linguistic matter, which they describe as the “direct replication of morphemes and phonological shapes from a source language”, or pattern replication where “patterns of distribution, of grammatical and semantic meaning, and of formal-syntactic arrangement at various levels (discourse, clause, phrase, or word) [...] are modelled on an external source.” In this paper we argue that the expression of motion in Sranan involves pattern replication, predominantly, where the Gbe languages of West Africa, which we illustrate with data from Ewe, serve as the external source on which the patterns are modelled. However, we show that lately, intense contact with Dutch is giving rise to changes involving the replication of linguistic matter or, to put it simply, lexical borrowing. We therefore present a situation where the type of contact-induced change that occurs in one semantic domain (i.e. motion) appears to evolve over time (cf. Borges 2013).

Within the field of creole studies, the possibility of morphosyntactic influence from the substrate has gained acceptance since the late nineties. However, opinions still differ as to its precise effect and extent. The most explicit proposal in this respect, put forward by Lefebvre (1998) and Lumsden (1999) with regard to Haitian French Creole, is that creoles may have come about by a process of relexification: the replacement of a substrate word form of a lexical entry or lemma with a form from the lexifier language, maintaining the morphosyntactic and semantic properties of the original lemma. By comparing the expression of motion in Sranan with that in Ewe, we show why relexification fails to account properly for the phenomenon and, instead, a pattern-replication account that allows for a shift towards the replication of morphemes and phonological shape handles the data better.

Sranan developed in the latter half of the 17th century on the plantations of Suriname (‘Dutch Guyana’), which was an English colony from 1651 until 1667 before it came under Dutch rule. While English has contributed the larger part of its basic vocabulary (Smith 2001), continuing contact with Dutch has resulted in,

in particular, the incorporation of Dutch-derived words over the course of time. Influence from Dutch concerning morphosyntactic properties has been less conspicuous. It appears, however, that certain recent changes, including those to be discussed below, can be attributed to contact with Dutch.

As for the substrate, among the languages that were spoken by West Africans in Suriname in the early colonial period and, thus, may have influenced the developing creole language, the most important ones can be assumed to have been Fon and Ewe, both belonging to the Gbe dialect cluster of Kwa languages, and Kikongo, a Bantu language (Smith 1987; Arends 1995). Influences from these languages have been identified at the levels of phonology, vocabulary, and lexical semantics (e.g. Huttar 1975; Smith 1987, 1997) as well as morphosyntax (e.g. Huttar 1981; Migge 1998, 2000; Smith 2001; Essegbey 2005). The general picture that arises is that whereas various African languages including Kikongo as well as Gbe languages have contributed to the Sranan lexicon, instances of morphosyntactic influence are rather attributable to Gbe. In what follows, we compare Sranan with Ewe, which can, on the aspects under discussion, be regarded as representative for other Gbe varieties, including Fon.

After a brief introduction of Talmy's typology of the expression of motion, it will be pointed out that languages that make use of serial verb constructions (SVCs), as is the case with Sranan as well as Ewe, do not easily fit into this typology. In Section 3, we show how the characteristic ways of expressing motion into and out of an enclosure in Sranan differ from those of English, and that the strategies used in Sranan rather resemble those of Ewe. Section 4 presents particular similarities between the lexicalisation patterns of Sranan and Ewe, as evidence for the influence of the latter on the former. Although the lexicalisation patterns of particular motion verbs may differ between the two languages, as will be illustrated in Section 5, this does not affect our conclusion that the typological make-up of Sranan regarding the expression of movement differs from that of English, and that this is attributable to influence from the Gbe languages during the early history of Sranan. Section 6 discusses some recent syntactic and semantic changes leading to a deviation from the original situation – changes that can be attributed to influence from Dutch. The paper concludes with a summary and brief discussion of our findings in Section 7.

The Sranan data come from a variety of written texts and descriptive works for various stages of the language (van Dyk ca 1765; Schumann 1781, 1783; Focke 1855; Donicie 1954; de Drie 1985; Wilner 1994), and from data collected by Essegbey from Sranan speakers in the Netherlands. The latter include texts elicited on the basis of *Frog, where are you?* (Mayer 1969), a picture story book with 24 pictures,

about a boy who lost his frog and went with his dog to look for it. For Ewe and Dutch, we have drawn on the fact that Essegbey is a native speaker of the former, and Bruyn of the latter.<sup>1</sup>

## 2. The typology of a complex-motion event

In this section we introduce Talmy's well-known typology (1985, 2000, 2007), noting the problems that have been raised in connection with it, and one particular solution that has been proposed. We begin with a brief recap of his concepts. The basic motion event, according to Talmy, consists of one object (the Figure) moving or located with respect to another object (the reference object or Ground). In addition to the Figure and Ground, the motion event has two other components, namely, Motion and Path. Motion refers to the occurrence or non-occurrence of translational motion (i.e. displacement over time), while Path refers to the path followed in translational motion or the site occupied with respect to the Ground (Talmy 2000: 25).

Talmy argues that the world's languages fall into two categories depending on whether the core schema of a complex event that is expressible by a single clause (referred to as a macro-event) is expressed by the main verb or a satellite. The core schema is defined as "the association function that sets the figural entity into a particular relationship with the ground entity" (Talmy 2000: 218), and the satellite as "the grammatical category of any constituent other than a nominal or prepositional-phrase complement that is in a sister relation to the verb root" (Talmy 2000: 222). In a motion event, the core schema is the Path, hence languages are distinguished as to whether they express this component with a verb or a satellite. The distinction is captured in the sentences below:

- (1) a. Juan *salió* de la sala.  
b. John went out of the room.

Example (1a) illustrates that, in Spanish, the Path is encoded by the verb *salió* 'exited', while in the English equivalent in (1b), it is encoded by the satellite *out*. Talmy refers to the Spanish-type language as verb-framed and the English type as satellite-framed. An important issue for Talmy's typology is that the expression must be characteristic of the language which, in his terms, means that it must be colloquial rather than literary, frequent in occurrence rather than occasional and pervasive rather than limited (Talmy 2000: 222). Thus, although English has the word *exit*, it is not used characteristically to express leaving a location.

---

1. With Sranan examples, written sources are indicated; the spelling is as in the sources. In the case of Ewe, high tone is marked by an acute accent.

Examples (1a) and (1b) express a basic motion event and, hence, contain only the four components which we listed above. Talmy notes that the motion event can be associated with an external event that often bears the relation of Manner or Cause to it. He refers to this external event as the co-event. Sentences (2a) and (2b), taken from Talmy, illustrate the distinction between a translational motion expression with Manner and one with Cause respectively:

- (2) a. The pencil rolled off the table.
- b. The pencil blew off the table.

(2a) expresses the manner in which the Figure moves off the table while (2b) expresses the cause for it doing so. In this paper, we shall be concerned mainly with the expression of translational motion that is association with manner, although we shall touch on the expression of a basic motion event as well.

With regard to such complex events which contain a co-event, Talmy (2000) writes:

Languages with a framing satellite regularly map the co-event into the main verb, which can thus be called the co-event verb. On the other hand, languages with a framing verb map the co-event either onto a satellite or into an adjunct, typically an adpositional phrase or a gerundive-type constituent. (Talmy 2000: 222)

The two modes of packaging are represented schematically by Slobin (2000) as in Figures (1a) and (1b) below which involve the expression of translational manner of motion in and out of a location:

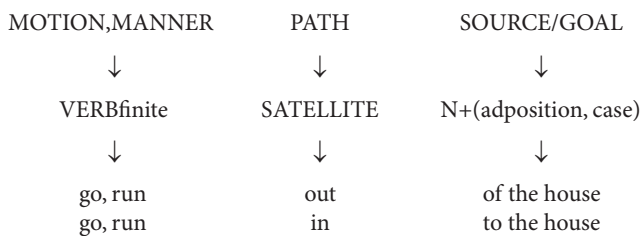


Figure 1a. Satellite-framed construction type

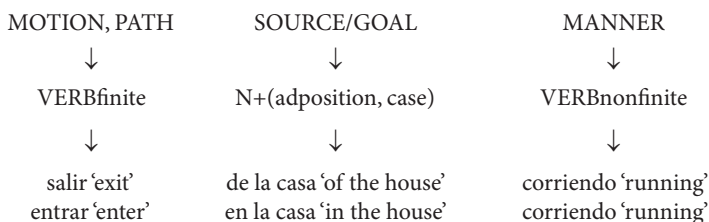


Figure 1b. Verb-framed construction type (Slobin 2000: 109)

In Figure (1a), the move-out and move-in information are contained in the satellites *out* and *in*, while the verbs *go* and *run* express deixis and manner of motion, respectively. In Figure (1b), the Path information is rather contained in the verbs *salir* ‘exit’ and *entrar* ‘enter’ while the Manner information is contained in the gerund *corriendo* ‘running’.

Various authors (cf. Slobin & Hoiting 1994; Schaefer & Gaines 1997; Zlatev & Yangklang 2004; Slobin 2004, 2006; Ameka & Essegbey 2013) have noted that serialising languages, which have two or more independent verbs occurring in a single clause, do not fit into Talmy’s typology. In such languages, both Manner and Path are expressed by verbs, none of which is subordinate to the other. This is illustrated by the example below from Ewe:

- (3) Kpé-á mli yi do-a me.  
 stone-DEF roll go hole-DEF in(side)  
 ‘The stone rolled into the hole.’

In (3), both the Manner and a component of the Path are expressed by independent verbs in a single clause. As we shall see in subsequent sections, Sranan, which is also a serialising language, expresses translational manner of motion in the same way. Slobin (2004, 2006) proposes that Talmy’s typology be expanded to include equipollently-framed languages. These are languages in which Path and Manner are expressed by equivalent grammatical forms. The typical construction types, depending on language, are:

MANNER VERB + PATH VERB. This is the construction type found in serial-verb languages found in Niger-Congo, Hmong-Mien, Sino-Tibetan, Tai-Kadai, Mon-Khmer, and Austronesian languages.

[MANNER + PATH] VERB: bipartite verb languages. These are found among Algonquian, Athabaskan, Hokan, and Klamath-Takelman languages.

MANNER PREVERB + PATH PREVERB + VERB: These occur in Jaminjung and other languages of Australia with small closed classes of inflecting verbs.

Talmy (2009) does not agree with this additional classification and argues that the constructions used by serialising languages such as Chinese are better treated as satellite-framed. Lambert-Brétière (2009) also argues that Fon is a satellite language. She bases her conclusions, among other reasons, on the fact that path verbs in Fon, unlike manner verbs, are chosen from a closed class. Van Putten (2017), however, argues that an analysis of motion expression discourse in Avatime, a Ghana Togo Mountain (GTM) language of the Kwa family spoken in Ghana, suggests that although it is equipollently framed because it is serialising, it behaves more like a verb-framed language when compared with languages like Chinese. We shall not focus on such details since they detract from the point we seek to make in this

paper. Instead, we show in the next section that in terms of Slobin's proposal, Sranan and Ewe would be classified as equipollently-framed languages to the extent that they both express Manner and Path with a verb.

### 3. Moving in and out

In this section we compare the expression of simple and complex motion in and out of a location in Sranan and its principal lexifier language English. On the surface, it appears as if the expression of simple motion into a location is the same in the two languages. However, motion out of a location and complex motions both in and out of a location show that both languages have different strategies. We show that in the expression of a complex motion event, Sranan, unlike English, which is satellite-framed, is equipollently-framed. In Section 3.2, we take a look at the expression of similar concepts in Ewe and show that the facts in Ewe are similar to those of Sranan.

#### 3.1 The expression of motion in English and Sranan

As discussed extensively by Talmy (1985, 2000) and several others, English is a satellite-framed language in that it expresses Path in a satellite. Thus, even though, as we noted above, the language has the verbs *enter* and *exit*, the characteristic way of expressing motion into and out of an enclosure involves the use of *into* and *out of*. Consider the expression of 'enter' in the English and Sranan examples below:

- (4) a. Kofi went into the room.  
 b. Kofi entered the room.
- (5) Kofi go na ini a kamra.  
 K go LOC in(side) DEF room  
 'Kofi went into the room.'

Sentences (4a) and (4b) show that English has the option between a deictic verb with the satellite *in(to)*, and a Path-conflating verb *enter* which also includes "boundary crossing" (à la Slobin 1997). By contrast, Sranan only uses the deictic-verb strategy in addition to a complex prepositional phrase made up of a general-purpose preposition *na* (or *a*) and a locational element *ini* 'in(side)'. This phrase is discussed more extensively below. Note that although the deictic-verb strategy in Sranan appears to be similar to that of its English counterpart exemplified by (4a), there is actually an important difference in that Sranan *ini* 'in(side)' does not express the same Path component as *in(to)* does in English. The difference is discussed in detail in Section 4. Now we turn to the expression of moving out in both languages:

- (6) a. Kofi went out of the room.  
 b. Kofi exited the room.
- (7) Kofi komoto/komopo na ini a kamra.  
 K come\_out LOC in(side) DEF room  
 ‘Kofi came out of the room.’

Just as we saw for the expression of movement into a location, that of movement out of a location also involves two strategies in English: a deictic verb with the satellite *out*, and the verb *exit*. We have already noted that *exit* is not the characteristic verb used to express movement out of a location in English. Instead, the deictic verb with the satellite is the preferred choice. Sranan uses the verb *komoto* or *komopo* ‘come out’ together with the same complex PP used in the expression of moving in. The form *komopo* is a fossilised reflex of English *come* plus *up* while *komoto* is fossilised *come* plus *out*. Both forms are phonological words in terms of stress. Furthermore, since English *out* by itself has no reflex in Sranan – that is, there is no corresponding free morpheme such as *\*oto* – the monomorphemic status is particularly clear with *komoto*. Although *opo* exists as an independent verb in Sranan, its meaning ‘rise, raise’ does not figure in *komopo* ‘come out’. Together with the fact that the verb ‘come’ has taken on the form *kon* in modern Sranan, it is clear that *komopo* too consists of just a single morpheme. While some researchers assign slightly different meanings to *komoto* and *komopo* (Sordam & Eersel 1989; Sebba 1987: 44) others simply treat them as variants of one verb meaning ‘come out, come from’ (Schumann 1783; Wilner 1994). Since both forms occur similarly in the construction under discussion, we follow the latter in treating the two verbs as variants, glossing both as ‘come out’. We shall use them interchangeably depending on the source from which the data are taken. Note that although etymologically, the forms contain deictic *come*, *komoto* and *komopo* do not have any deictic meaning. Instead, they express moving out. In this respect, the expression of moving out in Sranan is different from that in English, which uses the deictic verb *go* in combination with the satellite *out* to express the Path, as in (4a) above.

A comparison of the expression of complex motion which includes a Manner component in the two languages shows more clearly that the strategy in Sranan differs from that in English. Compare the expression of translational motion below:

- (8) a. Edgar kroipi go na ini a kamra.  
 E crawl go LOC in(side) DEF room  
 ‘Edgar crawled into the room.’
- b. Edgar kroipi kmopo na ini a kamra.  
 E crawl come\_out LOC in(side) DEF room  
 ‘Edgar crawled out of the room.’

These examples show that where translational motion includes a Manner component, Sranan uses an SVC. The first verb in this construction expresses manner while the second verb and the complex prepositional phrases are the same forms used in the expression of basic motion. In other words, a deictic verb and a complex prepositional phrase are added to a manner verb to express translational manner of motion into a location, while *komoto* or *komopo* and the complex PP are added to the verb to express motion out of a location. Although we are only concerned with these two paths, we should point out that other path verbs occur in the second-verb position in the SVC. These include *fadon* ‘fall’ and *opo* ‘rise’ (Sebba 1987). It is therefore clear that Sranan is equipollently-framed and not satellite-framed like English. In the next section, we discuss the expression of similar motion events in Ewe.

### 3.2 Moving in and out in Ewe

We begin this section with a look at the expression of basic motion in Ewe:

- (9) a. Kofi yi xɔ-a me.  
 K go room-DEF in(side)  
 ‘Kofi went into the room.’
- b. Kofi gé dɛ́ xɔ-a me.  
 K fall ALL room-DEF in(side)  
 ‘Kofi went into the room.’
- c. Kofi do le xɔ-a me.  
 K exit at room-DEF in(side)  
 ‘Kofi went out of the room.’

Examples (9a) and (9b) show that Ewe also has two ways of expressing movement into a location: one strategy involves the use of a deictic verb together with a phrase that is generally referred to in Ewe studies as the postpositional phrase (cf. Ameka 2003, but see below). The other strategy involves the use of a directional verb together with a complex PP headed by the allative preposition *dɛ́* (in some dialects *dɛ́* actually occurs as a verb meaning ‘enter’; cf. Aboh, Ameka & Essegbey 2002; Ameka 2003). Where Ewe uses a deictic verb to express movement into a location, as in (9a), it does not use a complex PP, and where it uses a complex PP, as in (9b), it has a directional verb. Since Sranan may use a deictic verb with a complex PP, as in (5) above, it looks as though both languages express this kind movement differently. However, as we show in the next section, the adpositional elements in both languages express the same thing. Before we get into that, note that the expression of movement out of a location is similar in the two languages: that is to say, in addition to a directional verb indicating ‘exit’, they both use a complex PP.



The similarity between the two languages is even more striking in their expression of complex translational motion. Consider the Ewe sentences below:

- (10) a. Kofi tá yi xɔ-a me.  
 K crawl go room-DEF in(side)  
 ‘Kofi crawled into the room.’
- b. Kofi tá do le xɔ-a me.  
 K crawl exit at room-DEF in(side)  
 ‘Kofi crawled out of the room.’

Note that Ewe too uses two verbs in an SVC to express translational motion, with the first verb expressing the manner of motion while the second verb expresses directional motion. The similarity between Ewe and Sranan, especially in movement out of a location that includes Manner is captured by Figure 2 below:

	MOTION, MANNER	PATH	SOURCE/GOAL			
	↓	↓	↓			
	VERB finite	VERB finite	N+(Postnominal item)			
	↓	↓	↓			
Sranan	<i>kroi</i> crawl	<i>komoto</i> exit	<i>na</i> LOC	<i>a</i> DEF	<i>oso</i> house	<i>ini</i> in(side)
Ewe	<i>tá</i> crawl	<i>do</i> exit	<i>le</i> LOC	<i>afé-á</i> house-DEF		<i>me</i> in(side)

Figure 2. Expressing complex translational motion in Sranan and Ewe

One might argue that the fact that both languages use an SVC to express translational motion that includes manner is not enough evidence to support the claim that the Sranan strategy is based on Gbe languages. This is more so since, as Essegbey (2001) argues, languages with different kinds of SVCs also express such meaning in similar fashion. More robust evidence that the Sranan strategy is modelled on that of Gbe languages (as illustrated specifically by Ewe) is provided by the strong similarity in the way both languages segment Path and Ground components, a similarity that cannot be attributed to their property as serialising languages. This is discussed in the next section.

#### 4. Path and Ground components

In the previous section, we mentioned that both Sranan and Ewe make use of a complex prepositional phrase in addition to the deictic or directional verbs. We explore this phenomenon in detail in this section. Among other things, we argue that both languages do not lexicalise ‘enter’ in a single lexical form but that this meaning is derived compositionally from the verb and the complex prepositional phrase. This is, however, not the case for the expression of motion out of an enclosure.

Talmy notes that Path is a composite constituent comprising Vector, Conformation and Deictic. “Vector comprises the basic types of arrival, traversal, and departure that a Figural schema can execute with respect to a Ground schema” (Talmy 2000: 53). The Conformation component “is a geometric complex that relates the fundamental Ground schema within a Motion-aspect formula to the schema for a full Ground object” (Talmy 2000: 54). Talmy notes further that each language lexicalises its own set of geometric complexes. What we intend to show here is that Ewe and Sranan adopt the same strategy in the lexicalisation of geometric complexes. To establish this, let us take another look at the expression of motion into an enclosure:

- (11) Kofi go na ini a kamra. (Sranan)  
 K go LOC in(side) DEF room  
 ‘Kofi went into the room.’
- (12) a. Kofi yi (dɛ́) xɔ-a me. (Ewe)  
 K go (ALL) room-DEF in(side)  
 ‘Kofi went into the room.’
- b. Kofi gé dɛ́ xɔ-a me.  
 K fall ALL room-DEF in(side)  
 ‘Kofi went into the room.’

A look at (11) and (12a) shows that, unlike Sranan, the use of a preposition is not obligatory in Ewe when the deictic verb is used to express Path. A second point is that while both languages have an element that we gloss as ‘in(side)’, that of Sranan precedes the Ground-expressing nominal while that of Ewe follows it. As will be discussed further in Section 6, Sranan also has – or at least had until recently – the possibility of having this element in postnominal position, as represented below:

- (13) Kofi go na a kamra ini.  
 K go LOC DEF room in(side)  
 ‘Kofi went into the room.’

Although there is an established tradition of calling such items postpositions in the Gbe literature (cf. Lefebvre & Brousseau 2002; Ameka 2003; Essegbey 2005), the situation is different in Sranan (cf. Plag 1998). In this paper, we shall not enter into the debate as to whether the postnominal elements should be considered postpositions or not. Also, for the purpose of the present discussion, we shall not differentiate between pre- and postnominal usage of items such as Sranan *ini*, referring to them as pre/post-nominal elements.

Observe that the pre/post-nominal elements *me* and *ini* in Ewe and Sranan, respectively express the Conformation ‘inside of’ while the nominal they occur with expresses the Ground. As such, these forms are different from the English satellite *in(to)* which occurs with the deictic verb *go* to express movement into a location. What this suggests is that Sranan and English lexicalise different components: while *in(to)* lexicalises a Vector component in English, Sranan *ini* lexicalises a Conformation. Observe further that Sranan and, in like manner, Ewe, do not have a verb that lexicalises motion into an enclosure. This is because deictic verbs do not express boundary crossing. Instead, the move-into-an-enclosure reading is compositionally derived from a combination of the verbs and the phrases containing pre/post-nominal elements. Note that in Ewe, the same goes for the *gé dɛ* expression, which is usually translated as ‘enter’, as well. The allative preposition in combination with the ‘fall’ verb does not encode boundary crossing. Instead, like the deictic verb, *gé dɛ* simply expresses motion towards a location. It is in combination with the Conformation component *me* ‘in(side)’ that the boundary crossing interpretation is derived. The fact that the pre/post-nominal elements express Conformation rather than Vector is seen more clearly when one considers the expression of motion out of a location in the two languages. These are repeated below:

- (14) a. Kofi do le xɔ-a me. (Ewe)  
 K exit at room-DEF in(side)  
 ‘Kofi went out of the room.’
- b. Kofi komoto na ini a kamra. (Sranan)  
 K come\_out LOC in(side) DEF room  
 ‘Kofi came out of the room.’

Unlike the deictic verbs discussed earlier, *do* and *komoto/komopo* express boundary crossing: they refer to motion out of an enclosure. Observe that both languages still have the same pre/post-nominal elements as when expressing motion into a location. This is evidence that these elements express a Conformation and not a Vector. Thus, the constructions in Sranan and Ewe may be characterised as “X moves out of the inside of Y”. The difference between lexicalisation in Ewe and Sranan, on the one hand, and English, on the other, is represented in Figure 3 (note that the Sranan preposition *na*, being semantically empty, is not included here).

Language	Vector		Conformation notion	Fundamental ground
	D-Move	Move TO	Which is at the inside of	room
English	<i>go</i>	<i>in(to)</i>		<i>room</i>
Sranan	<i>go</i>		<i>ini</i>	<i>kamra</i>
Ewe	<i>yi</i>		<i>me</i>	<i>xɔ</i>
	D-Move	Move FROM	Which is at the inside of	room
English	<i>go</i>	<i>out</i>		<i>room</i>
Sranan	<i>komoto/komopo</i>		<i>ini</i>	<i>kamra</i>
Ewe	<i>do</i>		<i>me</i>	<i>xɔ</i>

Figure 3. Lexicalisation of motion in and out

In order to best capture the expression of motion into a location, we decompose Vector into directional movement (D-Move), which does not cross a boundary, and Talmy's (2000) "Move TO" and "Move FROM" which we assume to express boundary crossing. This division enables us to best explain the difference between the satellite *in(to)* in English and the pre/postnominal element *ini* 'in(side)' in Sranan. While the former expresses boundary-crossing Path and Conformation (a "conflation" in the terms of Talmy), the latter only expresses Conformation. Furthermore, our representation enables us to capture the fact that Sranan and Ewe do not encode the Vector TO. As we stated above, this is compositionally derived from the deictic verb and the complement. The expression of moving out is somewhat different: both Sranan and Ewe use a boundary-crossing type of Path verb while English uses a deictic verb to express D-Move and a satellite to express the Vector. This is evidence then that the expression of motion in Sranan is modelled not on that of English but rather on that of Ewe. However, this does not mean that Sranan is exactly like Ewe. As we show in the next section, there are differences in the concepts that are lexicalised by specific words.

## 5. Sranan is not Ewe

Lefebvre (1998: 9) hypothesizes that "the creators of a creole language, adult native speakers of the substratum languages, use the properties of their native lexicon, the parametric values and semantic interpretation rules of their native grammars in creating creoles". This implies that the semantics as well as the syntax of the creole is based on that of the substrate. In this section, we show that although Ewe does play a significant role in the expression of motion in Sranan, relexification does not take place across the board as the quote from Lefebvre implies. This is evidenced by the fact that the components of motion that are conflated in a particular verb

are not necessarily the same across the two languages. This is illustrated here by comparing the equivalents of ‘climb’.

The equivalents of ‘climb’ in Sranan and Ewe appear to be similar because they both conflate Manner and Path. They are therefore of the type that Zlatev and Yangklang (2004) describe as Manner+Path verbs. There is, however, an important difference between the two verbs: while *líá* in Ewe always requires the expression of the fundamental Ground, *kren* in Sranan does not. Consider the examples below:

- (15) a. A kren go na tapu sodro. (Sranan)  
 3SG climb go LOC top attic  
 ‘He climbed upstairs.’  
 b. Mi kren kom na sodro, nou mi moe bákka. (Focke 1855)  
 1SG climb come LOC attic now 1SG must back  
 ‘I climbed upstairs, now I have to go down again.’
- (16) a. Kofi líá atrakpui-a yi xɔ-a tame. (Ewe)  
 K climb ladder-DEF go room-DEF top  
 ‘Kofi climbed up the ladder to the top of the roof.’  
 b. \*Kofi líá yi xɔ-a tame.  
 K climb go room-DEF top  
 ‘Kofi climbed up to the top of the roof.’

Both languages use SVCs to express this kind of translational motion, and in both languages, as with the constructions discussed earlier, the first verb expresses Manner. In this case, however, as we have already stated, the verb also conflates Path. Example (15b) shows that Sranan does not need the element *tapu* ‘top’ to encode that the motion expressed by *kren* is always upward. In Ewe too, *líá* expresses upward motion. The deictic verbs that occur in the second-verb position in both languages therefore simply indicate motion towards or away from deictic centre. The difference between the two verbs lies in the fact that while *kren* in Sranan does not require the expression of the object along which the Figure moves in order to get to the endpoint, this is obligatory in Ewe. Thus, *kren* can be characterised as ‘move up in a climbing manner’, and *líá* in Ewe as ‘move up Y in a climbing manner’.

## 6. Developments under the influence of Dutch

Notwithstanding differences such as those discussed in the previous section, it is clear that Sranan resembles the Gbe languages rather than English in the use of ‘in, inside’ as the Conformation component of a Path expressing ‘out of’, ‘from’, and, more generally, in that the Path itself is expressed by a verb. However, we show in this section that changes are going on in present-day Sranan that seem to lead the

way to a Germanic type of lexicalisation, with the Path being expressed by a Satellite rather than a verb. In order to appreciate this, it is necessary to take into account the broader developments regarding adpositional phrases in Sranan.

## 6.1 Prepositional phrases

From the earliest stages of Sranan onwards, locative complements may consist of a PP containing merely the all-purpose preposition *na* (in modern Sranan also *a*), which itself does not convey any particular meaning.<sup>2</sup> For example, *na foto* may have a goal interpretation ‘to town’ as well as a locational interpretation ‘in town, at the town’, depending on the context, including the verb with which it occurs. As we have seen above, a specifying locative element may co-occur with the preposition *na*, forming a complex PP with the locative item appearing before or after the noun. Example (17) provides an instance from an 18th-century source where *ini* is the Conformation component of the Path, while the verb *komoto* ‘come out of, come from’ lexicalises the Vector. (18) and (19) provide some more 18th-century examples of the pre- and postnominal types of complex PPs respectively. With the prenominal type, *fu* ‘for, of’ occasionally appears between the locative element and the nominal, as illustrated in (18b) (where <vo> = *fu*).

- (17) mi kommotto na inni djari. (Schumann 1783)  
 1SG come\_out LOC in(side) garden  
 ‘I’m coming out of the garden.’
- (18) a. hoe zomma kom na ini hosse (van Dyk ca 1765)  
 Q person come LOC in(side) house  
 ‘who is coming into the house?’  
 b. a trueh watra na inni vo wan tobbo (Schumann 1781)  
 3SG throw water LOC in(side) of INDEF tub  
 ‘he poured water in a tub.’  
 c. da umanpikin liddom na tappo bedi (Schumann 1781)  
 DEF woman-child lie(down) LOC top bed  
 ‘the daughter was lying down on the bed’
- (19) a. a komm na hosso inni (Schumann 1781)  
 3SG come LOC house in(side)  
 ‘he came into a house’

---

2. The origin of *na*, found in several creole languages of various lexical stock, is not yet established with certainty; suggested source forms include Portuguese *na* ‘in the (SG.FEM)’, and Igbo *ná*, a deverbal general locative preposition. It can be assumed to have existed in Sranan from an early stage.

- b. na dem heh bergi bakka (Schumann 1783)  
 LOC DEF.PL high mountain back  
 '[the source of this river is] behind the high mountains'

These examples, from the earliest substantial sources available for Sranan, show that the two types of complex PPs were firmly established by the 18th century already. In particular with respect to the postnominal type of complex PPs, the Gbe languages can be assumed to have provided the model (Bruyn 1996, 2003; Yakpo & Bruyn 2015). The origins of the – more frequently occurring<sup>3</sup> – prenominal type involve the interplay of a variety of factors and will not concern us here; nor will the subtle meaning differences that may obtain between pre- and postnominal complex PPs. For the present purpose, the important point is that until the middle of the 20th century, simple locative prepositions with semantic content did not occur in Sranan. This is not to say that the simple preposition as such did not exist. Apart from all-purpose *na*, there is *fu* 'for, of, from', which occasionally marks locative complements but without adding anything to the meaning already expressed by the verb(s). Furthermore, there are non-locative prepositions such as *nanga* 'with', and *sondro* 'without'. However, locative, and also temporal, complements involved either just *na*, or a complex PP containing a specifying locative element alongside *na*. The locative elements include *baka* 'back(part); behind' (< English *back*), *fesi* 'face; before' (< English *face*), and *tapu* 'top; on' (< English *top*). While some of these elements, viz. *ini* 'in(side)' (< English *in*), *ondro* 'under' (< English *under*), and *abra* 'opposite, across' (< English *over*) are derived from prepositions, the Sranan reflexes appeared only in complex PPs until recently.<sup>4</sup> Most of the English locative prepositions, however, are not retained at all in Sranan; thus, there are no morpho-phonological reflexes of, for example, *to*, *on*, *above*, *behind*, or *against*.

From the middle of the 20th century onwards, the locative items that until then were always accompanied by *na* in complex PPs began to be used as simple prepositions as well. Thus, nowadays we find, alongside complex PPs, PPs such as those in (20), which used to be impossible:

3. Yakpo, van den Berg & Borges (2015: 187) state that postpositional structures were the default option at an earlier stage of Sranan. Prenominal PPs are already more frequent in the earliest available sources however, that is, when a full noun is involved. With pronouns on the other hand, the postnominal type has always been the only option (Bruyn 2003).

4. For *ini* and *ondro*, a Dutch origin, though unlikely, cannot be excluded (cf. Dutch *in*, *onder*; although see Smith (1987: 43) for the derivation of Sranan *abra* from English 'over' rather than from Dutch *over*). This would not affect the present argument, however.

- (20) a. tap' a tafra  
           'on the table'  
       b. baka a oso  
           'behind the house'  
       c. ini a kamra  
           'in the room'

While this change could be interpreted as an internal development, it is highly likely that analogy with Dutch, which has simple locative prepositions and with which Sranan has been in heavy contact, has played a role (cf. Yakpo, van den Berg & Borges 2015). Influence from Dutch also appears to be responsible for changes concerning the expression of motion.

## 6.2 The lexicalisation of the Path

We have just seen that there is a gradual change in Sranan with respect to structural properties, in that the locative elements such as *tapu* 'top, on' and *ini* 'in(side)' may now function as prepositions by themselves. This appears to have paved the way for changes that are going on concerning semantic aspects of the expression of motion. The first thing to notice is that, as a plain preposition, *ini*, or its variants such as *nin* (contracted *na ini* nowadays functioning as a variant of *ini*), appears to be able to express Path in modern Sranan, in contrast to both its earlier Sranan and its Ewe counterparts. In (21a), where it is clear from the context that someone jumps into the boat, there is no deictic verb indicating that translational motion is involved. In earlier stages of the language, the interpretation of a sentence that similarly involves the verb 'jump' and the locative item *ini* but not a deictic verb such as *go*, would be one of self-contained motion, as in (21b), where the jumping is inside the belly. Note that (21b) resembles (22) from Ewe, where the verb *dzo* 'jump' does not include a Path; without a deictic verb a sentence such as (22) can only refer to self-contained motion:

- (21) a. a dyompo nin a boto (de Drie 1985)  
           3SG jump in(to) DEF boat  
           'he jumped into the boat'  
       b. da pikin djompo na inni va hem belle (Schumann 1781)  
           DEF child jump LOC in(side) of 3SG belly  
           'the child jumped inside her belly'
- (22) dɛví-á dzo le tɔdziúú-a me (Ewe)  
       child-DEF jump LOC boat-DEF in(side)  
       'the child jumped inside the boat'



The innovative use of *(n)in(i)* in cases such as (21a) clearly reduces the correspondence between Sranan and its original substrate. For some Sranan speakers, *(n)in(i)* cannot even appear anymore in sentences expressing motion out of a location. Rather than *na* plus *ini* as in (23a) (cf. (8b) and (17) above), these speakers use the simple preposition *fu* ‘for, of, from’ (23b), or *uit* ‘out (of)’, a recent Dutch borrowing (23c) (cf. Sebba 1987: 47):

- (23) a. a waka komoto na ini a oso  
 3SG walk come\_out LOC in(side) DEF house  
 ‘he walked out of the house’  
 b. a waka komoto fu a oso  
 3SG walk come\_out P DEF house  
 ‘he walked out of the house’  
 c. a komoto uit a batra go koyri  
 3SG come\_out out DEF bottle go “walk”  
 ‘it came out of the bottle and went away’

The fact that for some speakers *ini* cannot co-occur with a verb such as *komoto* that expresses a Move-FROM Path means that they take *ini* to express Move-TO. For these speakers the (new) meaning of *ini* clashes with that of *komoto*. While this deviates from the original Sranan pattern, it renders Sranan more similar to Dutch, which, like English, is satellite-framed. Clearly, this change in the lexicalisation pattern results from influence from Dutch.

The borrowing of *uit*, which has the specific meaning ‘out (of)’ demonstrates even more clearly the tendency for some Sranan speakers to have a satellite take part in the expression of the Vector. There is no morpheme in Sranan that is a reflex of English *out*. Thus, in the absence of an appropriate Sranan form, Dutch *uit* ‘out (of)’ is borrowed to express the Vector in a way that is similar to Dutch. This borrowing supports the assumption that Dutch influence is at issue.

The use of *uit* in cases such as (23c) leads, however, to a mixed pattern, to the extent that the Vector is now lexicalised by the satellite *uit* as well as by the verb *komoto*. The next step would be the use of *uit* without a Path-conflating verb being present, for example, *a waka uit a oso*, where the Path is not lexicalised by *waka* ‘walk, go’ but only by *uit*. While we did not attest such cases involving *uit*, (21a) above provides an example involving the verb *dyompo* ‘jump’, with the Path being lexicalised by *ini* ‘in(to)’.

Thus, several changes in Sranan can be attributed to influence from Dutch. First, the development whereby it became possible to use locative items such as *ini* ‘in’ and *baka* ‘behind’ as simple prepositions can be assumed to have proceeded under influence of Dutch. Second, at least for some younger speakers, the preposition *ini* can no longer express the Conformation component of a Path expressing

‘out of the inside of’. Instead, *fu* ‘of, for, from’ is used, or, and this is the third point of Dutch influence, the Dutch borrowing *uit* ‘out’. This borrowing supports the idea that some younger speakers tend to attribute to the preposition the role of lexicalising the Vector component of the Path rather than the Conformation component. The incorporation of Dutch *uit* as a preposition in Sranan is facilitated by the earlier development that led to the usage of simple locative prepositions. It should be stressed, however, that the older pattern still exists, and that those younger speakers that use *uit* rather than *ini* do so in combination with a Path-conflating verb.

It might be argued that the changes we have discussed are due to internal developments in Sranan. However, there is reason to believe that they were induced by contact with Dutch. From the middle of the 20th century, there have been a number of changes in morpho-syntactic areas that had been stable for 200 years. These include the replacement of *di(si)* with the question words *san* ‘what’ and *suma* ‘who’ as heads of the relative clause, and the obsolescence of resumptive pronouns with relativized PPs (Bruyn 1995). The rather rapid and chaotic way in which the latter developments have proceeded indicates that they were induced by contact with Dutch. A sociolinguistic factor that may account for the fact that Dutch influence on Sranan has taken effect more strongly since the mid-20th century rather than earlier is related to the emancipation of Sranan. From around the time that Suriname became an autonomous state within the Dutch kingdom in 1954, the functional distribution of Sranan and Dutch became less diglossic, with Sranan becoming more widely used by different ethnic groups than used to be the case, and the number of Sranan speakers speaking Dutch as a native language increasing significantly (Gobardhan-Rambocus 1995; de Kleine 1999). The extent of bilingualism, not only in Suriname but also among the Surinamese living in the Netherlands, certainly provided a context for Dutch to influence Sranan (cf. Thomason 2001). We argue that such influence extended to the lexicalisation of motion out of a location. Our position is supported by the fact that an actual Dutch form, namely *uit*, figures in the constructions of some younger speakers.

## 7. Conclusion

We hope to have shown that multiple effects of contact are at issue with regard to the expression of motion into and out of locations in Sranan.

Sranan, being a serialising language, employs typologically equivalent verbs to express Path and Manner of Motion. It is thus of the type which Slobin (2004) calls equipollently-framed. While Sranan clearly contrasts with its lexifier language, English, which is satellite-framed in this respect, it resembles Ewe and other Gbe languages that constituted the substrate during its developmental stages. The

similarity between Sranan and Ewe – and by extension, Gbe – is even more evident in the lexicalisation patterns of the Path and Ground components of translational motion involving manner. While both Sranan and Ewe have a verb expressing motion out of an enclosure, neither has a verb that, by itself, lexicalises motion into an enclosure. Furthermore, in both languages the expression of the notions ‘enter’ and ‘exit’ may involve an element ‘in(side)’ expressing the Conformation component of the Path, that is, it relates the Ground to the Path. This leads us to conclude that the way motion in and out of a location is lexicalised in Sranan is modelled on the Gbe substrate languages. However, as we have shown, Sranan is not exactly like Ewe in every respect. On the one hand, there are some differences in the use of prepositions, and in the position of items such as ‘in(side)’ vis-à-vis the nominal expressing the source or goal of the motion event. On the other hand, not all motion verbs have necessarily the same lexicalisation patterns in the two languages, as illustrated by the verb expressing ‘climb’. Thus, there has not been relexification of all individual lexical items that take part in the expression of movement. According to the relexification hypothesis (Lefebvre 1998; Lumsden 1999), word forms derived from the lexifier are incorporated into lexical entries of the substrate, with the morphosyntactic and semantic properties of the latter being preserved. Our comparison of the expression of motion in Sranan, English, and Ewe, suggests that, while the properties of some but not all lexical items in Sranan can be accounted for by relexification of particular Ewe forms, it is ultimately the strategy typically used to express motion in Sranan which is modelled on Ewe and other Gbe languages. Relexification here should therefore be taken to involve pattern replication rather than substance or matter copying (Aboh 2006, 2015a, 2015b; Matras & Sakel 2007; Bruyn 2009; Yakpo & Bruyn 2015).

While the typological make-up of Sranan regarding the expression of movement is similar to that of the Gbe languages, we have also shown that recent developments lead to a deviation from the original situation. Nowadays, the Vector component of the Path may, for some speakers, be lexicalised by a satellite. This is most clearly evidenced by the borrowing of the Dutch preposition *uit* ‘out’. This development must be viewed in relation to a change in Sranan whereby simple locative prepositions have become possible. The latter development, as well as the tendency for some speakers to have the satellite express the Vector rather than the Conformation component of the Path, can be attributed to influence from Dutch. The implication of the recent developments in the area of the expression of motion is that Sranan appears to be moving away from its original state, in which it was typologically closer to the Gbe languages, to become more like Dutch, and, for that matter, English, in the lexicalisation of the Path of motion.

## Acknowledgements

The paper reports on research supported by the Netherlands Organization for Scientific Research (NWO), grant nr 360-7-020, and Pieter Muysken's Spinoza Program Lexicon and Syntax. It was originally presented at the International Conference on Adpositions of Movement, Catholic University of Leuven, 14–16 January 2002. We thank two anonymous reviewers for their helpful comments on an earlier version of the paper, for a volume which, unfortunately, never appeared. We are further indebted to two anonymous reviewers who provided comments on the current version.

## Abbreviations

The following abbreviations are used in the glosses:

1/2/3	1st/2nd/3rd person	ALL	allative preposition
SG	singular	Q	question particle
PL	plural	DEF	definite determiner
P	preposition	INDEF	indefinite determiner (singular)
LOC	locative preposition		

## References

- Aboh, E. O. 2006. The role of the syntax-semantics interface in language transfer. In *L2 Acquisition and Creole Genesis: Dialogues* [Language Acquisition and Language Disorders 42], C. Lefebvre, L. White & C. Jourdan (eds), 221–252. Amsterdam: Benjamins.  
<https://doi.org/10.1075/lald.42.14abo>
- Aboh, E. O. 2015a. *The Emergence of Hybrid Grammars: Contact, Change and Creation*. Cambridge: CUP.
- Aboh, E. O. 2015b. The left periphery in the Surinamese creoles and Gbe: On the modularity of substrate transfer. In *Surviving the Middle Passage*, P. C. Muysken & N. S. H. Smith (eds), 323–368. Berlin: Mouton de Gruyter.
- Aboh, E., Ameka, F. & Essegbey, J. 2002. Moving from verbs to prepositions. Paper presented at the International Conference on Adpositions, Leuven, January 2002.
- Ameka, F. 2003. Prepositions and postpositions in Ewe(Gbe): Empirical and theoretical considerations. In *Typologie des langues d'Afrique et universaux de la grammaire, Vol. 2: Benue-kwa, Wolof*, P. Sauzet & A. Zribi-Hertz (eds), 41–67. Paris: l'Harmattan.
- Ameka, F. & Essegbey, J. 2013. Serialising languages: Satellite-framed, verb-framed or neither. *Ghana Journal of Linguistics* 2: 19–38.
- Arends, J. 1995. Demographic factors in the formation of Sranan. In *The Early Stages of Creolization* [Creole Language Library 13], J. Arends (ed.), 233–285. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/cll.13.11are>
- Borges, R. 2013. *The Life of Language: Dynamics of Language Contact in Suriname*. Utrecht: LOT.

- Bruyn, A. 1995. *Grammaticalization in Creoles: The Development of Determiners and Relative Clauses in Sranan*. Amsterdam: IFOTT.
- Bruyn, A. 1996. On identifying instances of grammaticalization in creole languages. In *Changing Meanings, Changing Functions: Papers Related to Grammaticalization in Contact Languages*, P. Baker & A. Sycia (eds), 29–46. London: University of Westminster Press.
- Bruyn, A. 2003. Grammaticalisation, réanalyse et influence substratique: Quelques cas de Sranan. In *Grammaticalisation et réanalyse: Approches de la variation créole et française*, S. Kriegel (ed.), 25–47. Paris: CNRS Éditions.
- Bruyn, A. 2009. Grammaticalization in creoles: Ordinary and not-so-ordinary cases. *Studies in Language* 33: 312–337. <https://doi.org/10.1075/sl.33.2.04bru>
- de Drie, A. 1985. *Sye! Arki tori!*, compiled by Trudi Guda. [Paramaribo:] Afdeling Cultuur Studies van het Ministerie van Onderwijs, Wetenschappen en Cultuur.
- de Kleine, C. M. 1999. *A Morphosyntactic Analysis of Surinamese Dutch as Spoken by the Creole Population of Paramaribo, Suriname*. Ann Arbor MI: University Microfilms International.
- Doncic, A. 1954. *De creoletaal van Suriname. Spraakkunst*. Paramaribo: Radhakishun.
- Essegbey, J. 2001. Motion expression in Sranan, evidence for substrate influence? Paper presented at the Meeting of the Society for Pidgin and Creole Linguistics, Coimbra.
- Essegbey, J. 2005. The “basic locative construction” in Gbe languages and Surinamese creoles. *Journal of Pidgin and Creole Languages* 20: 229–267. <https://doi.org/10.1075/jpcl.20.2.02ess>
- Focke, H. C. 1855. *Neger-Engelsch woordenboek*. Leiden: P.H. van den Heuvel.
- Gobardhan-Rambocus, L. 1995. Het Sranan: Ontwikkeling en emancipatie. In *De erfenis van de slavernij*, L. Gobardhan-Rambocus, M. S. Hasankhan & J. L. Egger (eds), 147–71. Paramaribo: Anton de Kom Universiteit.
- Huttar, G. L. 1975. Sources of creole semantic structures. *Language* 51: 684–695.
- Huttar, G. L. 1981. Some Kwa-like features of Djuka syntax. *Studies in African Linguistics* 12: 291–323.
- Lambert-Brétière, R. 2009. Serializing languages as satellite-framed: the case of Fon. *Annual Review of Cognitive Linguistics* 7: 1–29.
- Lefebvre, C. 1998. *Creole Genesis and the Acquisition of Grammar*. Cambridge: CUP.
- Lefebvre, C. & Brousseau, A.-M. 2002. *A Grammar of Fongbe*. Berlin: Mouton de Gruyter.
- Lumsden, J. S. 1999. The role of relexification in creole genesis. *Journal of Pidgin and Creole Languages* 14: 225–258. <https://doi.org/10.1075/jpcl.14.2.01lum>
- Matras, Y. & Sakel, J. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31: 829–865. <https://doi.org/10.1075/sl.31.4.05mat>
- Mayer, M. 1969. *Frog, Where Are You?* New York NY: Dial Press.
- Migge, B. 1998. Substrate influence in creole formation: The origin of *give*-type serial verb constructions in the Surinamese Plantation Creole. *Journal of Pidgin and Creole Languages* 13: 215–265. <https://doi.org/10.1075/jpcl.13.2.02mig>
- Migge, B. 2000. The origin of the syntax and semantics of property items in the Surinamese Plantation Creole. In *Language Change and Language Contact in Pidgins and Creoles* [Creole Language Library, 21] J. McWhorter (ed), 201–234. Amsterdam: Benjamins. <https://doi.org/10.1075/cll.21.08mig>
- Plag, I. 1998. The syntax of some locative expressions in Sranan. Prepositions, postpositions or noun? *Journal of Pidgin and Creole Languages* 13: 335–353. <https://doi.org/10.1075/jpcl.13.2.05pla>
- Schaefer, R. P. & Gaines, R. 1997. Toward a typology of directional motion for African languages. *Studies in African Linguistics* 26: 193–220.

- Schumann, C. L. 1781. *Die Geschichte unsers Herrn und Heilandes Jesu Christi, aus den vier Evangelisten zusammengezogen*, durch Samuel Lieberkühn; in Neger-Englische Sprache übersezt, zum Gebrauch bey der Neger-Gemeine. MS Paramaribo [Moravian Archives Utrecht 12/617].
- Schumann, C. L. 1783. *Neger-Englisches Wörter-Buch*. Editio tertia. MS Paramaribo [Moravian Archives Paramaribo 12/648].
- Sebba, M. 1987. *The Syntax of Serial Verbs. An Investigation into Serialisation in Sranan and Other Languages* [Creole Language Library 2]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.2>
- Slobin, D. I. 1997. Mind, code, and text. In *Essays on Language Function and Language Type: Dedicated to T. Givón*, J. L. Bybee, J. Haiman & S. A. Thompson (eds), 437–467. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.82>
- Slobin, D. I. 2000. Verbalized events: A dynamic approach to linguistic relativity and determinism. In *Evidence for Linguistic Relativity* [Current Issues in Linguistic Theory 198], S. Niemeier & R. Dirven (eds), 107–138. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.198.10slo>
- Slobin, D. I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In *Relating Events in Narrative, Vol. 2: Typological and Contextual Perspectives*, S. Strömquist & L. Verhoeven (eds), 219–257. Mahwah NJ: Lawrence Erlbaum Associates.
- Slobin, D. I. 2006. What makes manner of motion salient? Explorations in linguistic typology, discourse and cognition. In *Space in Language: Linguistic Systems and Cognitive Categories* [Typological Studies in Language 66], M. Hickmann & S. Robert (eds), 60–81. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.66.05slo>
- Slobin, D. I. & Hoiting, N. 1994. Reference to movement in spoken and signed languages: Typological considerations. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society, General Session*, K. E. Moore, D. A. Peterson & C. Wentum (eds), 437–505. Berkeley CA: BLS.
- Smith, N. S. H. 1987. The Genesis of the Creole Languages of Surinam. PhD dissertation, University of Amsterdam.
- Smith, N. S. H. 1997. Ingredient X: The common core of African words in the Atlantic creoles. Paper presented at the meeting of Society for Pidgin and Creole Linguistics, London.
- Smith, N. S. H. 2001. Voodoo Chile: Differential substrate effects in Saramaccan and Haitian. In *Creolization and Contact* [Creole Language Library 23], N. Smith & T. Veenstra (eds), 43–80. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.23.03smi>
- Sordam, M. & Eersel, H. 1989. *Surinaams woordenboek/Sranantongo*, 2nd edn. Baarn: Bosch & Keuning.
- Talmy, L. 1985. Lexicalization patterns. In *Language Typology and Syntactic Description, Vol 3: Grammatical Categories and the Lexicon*, T. Shopen (ed.), 57–149. Cambridge: CUP.
- Talmy, L. 2000. *Toward a Cognitive Semantics, Vol 2: Typology and Process in Concept Structuring*. Cambridge MA: The MIT Press.
- Talmy, L. 2007. Lexical typologies. In *Language Typology and Syntactic Description, Vol. 3: Grammatical Categories and the Lexicon*, 2nd edn, T. Shopen (ed), 66–168. Cambridge: CUP.
- Talmy, L. 2009. Main verb properties and equipollent framing. In *Crosslinguistic Approaches to the Psychology of Language. Research in the Tradition of Dan Isaac Slobin*, J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & Ş. Özçalışkan (eds), 389–402. New York NY: Taylor & Francis.
- Thomason, S. G. 2001. *Language Contact*. Edinburgh: EUP.

- van Dyk, P. [ca 1765]. *Nieuwe en nooit bevoorens geziene onderwyzinge in het Bastert Engels, of Neeger Engels, zoo als het zelve in de Hollandsze Colonien gebruikt word*. Amsterdam: Erven de Weduwe Jacobus van Egmont.
- van Putten, S. 2017. Motion in serializing languages revisited: The case of Avatime. *STUF* 70: 65–83.
- Wilner, J. (ed.). 1994. *Wortubuku fu Sranan Tongo*, 3rd trial ed. Paramaribo: Summer Institute of Linguistics.
- Yakpo, K. & Bruyn, A. 2015. Trans-Atlantic patterns: The relexification of locative constructions in Sranan. In *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*, P. C. Muysken & N. S. H. Smith (eds), 135–174. Berlin: Mouton de Gruyter.
- Yakpo, K., van den Berg, M. C. & Borges, R. 2015. On the linguistic consequences of language contact in Suriname: The case of convergence. In *In and out of Suriname*, E. B. Carlin, I. Leglise, B. Migge & P. B. Tjon Sie Fat (eds), 164–195. Leiden: Brill.
- Zlatev, J. & Yangklang, P. 2004. A third way to travel: The place of Thai in motion event typology. In *Relating Events in Narrative, Vol. 2: Typological and Contextual Perspectives*, S. Strömquist & L. Verhoeven (eds), 159–190. Mahwah NJ: Lawrence Erlbaum Associates.

# Sociolinguistic characteristics of the English-lexifier contact languages of West Africa

Kofi Yakpo

The University of Hong Kong

This chapter provides a comparison of key sociolinguistic characteristics of Nigerian Pidgin, Cameroon Pidgin, Ghanaian Pidgin English, Pichi (Equatorial Guinea) and Krio (Sierra Leone). In the past few decades, these African English-lexifier contact languages (AECs) have seen an exponential growth in speaker numbers and an expansion into domains once reserved for English and non-creole African languages. All AECs nevertheless still struggle with a low sociolinguistic prestige and the absence of corpus and status planning initiatives by state actors. Overall, the potential of these languages remains relatively untapped across the region for education, political participation, economic, and cultural activity. At the same time, the impact of the AECs on smaller languages through contact and shift to the AECs is also likely to make itself felt in coming decades.

**Keywords:** English-lexifier creoles, language contact, language policies, language ideologies, West Africa, sociolinguistic domains

## 1. Introduction

If Pieter Muysken's early research in creole and contact linguistics was dominated by formal approaches (e.g. Muysken 1981a, 1981b; Muysken & Smith 1986), his work in the last two decades or so has been thoroughly interdisciplinary, integrating language typology, psycholinguistics, historical linguistics and sociolinguistics (e.g. in the "contact scenario" framework; cf. Muysken 2008). Through his work, he has also supported language vitalization, awareness-building, and information dissemination activities on the Indigenous languages of the South American Andes region (e.g. through work on the four-volume Crevels & Muysken 2009). With his undogmatic and holistic approach, Pieter has always been a step ahead in foreseeing



the necessity to integrate creole linguistics into linguistic typology (e.g. Yakpo & Muysken 2017), areal linguistics (Muysken 2008; Muysken & Smith 2015), historical linguistics (e.g. van Sluijs, van den Berg & Muysken 2016) and sociolinguistics. Current societal trends in the regions of Africa in which English-lexifier contact languages are underline show the necessity for a broader perspective. A tectonic demographic shift is underway on the African continent, entailing a corresponding growth in speaker numbers of the English-lexifier contact languages of West Africa, and foreseeable changes in their social functions and uses.

The population of Nigeria, where Nigerian Pidgin, the largest English-lexifier contact language on the globe is spoken, is projected to reach about 750 million in the next eighty years (United Nations 2017: 21). In view of such demographics, it is possible that the combined speaker numbers of English-lexifier contact languages (AECs) in Africa and the Caribbean will reach half a billion by 2100, making AECs the most widely spoken African language continuum, and one of the largest language clusters of the world in speaker numbers by the turn of the 22nd century. The implications of the status and uses of these languages for education, political participation, economic development, cultural expression and linguistic diversity, language contact and change, are therefore highly significant.

In this chapter, I attempt to contribute to a better understanding of these societal dynamics. I do so by providing a first comparative view of the sociolinguistic situation of the five major African AECs: Cameroon Pidgin, Nigerian Pidgin, Pichi (Equatorial Guinea), Krio (Sierra Leone), and Ghanaian Pidgin English (Ghana). I characterise the respective speaker communities, identify domains of use and address critical aspects of language policies and language ideologies.

In spite of their substantial expansion as first and second languages, the African AECs have received little attention from state institutions. The uses of these languages differ across the five countries. Their status, however, shows strong similarities. What unites all five countries is the general reluctance to furnish organised state responses to linguistic diversity on the one hand, and the presence of largely “exoglossic” (Heine 1992) language policies on the other. Such policies exclusively rely on a colonial language in official functions, privileging the use of English and Spanish (the latter for Equatorial Guinea) in formal domains including administration and the educational system. On a whole, policy makers have not responded to the possibilities that the emergence of the national AEC varieties as important lingua francas have raised for access to formal education and literacy, cultural vibrancy, nation-building, political participation and economic growth, and the economic and political integration of West Africa.

The structure of this chapter is as follows: In Section 2, I provide an overview of the English-lexifier contact languages spoken West Africa. In Section 3, I provide

a comparative perspective by looking at the speaker communities, language uses and domains, and language policies vis-à-vis Nigerian Pidgin, Cameroon Pidgin, Krio (Sierra Leone) and Ghanaian Pidgin English. Section 4 concludes this article. Where not indicated otherwise, the observations and data stem from many years of research throughout West Africa. This data consists, among other genres, of many hours of about fifty thousand words of recorded conversations on a diverse range of topics including views on grammar, language attitudes and uses.

## 2. The English-lexifier contact languages of West Africa

In the following, I use the term “(Afro-Caribbean) English-lexifier contact languages” (henceforth AECs) to designate the chain of English-lexifier contact languages spoken in the Americas and in Africa (also referred to as Afro-Caribbean English-lexifier Creoles, cf. Faraclas 2004; Yakpo 2012a; and Atlantic English-lexifier Creoles, cf. Arends, Muysken & Smith 1995). The evidence that these languages owe similarities between each other to a common ancestry in one or several proto-languages spoken in the Eastern Caribbean in the early 17th century, and/or along the West African coast is compelling (Hancock 1987; Smith 2015), even if the exact historical details are yet unclear. The evidence for the relatedness of the African AECs is even more convincing (Huber & Görlach 1996; Huber 1999, 2000). The African AECs are characterised by a large number of lexical, structural, and phonological similarities (cf. Yakpo 2012b, 2013b, 2019).

AECs serve as the established community languages and L1s to the majority populations of various Caribbean nations, among them Jamaica, Guyana, Suriname, and Trinidad & Tobago. The African AECs, however, have proportionally much larger L2 (or better, LX for “additional language”) than L1 communities. Differences in the proportions of L1 and L2 speakers between West Africa and the Caribbean have prompted some sources to give the label “pidgincreole” to West African AECs like Nigerian Pidgin and Ghanaian Pidgin English (e.g. Bakker 2008; Corum 2015; Fonka 2016). I, however, dispense with the term “pidgincreole” in this chapter. There has not yet been a systematic attempt to show that the predominant use of the African AECs as L2s has any structural-lexical ramifications. What we instead find are very stable grammatical systems that do not show out-of-the ordinary features distinguishing them from other languages spoken alongside them. I therefore settle for “(Afro-Caribbean) English-lexifier contact languages” (abbreviated AECs) in this chapter.

The following table provides estimated speaker numbers of the AEC varieties of West Africa and the Caribbean (cf. Yakpo 2016a):

**Table 1.** Speaker numbers of Afro-Caribbean English-lexifier contact languages

Language	Country spoken	Speaker estimate	Sources
Nigerian Pidgin	Nigeria	75–80 million	Ihemere 2006
Cameroon Pidgin	Cameroon	10 million	Lewis, Simons & Fennig 2013
Krio	Sierra Leone	6 million	Finney 2011
Aku	Gambia	14,000	Juffermans & McGlynn 2010
Ghanaian Pidgin English	Ghana	5 million	Huber 2013
Pichi	Equatorial Guinea	100–150,000	Yakpo 2013b
Caribbean (combined)	Insular & continental Caribbean	6 million	Conservative estimate based on population sizes
<b>Total speaker no.</b>		<b>about 110 million</b>	

Based on estimated speaker numbers, the different varieties of AECs together count around 110 million speakers in Africa and the Americas. The AECs form the largest linguistic continuum of West Africa, before the two largest indigenous international languages of West Africa, the Maninkakan-Bamanankan-Jula (Manding) cluster (at least 20 million speakers, cf. Vydrin 2013: 76) and Hausa (presently significantly more speakers than the figure of 40 million given in (Newman 1987: 621)). It is also almost certain that the speaker numbers of the African AECs surpass those of the colonial official languages of the region, English, French, Portuguese, and Spanish put together. For example, Adegbiya (2004: 195) gives the figure of twenty percent of speakers of (Nigerian) Standard English over the total population of Nigeria, without, however, providing further details about the nature of “Standard” English in the country, nor about the relationship between competence in English and in Nigerian Pidgin.

In my experience of observing and participating in interactions between AEC speakers from different West African countries, mutual intelligibility is very high. There is, however, a split between Krio and its offshoots Pichi and Aku on the one hand, and Ghanaian, Nigerian and Cameroon Pidgin on the other, which can be argued for on phonological and structural grounds (Yakpo 2013a, for a comparison of selected features of the West African AECs). Mutual intelligibility is lowest between Pichi (Equatorial Guinea) and all other West African AECs. This is a consequence of the unique trajectory of Pichi: It has been isolated from its lexicon-providing language English for two centuries and therefore escaped lexical and structural convergence with English (cf. Yakpo 2017). It has also developed phonological idiosyncrasies due to adstratal influence from the Bantu language Bubi (Yakpo 2013b: 287–90), and has been influenced extensively by Equatorial

Guinea's colonial and official language Spanish (Yakpo 2009b, 2018). I show in the next section that the linguistic ecologies the five AECs are spoken in are characterised by both similarities and differences.

### 3. The sociolinguistic situation of the West African AECs

I will now provide an overview of the status quo of the five largest AECs of West Africa, namely Nigerian Pidgin (Nigeria), Cameroon Pidgin (Cameroon), Krio (Sierra Leone), Ghanaian Pidgin (Ghana) and Pichi (Equatorial Guinea). Aku, an offshoot of Krio spoken in Banjul, the capital of The Gambia, has been excluded because there is very little sociolinguistic and linguistic data available. I look at the nature of the speaker communities, identify domains of use, and describe the status of these languages as reflected in language policies in these five countries. The situation of these languages is similar in a number of respects. One of them is their impressive expansion and their conquest of domains once reserved for English and non-creole African languages, such as, for example, media and broadcasting, religious practice, and basic literacy teaching. At the same time, all five languages still struggle with similar prestige problems, and ideas about the inferiority of these languages (to both European and other African languages) are still being reproduced and passed down in variations and gradations by post-colonial intellectual elites (for the situation in the Caribbean, cf. Devonish 1986, 2010).

#### 3.1 Nigerian Pidgin

Nigeria is of pivotal importance when assessing the significance and status of the West African AECs. With an estimated 70–80 million speakers, Nigerian Pidgin (also referred to by many Nigerians as “Naija”) is the largest European-lexifier contact language on the globe (Faraclas 2013: 176). It is probably the only Nigerian language that most Nigerians can communicate in to some degree. In the West African sub-region, Nigeria has been at the forefront of status and corpus planning for its autochthonous languages, introducing them into school and university curricula as well as officially declaring Hausa, Yoruba, Igbo as languages of parliamentary debate in the 2nd Republic Constitution of 1979 (cf. Adegbija 2004). Today, the country has, by comparison with other African nations a relatively vibrant media, music, film and print output in its African languages. How is Nigerian Pidgin faring in this context?

Nigerian Pidgin is probably the most widely used language in the media besides English. Knowledge about the language has been carried far beyond Nigeria's

borders, through the country's burgeoning pop music industry and its presence and use, often in alternation with English and other languages of Nigeria, in radio programming, and in "Nollywood" films, produced by the country's prolific and internationally successful film industry.<sup>1</sup> Nigerian Pidgin has also been used in literature in various forms, ranging from dialogues in novels and cartoons, to poetry and plays (e.g. in a recent Nigerian Pidgin adaptation of *Romeo and Juliet* by Bernard Ogini, see Gift 2019). In 2017, BBC Africa began a news service featuring print and audio-visual content in (Nigerian) Pidgin, largely produced by Nigerian journalists, (see <<https://www.bbc.com/pidgin>>. The spelling used is etymological, and English-oriented for words not derived from English, thus reiterating long-standing practice in the region by literate Pidgin speakers in spontaneous graphicisation found in print media, books, letters, and more recently on social media.

Nigerian Pidgin has also been enjoying increasing recognition by intellectuals, both for its immense practicality as a means of communication in Nigeria, and for its unifying potential in being less loaded with ethnic affiliation than other languages in a nation, in which ethnicity is highly politicised (cf. e.g. Momoh & Adejumbi 2002 for an overview). The foundation of the *Naija Langwej Akademi* (Academy for Nigerian Pidgin) in 2010 ushered in academic involvement in corpus and status planning initiatives independently of state support, including the promotion of the glossonym "Naija". A practical standard orthography was also recently developed by the Akademi, which has the potential to serve as a regional standard for the graphicisation of other West African AECs, see <<https://www.ifra-nigeria.org/naija-corner/naija-langwej-akademi>>. Unfortunately, neither this orthography nor previous ones (e.g. Faraclas 1986) have been acknowledged and accepted outside academia, nor even by the media, as the BBC experience mentioned above shows.

Nigerian Pidgin is an important business language in Nigeria's "informal" sector, including individual, small and medium-sized businesses both within Nigeria's borders and in trans-border commerce with countries in the region including Cameroon, Benin, Togo, and Ghana. Equally, Nigerian Pidgin has long been used as a language of interaction and command in military and police barracks, which have been focal points in the spread of the language into Nigeria's urban areas.

---

1. The use of Nigerian Pidgin in pop music increased exponentially after Nigerian independence in 1960. People in other parts of Africa and the world were familiarised with it through international hits like *sweet mother* by Prince Nico Mbarga (1976) <<http://youtu.be/3mecN-rIaWOA>>, and the complete works of Fela Kuti, e.g. *colonial mentality* (1977) <<http://youtu.be/WKm7tWInfnU>>. Today, songs sung in Nigerian Pidgin dominate Nigeria's and West Africa's pop music scene and are increasingly penetrating the global pop music market, see e.g. Tekno's 2016 world hit *Pana* <<https://youtu.be/8YhAFBOSk1U>>.

Last but not least, Nigerian Pidgin has a growing L1 community, particularly in the highly multilingual Delta region in the south of the country (Adebija 2004: 226).

Nevertheless, there has so far been no state recognition of Nigerian Pidgin, nor any other serious attempts at status and corpus planning (Agheysi 1988; Igboanusi 2008). As a consequence, Nigerian Pidgin is the only large Nigerian language to yet have no officially recognised orthography (Ojarikre 2013), nor regulated presence in the educational system or other institutional contexts (Gani-Ikilama 1990). That said, Nigerian Pidgin has served as a language in grass-roots literacy campaigns (Faraclas 1986), and is used in proselytization activities by Christian groups (for a translation of the Christian New Testament into Nigerian Pidgin, cf. <<http://nigerianpidgin-bible.yolasite.com>>. Nigerian Pidgin is also used by Nigerians in interactions with the administration and is hugely popular as a youth sociolect and language of identity for Nigerians of all class backgrounds and age groups. Consequently, a closer look reveals a broad range of attitudes towards Nigerian Pidgin, reflecting the kind of ambivalence to the language found towards other AECs; an ambivalence that reflects the tension between practicality and the identitarian value of Nigerian Pidgin on the one hand, and the reproduction of the cultural and political legacy of colonial rule on the other, with its privileging of English as the *ne plus ultra* of upward social mobility.

### 3.2 Cameroon Pidgin

The sociolinguistic situation of Cameroonian Pidgin (also referred to as “Kamtok” by some academics and language activists, cf. Ngefacs 2010: 152, for the rationale in choosing this designation) is very similar to that of Nigerian Pidgin. Cameroon Pidgin is probably the most widely spoken language of Cameroon, although national figures are difficult to come by due to the absence of comprehensive linguistic surveys. Extrapolations of speaker numbers range between a quarter to half of the Cameroonian population, hence up to 10 million (cf. e.g. the entry “Pidgin, Cameroon” in Lewis, Simons & Fennig 2013). Cameroon Pidgin is the first and dominant language of a growing proportion of inhabitants of the so-called “Anglophone” or “English-speaking” provinces of western Cameroon (Schröder 2003: 82ff.). The spread of Cameroon Pidgin into the “francophone” or “French-speaking” provinces of Cameroon has prompted Cameroonian scholars and lay people to emphasise the value of the language in bridging the political divisions between “Anglophone” and “francophone” Cameroon (e.g. Neba, Chibaka & Atindogbé 2006).

The expansion of Cameroon Pidgin has followed similar pathways to Nigeria. The domains of use of Cameroon Pidgin have been continually expanding since independence, from market-place communication, via use in popular music and

oral genres like Cameroonian pop music, comedy, radio programming, advertising (Schröder 2003: 123ff.; Anchimbe 2013: 172–73), to its appropriation as a language of youth identity and counterculture on school and university campuses, despite constant attempts to ban its use in these institutions (Alobwede D’Epie 1998). Factors such as these have led to Cameroon Pidgin increasingly penetrating the traditionally “francophone” cities of central and south-eastern Cameroon beyond Douala, one of its traditional stronghold in the “French-speaking” part of the country. According to Mbangwana (1983) the number of four-year olds using Cameroon Pidgin in 1983 as an L1 ranged between 20–30% in the major urban centres of “anglophone” Cameroon (Bamenda, 22%, Buea, 26%, Limbe, 31%). In a more recent survey by Schröder (2003: 83ff.) these percentages have risen considerably (Bamenda, 36%, Buea, 42%), and show surprisingly high double digits of Cameroon Pidgin L1 speakers in Cameroon’s capital Yaoundé (30%) located in the “francophone” heartlands of the country.

Cameroon Pidgin has conquered a broad range of additional domains in the last few decades, halting only before the most formal, institutionalised ones: written literary production, administrative and legal writing, in-class interaction, and big business interaction in the formal sector of the economy (Schröder 2003: 181; Anchimbe 2013: 176ff). This notwithstanding, an institutional recognition of Cameroon Pidgin as a language in its own right and a suitable subject of language policy is not on the horizon. There is a relatively large body of descriptive work on Cameroon Pidgin (among which major works like Kouega 2008; Nkengasong 2016; Ngefac 2016; Ayafor & Green 2017), as well as Christian missionary initiatives, e.g. the Cameroon Pidgin New Testament (Molindo 1996). As is the case with the other AECs of West Africa, substantial sections of the traditional intellectual and political elites of the country see Cameroon Pidgin as a deficient form of English. A recurring theme in the discourses on Cameroon Pidgin is its supposed potential to hamper the acquisition of standard English during secondary and tertiary education (Anchimbe 2013: 11–12).

In Cameroon, we therefore find the usual ambivalence towards the language as in Nigeria. On one side, Cameroon Pidgin serves as a vehicle of regional identity (i.e. the language of unity of the linguistically highly diverse two “anglophone” provinces of Cameroon) and urban identity. On the other side, many Cameroonians, and members of the elite in particular, reproduce a cultural, political, and economic system that privileges English (and French) at the expense of Cameroon Pidgin and other African languages in order to be socially and economically successful. On such an ideological and political backdrop it is unlikely that any form of officialising policy towards Cameroon Pidgin will be undertaken in the near future, while the language continues to naturally expand its domains of use.

### 3.3 Krio (Sierra Leone)

Krio is spoken by the Krio people, to whom it has served as the primary language of community and identity since the emergence of Krio in the late 18th and early 19th century. Through the 19th century presence of Krio people along the west coast of Africa, the Krio language deeply influenced, merged with or replaced pre-existing English-lexicon contact languages in Ghana, Nigeria and Cameroon (cf. e.g. Huber 1995; Huber & Görlach 1996). Through the course of time, Krio has been continuously acquiring additional L1 speakers of diverse ethno-linguistic backgrounds in Freetown, the capital of the country and traditional settlement area of the Krio people.<sup>2</sup> The number of L1 Krio speakers is pegged at somewhere between five and ten per cent of the country's population of six million, but Krio is also the most widely spoken L2 in Sierra Leone, serving as the primary interethnic vehicular language, with a high degree of competence ascribed to virtually the entire population of the country (Finney 2013: 157).

Sierra Leone Krio boasts the longest history of corpus and status planning of all African AECs. The language has enjoyed a comparatively high prestige compared to the other AECs because it has been associated with the Krio ethnic group, which dominated Sierra Leonean society under British colonial rule in the 19th and 20th centuries. Krio therefore received more attention by Sierra Leonean and foreign linguists than other AECs of the region, extending to socio-historical and descriptive work (e.g. Berry 1961; Jones 1968; Hancock 1971 for some earlier works), and the most extensive dictionary of an African AEC to date (Fyle & Jones 1980). The decades after independence in 1961 until the 1991–2002 civil war saw a significant development of Krio literature, chiefly in the form of theatre plays and short stories (cf. e.g. Decker 1988) as well as a recent bible translation (Bible Society in Sierra Leone 2013). Citizens' initiatives at uplifting the status of Krio coincided with state and donor-supported status planning for Krio and other languages of Sierra Leone (i.e. Mende, Temne, Limba) being targeted for inclusion as a subject into school and university curricula (cf. e.g. Anonymous 1984). The relatively high status of Krio also can be seen from the fact that the Universal Declaration of Human Rights was

---

2. The concept of an L1 and L2 is, of course, a problematic one in the multilingual societies of West Africa. I employ the term in this section and following ones on the basis of usage as one of *several primary* languages of socialisation inside and outside of the family from early childhood onwards. According to this definition, an L2 is any language mainly acquired outside of the home, and at later points in life. In the West African contexts I am familiar with, the acquisition of L2s in addition to (various) L1(s) often begins on the cusp of, or during adolescence, with the onset of greater personal mobility and the broadening of personal networks of an individual speaker.



translated into Krio by a state body, the National Commission for Democracy and Human Rights of Sierra Leone.

Much of the administrative and educational infrastructure of Sierra Leone was wiped out during the civil war. But internal displacement from the linguistically more homogenous rural areas during the war, and settlement, population mixture, and accelerated urbanisation in the post-war period have favoured the expansion of Krio as an L2 lingua franca. An increasingly positive attitude towards Krio has concomitantly developed (Finney 2013: 158). The language is nowadays also widely used in broadcasting, TV, advertising, pop music and for civic and health education. Besides the increased practical value of Krio as a nation-wide medium of communication, the popularity of Krio is also due to its relative identitarian “neutrality”, a characterisation that is often also heard with respect to the other African AECs. The lingua franca status of Krio as everybody’s language is perceived to transcend the ethnic divisions fomented by politicians and military leaders before and during the civil war. Krio did not always have a more neutral position in the country. For the longest time after the colonization and creation of Sierra Leone by the British, the country was dominated politically, economically and intellectually by the numerically small Krio people. Krio hegemony, however, saw a steady decline after Sierra Leone’s independence, accompanied by the rise of ethnically more heterogeneous elites (Wyse 1991), thereby possibly contributing to a change in attitudes towards Krio. Today, Krio is spoken as an L1 or L2 by virtually the entire population of the country, thereby even outranking the proportion of Nigerian Pidgin speakers in Nigeria.

In sum, Krio differs quite markedly from the other West African AECs in a number of key socio-linguistic characteristics. Of course, negative attitudes towards Krio still persist due to similar dynamics as with the other West African AECs. Krio nevertheless has a higher prestige than the other AECs, has seen some degree of official recognition and has been the subject of policy making. Krio is also more widely used than other African AECs as a written medium both in secular and religious writing, and with a widely accepted standard orthography. This makes Sierra Leone a potential benchmark for the future status and corpus planning initiatives in the other countries discussed in this chapter.

### 3.4 Ghanaian Pidgin English

If the status of Krio in Sierra Leone is relatively high, the opposite can be said of that of Ghanaian Pidgin English. Ghanaian Pidgin English is not officially recognised in any form, neither legal, nor *de facto*. It is less present in the audio-visual media, or in advertisements than the other AECs (except perhaps Pichi, see 3.5).

Even in pop culture, where the AECs have probably made their greatest advances outside of the interpersonal domain in all AEC-speaking countries, Ghanaian Pidgin English has made less dramatic an incursion than in Nigeria. The majority of songs released in Ghana's highly productive pop music industry are still sung in Akan even if Ghanaian Pidgin English is omnipresent in Hiplife, the Ghanaian offshoot of Hip Hop music (e.g. the complete works of *Fokn Bois*, cf. <[http://youtu.be/R\\_YsQK2Yo3c](http://youtu.be/R_YsQK2Yo3c)>).

A specific set of circumstances renders Ghana a special case with respect to the status and uses of the AEC variety spoken in this country. Firstly, Ghanaian Pidgin English is a language that has been classified as a pidgin (hence a contact language without L1 speakers) rather than a creole (a contact language with at least some L1 speakers) (Huber 1995; Dako 2002). Ghanaian Pidgin English is indeed used in fewer domains than the other three AECs treated above. The lack of sociolinguistic data, however, makes it impossible to determine whether Ghanaian Pidgin English does not also serve as an L1. I have anecdotal evidence for Ghanaian Pidgin being used as a language of the home for vertical communication between fathers and sons, and the language is certainly used for horizontal communication between couples as well as siblings in urban homes from school-going age onwards. All researchers working on Ghanaian Pidgin English, however, agree that the language does not serve as the *primary* language of any ethnic community or region of Ghana.

The absence of a deep family entrenchment may be the reason why Ghanaian Pidgin speakers appear to draw more extensively than those of other African AEC varieties, on lexical items, idioms, and grammatical structures from Ghanaian colloquial English *and* Ghanaian languages (see e.g. Osei-Tutu 2016), why they practice an exceptionally high degree of conventionalised codeswitching with Ghanaian English and other Ghanaian languages, and why Ghanaian Pidgin appear to have incorporated a higher proportion of youth language lexicon, Nigerian Pidginisms, Jamaicanisms, African-American English, and international English slang items into its core system.<sup>3</sup>

---

3. We find, for example, the ubiquitous term of address *chale* [tʃàlɛ́] 'dude, man', a shibboleth for Ghanaian Pidgin speakers. The term goes back to the North American slang term of address *charlie*. It is unclear how and when it arrived in Ghana. Nigerian Pidgin has provided the modal question particle *àbí* (Nordén 2016: 22), and the (pseudo-)Jamaicanisms *wànà* '1PL.POSS' and *dèmà* '3PL.POSS' have been fully integrated into the pronominal system of the Student Pidgin variety of Ghanaian Pidgin English, and have spread to other groups of speakers. For the extensive presence of youth language features into Ghanaian Pidgin, see Osei-Tutu (2008).

For most Ghanaians the encounter with Ghanaian Pidgin English begins during late childhood and adolescence, a time when the language becomes the primary medium of peer group interaction in schools, and in boarding schools in particular (Dako 2002). Ghanaian Pidgin English enjoys popularity with pupils as the icon of a linguistic counterculture against the top-down enforcement of English in classrooms. The language serves as a marker of youth identity and self-identification, increasingly too for girls and young women (Dako 2013). The tendency to use Ghanaian Pidgin English in peer group interactions continues unabated in tertiary institutions, where most students live together in on- and off-campus boarding houses. For the generation that introduced Ghanaian Pidgin English into educational institutions in the 1970s and 80s, and carried it back into their families and neighbourhoods, the language still functions as a “youth” sociolect, and marker of group identity in spite of the now advanced age of these speakers. Due to its adoption and creative expansion by the urban secondary and tertiary educated class of Ghana, a fluent and versatile command of Ghanaian Pidgin English has, paradoxically it may seem, become a marker of membership in the more affluent socio-economic class. At the same time, mastery of the language also provides these very speakers with an important linguistic resource to navigate the “informal” economy in a society characterised by economic uncertainty.

During several research trips to Ghana between 2016 and 2019, I observed Ghanaian Pidgin being spoken in many more areas and domains than I had previously thought, and by all generations but the most elderly, and across all social classes and in contexts where the interactants had shared competence in other Ghanaian languages. I heard Ghanaian Pidgin being used among young men in cities and towns all over the country, and in remote rural areas, including my own village Kute Buem, an isolated hamlet in the Volta Region. In a country characterised by rural-urban (circular) migration, competence in Ghanaian Pidgin signals acquaintance of, and experience with urban life in the capital Accra and other cities, which can constitute important social and economic capital in rural areas.

This development is interesting because Ghanaians of different linguistic backgrounds can usually resort to languages other than Ghanaian Pidgin English in order to communicate effectively. The numerical preponderance of Akan speakers in Ghana has favoured the expansion of Akan as an L2 lingua franca and koine since Ghana’s independence (Yankson 2018). The usage domains of Akan have expanded dramatically since then and it is increasingly penetrating many high domains. There are, for example, Akan-only TV channels and radio stations, and there is extensive Akan-English codeswitching on many primetime TV and news shows. On such a background, it is very likely that Akan and other large Ghanaian languages (e.g. Ewe in the Volta Region and Hausa in the northern regions of the country) will continue

to play a lingua franca role, but Pidgin English will make itself increasingly felt in the Ghanaian linguistic landscape.

### 3.5 Pichi (Equatorial Guinea)

Up to seventy per cent of the population of Bioko island, hence up to 150,000 speakers regularly use Pichi at various levels of nativisation and in various multilingual and multilectal constellations in and outside their homes (Yakpo 2013a: 194). The speaker community of Pichi is ethnically and culturally diverse and includes the traditional creole community of Bioko (the “Fernandinos”, see Lynn 1984) as well as ethnically mixed families. The largest numbers of speakers are probably people who self-identify as ethnic Bube, but have shifted to Pichi as a primary language (referred to as “nuevos criollos” [New Creoles] by Morgades Besari (2011); see also Bolekia Boleká 2005). Pichi is the most widely spoken African language on Bioko, while Fang is used by probably ninety per cent of the population in the continental part of the country. Both Pichi and Fang probably also serve as the most important international languages to common Equatoguineans. Pichi has served as the primary language of interaction for Equatoguineans in dealings with Nigerians, Cameroonians and Ghanaians since at least the beginning of the twentieth century (cf. Lipski 1992). Likewise, “Fang”, a term for a continuum of closely related, mutually intelligible lects, serves as an international language across continental Equatorial Guinea, southern Cameroon, Gabon, and the border zone of The Republic of Congo (see Perrois 1972: 102–105, on the linguistic relations between these varieties).

The political and legal context differs markedly from the realities of language use in Equatorial Guinea. The constitution of Equatorial Guinea (Article 4) declares Spanish, French and Portuguese “official languages” while the “indigenous languages are recognised as an integral part of the national culture”. Three European languages are therefore given legal recognition as languages of government business and officialdom even if neither of the co-official languages French and Portuguese play any significant role for everyday communication among Equatoguineans. From the primary to the tertiary levels, instruction is given exclusively in Spanish, the only *de facto* official language of the country. There is no legally or politically defined role for education in the African mother tongues of the majority of Equatoguinean children (Yakpo 2016a). Equatorial Guinea therefore features an overtly exoglossic language policy that uniquely privileges the colonial language Spanish. Nevertheless, Fang in particular, is increasingly heard in semi-formal contexts, and Fang and other African languages have slots in national radio programming. Some of the indigenous languages of the country have more or less

widely accepted known orthographies, and there is a modest religious literature (e.g. Asociación Cristiana de Traducciones Biblicas n.d.). Pichi is, in contrast, absent from print, online and audio-visual media, and is not normally employed in church, larger social gatherings (e.g. weddings) or other semi-formal settings. Pichi is equally very little present in contemporary pop music culture, a somewhat surprising fact in view of the prominence of music sung in Nigerian Pidgin in the soundscape of Malabo. Interactions in Pichi are therefore almost entirely relegated to the interpersonal domain. Finally, prior to the publication of Yakpo 2009a and 2010, the only existing descriptive work available was a modest grammar sketch and wordlist by de Zarco (1938).

In Equatorial Guinea, as in other African countries, a complex conglomerate of ideological, political and economic factors, socio-cultural dependence on the former coloniser, elitism, a lack of political vision among the ruling elites, weak state institutions and the lack of resources, administrative and technical expertise seems to be responsible for the neglect of African languages by state institutions, and the continuation of colonial language policies. All the indigenous languages of Equatorial Guinea therefore suffer from this status quo. But prevailing linguistic ideologies contribute to the particularly disadvantaged position of Pichi in the linguistic setting of Equatorial Guinea (see Yakpo 2016b).

### 3.6 Summary of findings

In the preceding sections, I have attempted to describe certain sociolinguistic characteristics of the speaker communities of the five major African AECs. Table 2 below provides a summary of these characteristics.

The Krio language on the very right of Table 2 features the largest proportion of speakers in the country it is spoken in (95% of the population of Sierra Leone is estimated to speak Krio). Krio has also conquered the largest number of domains of use of all AECs and has been the subject of status and corpus planning to a higher degree than the other languages. Pichi occupies the opposite end of the cline, with a relatively small proportion of speakers, although Pichi probably is the second largest African language of Equatorial Guinea. Pichi also has the lowest official status and domains of use outside of the interpersonal sphere. Nigerian Pidgin and Cameroon Pidgin fall in the middle range of the spectrum, with proportions of speakers significantly lower than Krio but still substantial. Both languages are probably spoken by a larger percentage of people in their respective homelands than any other African language. Nigerian Pidgin and Cameroon Pidgin are also similar in their sociolinguistic profile in that both languages are used in all domains but the most formal types of (written) literature. Nevertheless, Cameroon Pidgin

Table 2. Comparison of selected socio-linguistic characteristics of West African AECs

Context	Characteristic	Pichi	Ghanaian Pidgin	Cameroon Pidgin	Nigerian Pidgin	Krio
a. Speaker community	L1 community	√	×	√	√	√
	More L2 than L1 speakers	√	√	√	√	√
	Ethnic community language	√	×	√/×	√/×	√
	Proportion of (L1 & L2) speakers	~15%	~20%	~40%	~50%	~95%
b. Status planning & corpus work	Legal recognition	×	×	×	×	√
	Standard orthography	×	×	×	√/×	√
	Medium of instruction	×	×	×	×	×
	School subject	×	×	×	×	√
	Description/ documentation	√	√	√	√	√
c. Domains of use	Creative literature	×	×	√/×	√/×	√
	Religion, written & oral	×	×	√	√	√
	Grass-roots education	×	×	√	√	√
	Political campaigning	×	×	√	√	√
	TV/Radio broadcasting	×	×	√	√	√
	Pop music	×	√	√	√	√
TV/Radio/advertising	×	√	√	√	√	

and Nigerian Pidgin share the plight of all other AECs save Krio, namely the lack of any status and corpus planning, however modest, by the state.

Ghanaian Pidgin occupies an intermediate position between Nigerian Pidgin and Cameroon Pidgin on the one hand, and Pichi on the other. Contrary to Pichi, Ghanaian Pidgin has no native speaker community. The lack of any institutionally driven form of status or corpus planning therefore represents less of a “reality gap” than the absence of any form of institutional support for Pichi with its well-established native speaker community. The only other language to serve as an “ethnic community language” besides Pichi is Krio. The difference between this characteristic and the one in the first line (“L1 community”) lies in the fact that Pichi and Krio are perceived in Equatorial Guinea and Sierra Leone respectively as the primary languages or L1s of specific ethno-linguistic communities (the Krio people in Sierra Leone and the Fernandino and Bubi peoples in Equatorial Guinea). Although there are large and growing L1 communities of Nigerian Pidgin and Cameroon Pidgin as well, these languages do not serve as markers of ethnic identity. These languages are therefore still primarily considered to be lingua francas

without highly visible native speaker communities. At the same time, all five languages, including Pichi, have more L2 than L1 speakers, a fact that is indicative of the dramatic expansion of these languages since their emergence.

Other than that, the sociolinguistic characteristics of Pichi and Krio could not be more dissimilar, despite them being the two most closely related languages in Table 2. Krio is the only AEC to have been the subject of considerable status planning and corpus work, whether initiated by state institutions or through private and academic initiative. All other languages are characterised by less activity in this regard, and Pichi once more features the least activity even if Pichi and Nigerian Pidgin feature the most extensive grammatical descriptions to-date of any West African AEC (Yakpo 2009a, 2019; Faraclas 1996).

On the whole, Pichi appears as the West African AEC with the lowest status although it serves as an important L1 in Equatorial Guinea. The most striking aspect of the marginalisation of Pichi is its conspicuous absence, some exceptions notwithstanding, from relatively unregulated domains such as pop music, advertising and (private) broadcasting, domains in which the other West African AECs have made their greatest advances. Even Ghanaian Pidgin, which does not function as an L1 nor extensively as a lingua franca, surpasses Pichi in the number of domains in which it is used. However, Equatorial Guinea is generally characterised by the lack of an organised state response to linguistic diversity within its borders. The country has the most profoundly exoglossic language policy in the region, with reliance on the European colonial language Spanish in all but the most informal domains.

Overall, the discrepancies between the status and actual uses of the West African AECs leaves the impression that the potential of these languages remains untapped by official policies across the region for education, political participation, economic activity and cultural production. African countries record some of the lowest literacy rates, the highest school dropout rates, and the lowest tertiary education rates in the world in spite of some of the highest per capita investments in education in relation to GDP (UNESCO 2018; The World Bank 2018).

At least part of this predicament has been attributed to the continuing predominance of European colonial languages as languages of instruction and literacy (see, e.g., the contributions in Baldauf & Kaplan 2004, 2007). The cognitive advantages of employing African languages for education have been amply demonstrated by study upon study in the last five decades (cf. Brock-Utne 2010). Contrary to received opinion in West Africa, the AECs might have the additional advantage, if used as languages of instruction, of facilitating the acquisition of Standard English due to lexical affinities between the two languages, and their often seamless integration in codemixing. The potential benefits of using AECs and other African languages for civic and health communication, the interaction of citizens with administrative

institutions, political campaigning, in the courtroom and legal for practice for the construction of more transparent and participatory political cultures are almost self-evident. Less fast-lived forms of cultural production beyond pop music and cinema, like literary production and drama could be reinvigorated. This would, however, require concerted efforts at corpus and status planning, ideally involving the exchange of knowledge, experiences and practices between the countries in which the AECs are spoken. In coming decades, the AECs are also likely to contribute significantly to language contact and change, and become the target of language shift from other African languages spoken in the cities and in rural areas with a high density of smaller languages (e.g. in the Niger Delta federal states of Nigeria or in the Cameroonian Grassfields region). If linguistic diversity is to be maintained and nurtured in the future, this will make structured and concerted approaches to the status and uses of the AECs all the more important.

#### 4. Conclusion

The West African AECs facilitate communication across national, regional and ethnic boundaries in a culturally and linguistically highly diverse part of the world. In addition, various socio-cultural factors contribute to the expansion these languages have seen in the last few decades: the AECs are first and foremost languages of the city, and therefore have the lustre of urban life; they serve as youth sociolects in a region with the youngest population on earth; embody a tolerance for variation as well as ethnic indeterminacy and openness; and symbolise and express in their dynamism, adaptability and spatial distribution the African experience of the 21st century, characterised by rapid socio-economic and cultural change, and accelerated urbanisation.

The expansion of the AECs has taken place despite administrative and institutional practices ranging from inertia and indifference (e.g. Equatorial Guinea), to hostility (e.g. the banning of Cameroon Pidgin in secondary and tertiary education institutions). The absence of, or negative state responses to the African English-lexifier creoles and pidgins is the consequence of a complex constellation of economic and political factors that other African languages are also subjected to. Ideological factors also play a significant role in negative attitudes of sections of the elites and decision makers vis-à-vis the AECs. Languages like Nigerian Pidgin and Cameroon Pidgin are still widely seen as forms of “broken English” without grammatical structure and unfit for use other than in the streets.

In spite of the lack of state support in all five countries, the utility of these languages for communication and their identitarian appeal have nevertheless led



to their conquest of new domains. The increased presence of these languages in the public sphere and in the media is contributing to shifts in language attitudes. The AECs can count on a growing language loyalty of the younger generations. This might be an epiphenomenon of an increasing desire of young Africans to forge and enact a distinctly “glocal” identity of their own. With their rootedness in local linguistic practices and simultaneous international references the AECs are particularly suited to express the *Lebensgefühl* of African modernity. The African AECs will therefore continue expanding their domains and social functions with or without state support, thereby forcing language policy choices on decision makers and language stakeholders *nolens volens*.

## Acknowledgements

I am grateful to two anonymous reviewers for their comments and suggestions for improvement, as well as valuable input from the audience during a talk given at the West African Languages Congress & Linguistics Association of Ghana Conference 2017 in Winneba, Ghana. This chapter contains material from Yakpo 2016b.

## Funding

I herewith acknowledge the support of grant no. 17608819 of the Research Grants Council of the Government of Hong Kong.

## References

- Adebija, E. 2004. Language policy and planning in Nigeria. *Current Issues in Language Planning* 5(3): 181–246.
- Agheyisi, R. N. 1988. The standardization of Nigerian Pidgin English. *English World-Wide* 9(2): 227–241. <https://doi.org/10.1075/eww.9.2.06agh>
- Alobwede D’Epie, C. 1998. Banning Pidgin English in Cameroon? *English Today* 14(1): 54–60. <https://doi.org/10.1017/S0266078400000742>
- Anchimbe, E. A. 2013. *Language Policy and Identity Construction: The Dynamics of Cameroon’s Multilingualism* [Impact: Studies in Language and Society 32]. Amsterdam: John Benjamins. <https://doi.org/10.1075/impact.32>
- Anonymous. 1984. *Revised Kono, Krio, Limba, Mende and Themne Orthographies – Workshop Held during April 24–27, 1984*. Freetown, Sierra Leone: Ministry of Education.
- Arends, J., Muysken, P. C. & Smith, N. S. H. (eds). 1995. *Pidgins and Creoles: An Introduction* [Creole Language Library 15]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15>
- Asociación Cristiana de Traducciones Bíblicas. n.d. *Kálara Nzamá Farj Guinea Ecuatorial*. Bata: n.p.

- Ayafor, M. & Green, M. J. 2017. *Cameroon Pidgin English: A Comprehensive Grammar* [London Oriental and African Language Library 20]. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/loall.20>
- Bakker, P. 2008. Pidgins versus creoles and pidgincreoles. In *The Handbook of Pidgin and Creole Studies*, S. Kouwenberg & J. V. Singler (eds), 130–57. Oxford: Blackwell.
- Baldauf, R. B. & Kaplan, R. B. 2004. *Language Planning and Policy in Africa*, Vol. 1. Clevedon: Multilingual Matters.
- Baldauf, R. B. & Kaplan, R. B. 2007. *Language Planning and Policy in Africa*, Vol. 2. Clevedon: Multilingual Matters.
- Berry, J. 1961. English loanwords and adaptations in Sierra Leone Krio. In *Creole Language Studies II*, R. B. le Page (ed.), 1–16. London: Macmillan.
- Bible Society in Sierra Leone. 2013. *Krio fɔs oli Baybul*.
- Bolekia Boleká, J. 2005. *La francofonía: El nuevo rostro del colonialismo en África*. Salamanca: Signum S.G.E.
- Brock-Utne, B. 2010. Research and policy on the language of instruction issue in Africa. *International Journal of Educational Development* 30(6): 636–645.
- Corum, M. 2015. *Substrate and Adstrate: The Origins of Spatial Semantics in West African Pidgincreoles*. Berlin: De Gruyter Mouton.
- Crevels, M. & Muysken, P. C. (eds). 2009. *Lenguas de Bolivia*, 4 Vols. La Paz: Plural editores.
- Dako, K. 2002. Student Pidgin (SP): The language of the educated male elite. *Research Review from the Institute of African Studies (IAS), University of Ghana* 18(2): 53–62.
- Dako, K. 2013. Student Pidgin: A masculine code encroached on by young women. In *Gender and Language in Sub-Saharan Africa: Tradition, Struggle and Change* [Impact: Studies in Language and Society 33], L. L. Atanga, S. E. Ellece, L. Litosseliti & J. Sunderland (eds), 217–231. Amsterdam: John Benjamins. <https://doi.org/10.1075/impact.33>
- Decker, T. 1988. *Juliohs Siza*. Edited by N. Shrimpton and N. Sulayman, Vol. 4 [Krio Publications Series]. Umeå: Umeå University.
- Devonish, H. 1986. *Language and Liberation: Creole Language Politics in the Caribbean*. London: Karia Press.
- Devonish, H. 2010. The language heritage of the Caribbean: Linguistic genocide and resistance. *Glossa* 5(1): 1–26.
- Faraclas, N. G. 2004. Nigerian Pidgin English: Morphology and syntax. In *A Handbook of Varieties of English, Vol 2: Morphology and Syntax*, B. Kortmann, E. W. Schneider, K. Burridge, T. Mesthrie & C. Upton (eds), 828–853. Berlin: Mouton de Gruyter.
- Faraclas, N. G. 1996. *Nigerian Pidgin*. London: Routledge.
- Faraclas, N. G. 2013. Nigerian Pidgin. In *The Survey of Pidgin and Creole Languages: English-based and Dutch-based Languages*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 176–184. Oxford: OUP. <<https://apics-online.info/surveys/17>> (9 June 2020).
- Faraclas, N. G. 1986. *Reading and Writing Nigerian Pidgin*. Port Harcourt: Rivers Readers Project (RRP).
- Finney, M. A. 2011. Krio (Sierra Leone Creole). In *Electronic World Atlas of Varieties of English*, Bernd Kortmann & Kerstin Lunkenheimer (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://ewave-atlas.org/languages/36>> (23 April 2020).
- Finney, M. A. 2013. Krio. In *The Survey of Pidgin and Creole Languages: English-based and Dutch-based Languages*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 1:157–166. Oxford: OUP. <<https://apics-online.info/surveys/15>> (23 April 2020).

- Fonka, H. M. 2016. Is Cameroon Pidgin English a pidgin, a pidgincreole or a creole? Yaoundé.
- Fyle, C. N. & Jones, E. D. 1980. *A Krio-English Dictionary*. Oxford: OUP.
- Gani-Ikilama, T. O. 1990. Use of Nigerian Pidgin in education: Why not? In *Multilingualism, Minority Languages and Language Policy in Nigeria*, E. N. Emenanjo (ed.), 219–227. Agbor, Nigeria: Central Books.
- Gift, A. 2019. Romeo and Juliet don land for pidgin. *BBC News Pidgin*, July 4, 2019, sec. Tori. <<https://www.bbc.com/pidgin/media-48851624>> (26 September 2019).
- Hancock, I. F. 1971. A Study of the Sources and Development of the Lexicon of Sierra Leone Krio. PhD dissertation, University of London.
- Hancock, I. F. 1987. A preliminary classification of Anglophone Atlantic creoles, with syntactic data from thirty-three representative dialects. In *Pidgin and Creole Languages: Essays in Memory of John Reinecke*, G. G. Gilbert (ed.), 264–333. Honolulu HI: University of Hawai'i Press.
- Heine, B. 1992. Language policies in Africa. In *Language and Society in Africa: The Theory and Practice of Sociolinguistics*, R. K. Herbert (ed), 23–35. Johannesburg: Witwatersrand University Press.
- Huber, M. 1995. Ghanaian Pidgin English: An overview. *English World-Wide* 16(2): 215–249. <https://doi.org/10.1075/eww.16.2.04hub>
- Huber, M. 1999. *Ghanaian Pidgin English in Its West African Context: A Sociohistorical and Structural Analysis* [Varieties of English Around the World G24]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g24>
- Huber, M. 2000. Restructuring in vitro? Evidence from early Krio. In *Degrees of Restructuring in Creole Languages* [Creole Language Library 22], I. Neumann-Holzschuh & E. W. Schneider (eds), 275–307. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.22>
- Huber, M. 2013. Ghanaian Pidgin English. In *The Survey of Pidgin and Creole Languages: English-based and Dutch-based Languages*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 1:167–175. Oxford: OUP. <<https://apics-online.info/surveys/16>> (23 April 2020).
- Huber, M. & Görlach, M. 1996. West African Pidgin English. *English World-Wide* 17(2): 239–258. <https://doi.org/10.1075/eww.17.2.07hub>
- Igboanusi, H. 2008. Empowering Nigerian Pidgin: A challenge for status planning? *World Englishes* 27(1): 68–82. <https://doi.org/10.1111/j.1467-971X.2008.00536.x>
- Ihemere, K. U. 2006. A basic description and analytic treatment of noun clauses in Nigerian Pidgin. *Nordic journal of African Studies* 15(3): 296–313.
- Jones, E. D. 1968. Some tense, mode and aspects markers in Krio. *African Language Review* 7: 86–89.
- Juffermans, K. & McGlynn, C. 2010. A sociolinguistic profile of The Gambia. *Sociolinguistic Studies* 3(3): 329–355.
- Kouega, J.-P. 2008. *A Dictionary of Cameroon Pidgin English Usage: Pronunciation, Grammar and Vocabulary* [Lincom Studies in Pidgin & Creole Linguistics 9]. Munich: Lincom.
- Lewis, M. P., Simons, G. F. & Fennig, C. D. (eds). 2013. *Ethnologue: Languages of the World*. Dallas TX: SIL International. <<http://www.ethnologue.com>> (30 May 2013).
- Lipski, J. M. 1992. Pidgin English usage in Equatorial Guinea (Fernando Poo). *English World-Wide* 13(1): 33–57. <https://doi.org/10.1075/eww.13.1.03lip>
- Lynn, M. 1984. Commerce, christianity and the origins of the “Creoles” of Fernando Po. *The Journal of African History* 25(3): 257–278.
- Mbangwana, P. 1983. The scope and role of Pidgin English in Cameroon. In *A Sociolinguistic Profile of Urban Centers in Cameroon*, E. L. Koenig, E. N. Chia & J. F. Povey (eds), 79–92. Los Angeles CA: Crossroad Press.

- Molindo, E. A. 1996. *Cameroon Pidgin Bible*. Yaoundé: n.p.
- Momoh, A. & Adejumobi, S. 2002. *The National Question in Nigeria: Comparative Perspectives*. Aldershot Burlington: Ashgate.
- Morgades Besari, T. 2011. Los criollos (Fernandinos-Kriös) de Guinea Ecuatorial. *La Gaceta de Guinea Ecuatorial*, April 2011, No. 162 edition, sec. Misceláneas.
- Muysken, P. C. (ed.). 1981a. *Generative Studies on Creole Languages* [Studies in Generative Grammar 6]. Dordrecht: Foris.
- Muysken, P. C. 1981b. Halfway between Quechua and Spanish: The case for relexification. In *Historicity and Variation in Creole Studies*, A. Valdman & A. Highfield (eds), 52–78. Ann Arbor MI: Karoma.
- Muysken, P. C. (ed.). 2008. *From Linguistic Areas to Areal Linguistics* [Studies in Language Companion Series 90]. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.90>
- Muysken, P. C. & Smith, N. S. H. (eds). 1986. *Substrata Versus Universals in Creole Genesis: Papers from the Amsterdam Creole Workshop, April 1985* [Creole Language Library 1]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cl.1>
- Muysken, P. C. & Smith, N. S. H. (eds). 2015. *Surviving the Middle Passage: The West Africa-Surinam Sprachbund* [Trends in Linguistics, Studies and Monographs (TiLSM) 275]. Berlin: De Gruyter Mouton.
- Neba, A. N., Chibaka, E. F. & Atindogbé, G. 2006. Cameroon Pidgin English (CPE) as a tool for empowerment and national development. *African Study Monographs* 27: 39–61.
- Newman, P. 1987. Hausa and the Chadic languages. In *The World's Major Languages*, B. Comrie (ed.), 705–723. Oxford: OUP.
- Ngefacs, A. 2010. Linguistic choices in postcolonial multilingual Cameroon. *Nordic Journal of African Studies* 19(3): 149–164.
- Ngefacs, A. 2016. *Sociolinguistic and Structural Aspects of Cameroon Creole English*. Newcastle upon Tyne: Cambridge Scholars.
- Nkengasong, N. 2016. *A Grammar of Cameroonian Pidgin*. Newcastle upon Tyne: Cambridge Scholars.
- Nordén, A. H. 2016. Epistemic Modality in Ghanaian Pidgin English. MA thesis, University of Stockholm. <<http://www.diva-portal.org/smash/get/diva2:940706/FULLTEXT01.pdf>> (26 September 2019).
- Ojarikre, A. 2013. Perspectives and problems of codifying Nigerian Pidgin English orthography. *Developing Country Studies* 3(12): 126–133.
- Osei-Tutu, K. O. A. 2008. Exploring Meaning in Students' P(SP). MPhil thesis, University of Ghana, Legon.
- Osei-Tutu, K. O. A. 2016. Lexical borrowing in Ghanaian Student Pidgin (GSP) – The case of Akan loanwords and loan translations. In *Celebrating Multiple Identities: Opting out of Neocolonial Monolingualism, Monoculturalism and Mono-identification in the Greater Caribbean*, N. Faracis, R. Severing, C. Weijer, E. Echteld, W. Rutgers & R. Dupey (eds), 47–54. Willemstad: University of Curaçao and Fundashon pa Planifikashon di Idioma.
- Perrois, L. 1972. *La statuaire Fang, Gabon*. Paris: O.R.S.T.O.M.
- Schröder, A. 2003. *Status, Functions, and Prospects of Pidgin English: An Empirical Approach to Language Dynamics in Cameroon*. Tübingen: Gunter Narr.
- Smith, N. S. H. 2015. Ingredient X: The shared African lexical element in the English-lexifier Atlantic Creoles, and the theory of rapid creolization. In *Surviving the Middle Passage: The West Africa-Surinam Sprachbund* [Trends in Linguistics, Studies and Monographs (TiLSM) 275], P. C. Muysken & N. S. H. Smith (eds), 67–106. Berlin: De Gruyter Mouton.

- The World Bank. 2018. Education statistics. *World Bank Open Data*, 2018. <<http://datatopics.worldbank.org/education/>> (4 April 2018).
- UNESCO. 2018. Global Education Monitoring Report. *2017/8 GEM Report – Accountability in Education: Meeting Our Commitments*, 2018. <<https://en.unesco.org/gem-report/statistics>> (4 April 2018).
- United Nations. 2017. World population prospects: The 2017 revision, key findings and advance tables [Working Paper ESA/P/WP/248]. New York: United Nations Department of Economic and Social Affairs, Population Division. <[https://population.un.org/wpp/Publications/Files/WPP2017\\_KeyFindings.pdf](https://population.un.org/wpp/Publications/Files/WPP2017_KeyFindings.pdf)> (24 April 2020).
- van Sluijs, R., van den Berg, M. C. & Muysken, P. C. 2016. Exploring genealogical blends: The Surinamese Creole cluster and the Virgin Island Dutch Creole cluster. *Lingua* 178: 84–103. <https://doi.org/10.1016/j.lingua.2015.12.004>
- Vydrin, V. 2013. Bamana Reference Corpus (BRC). *Procedia – Social and Behavioral Sciences* 95: 75–80. <https://doi.org/10.1016/j.sbspro.2013.10.624>
- Wyse, A. J. G. 1991. *The Krio of Sierra Leone: An Interpretive History*. Washington DC: Howard University Press.
- Yakpo, K. 2009a. A Grammar of Pichi. PhD dissertation, Radboud University Nijmegen. <https://doi.org/10.5281/zenodo.3740353> (27 March 2020).
- Yakpo, K. 2009b. Complexity revisited: Pichi (Equatorial Guinea) and Spanish in contact. In *Simplicity and Complexity in Creoles and Pidgins*, N. G. Faraclas & T. Klein (eds), 183–215. London: Battlebridge.
- Yakpo, K. 2010. *Gramática del Pichi* [Laboratorio de Recursos Orales 13]. Barcelona: Ceiba Ediciones.
- Yakpo, K. 2012a. Betwixt and between: Causatives in the English-lexicon creoles of West Africa and the Caribbean. In *Analytical Causatives: From “Make” to “Laskma”*, J. Leino & R. von Waldenfels (eds), 9–39. Munich: Lincom.
- Yakpo, K. 2012b. Reiteration in Pichi: Forms, functions and areal-typological perspectives. In *The Morphosyntax of Reiteration in Creole and Non-creole Languages* [Creole Language Library 43], E. O. Aboh, N. S. H. Smith & A. Zribi-Hertz (eds), 251–284. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.43.08yak>
- Yakpo, K. 2013a. Pichi. In *The Atlas of Pidgin and Creole Language Structures: English-based and Dutch-based Languages*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 1:194–205. Oxford: OUP. <<http://apics-online.info/contributions/19>> (2 March 2018).
- Yakpo, K. 2013b. Wayward daughter: Language contact in the emergence of Pichi (Equatorial Guinea). *Journal of African Languages and Linguistics* 34(2): 275–299. <https://doi.org/10.1515/jall-2013-0009>
- Yakpo, K. 2016a. O estatuto do pichi na Guiné Equatorial. *PLATÔ (Revista Digital do Instituto Internacional da Língua Portuguesa)*, *Glotopolítica na Guiné Equatorial* 3(6): 20–40.
- Yakpo, K. 2016b. “The only language we speak really well”. The English creoles of Equatorial Guinea and West Africa at the intersection of language ideologies and language policies. In *Exploring Glottopolitical Dynamics in Africa: The Spanish Colonial Past and Beyond*, S. Castillo-Rodríguez & L. Morgenthaler García (eds). Special issue of *International Journal of the Sociology of Language* 239: 211–233. <https://doi.org/10.1515/ijsl-2016-0010>
- Yakpo, K. 2017. Towards a model of language contact and change in the English-lexifier creoles of Africa and the Caribbean. *English World-Wide* 38(1): 50–76. <https://doi.org/10.1075/eww.38.1.04yak>

- Yakpo, K. 2018. ¿El nacimiento de una lengua afrohispana?: La influencia del español en el idioma criollo inglés de Guinea Ecuatorial. In *Trans-afrohispanismos: Puentes culturales críticos entre África, Latinoamérica y España* [Foro Hispánico 58], D. Odartey-Wellington (ed.), 243–259. Leiden: Brill. [https://doi.org/10.1163/9789004364080\\_015](https://doi.org/10.1163/9789004364080_015)
- Yakpo, K. 2019. Inheritance, contact, convergence: Pronominal allomorphy in the African English-lexifier Creoles. *English World-Wide* 40(2): 201–225. <https://doi.org/10.1075/eww.00028.yak>
- Yakpo, K. & Muysken, P. C. (eds). 2017. *Boundaries and Bridges: Language Contact in Multilingual Ecologies* [Language Contact and Bilingualism 14]. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9781614514886>
- Yankson, S. 2018. Language Contact and Change in Linguistically Heterogeneous Urban Communities. The Case of Akan in Accra. PhD dissertation, Radboud University Nijmegen.
- de Zarco, Mariano. 1938. *Dialecto Inglés-Africano o Broken-English de La Colonia Española Del Golfo de Guinea*. Turnhout, Belgium: H. Proost.



# The quest for non-European creoles

## Is Kukama (Brazil, Peru) a creole language?

Peter Bakker  
Aarhus University

Kukama has been classified firstly, as a Tupi-Guarani language, secondly, as a language that has undergone massive contact-induced change, and thirdly, as a creole. These different positions are surveyed against definitions of mixed languages and creoles, and properties of Tupi-Guarani languages. We focus especially on its possible status as a creole language, as there is a need among creolists to expand the range of creoles with a broad set of lexifiers and substrates, especially after recent research has revealed a typological profile for creoles. With typological, historical and historical-comparative arguments, it is concluded that the original grammatical system of Kukama was strongly reduced, and subsequently innovated and expanded with mostly Tupi-Guarani roots, a process parallel to that gone through by accepted creoles. As is commonly the case in creoles, very little beyond the lexicon is inherited from its lexifier, a language close to Tupinamba. However, the Kukama language does not quite fit two proposals for creole prototypes, perhaps because the processes of loss and innovation took place in a much more remote past than for identified creoles. In this paper, it will be discussed whether Kukama is a language that defies classification within the current typology of contact languages, or whether it is deviant within current family classifications.

**Keywords:** creole languages, Tupi-Guaraní, Kukama, typology, historical linguistics, mixed languages, grammaticalization, creolization

### 1. Introduction

Pieter Muysken has had a long-term interest in disparate fields of linguistics, among them mixed languages, the languages of South America, and creoles. It is an honor for me to contribute to this volume dedicated to my *Doktorvater*. My contribution straddles all three areas of his interests. In other words, this contribution is about the South American indigenous language Kukama (earlier also called Cocama or



Kokama), and its status as a mixed language or as a creole. Kukama is taken here as a general term for Kukama proper and its related dialect Kukamiria (hence κκ in some of the quotes). κκ is closely related to a third language, Omagua, with quite similar grammatical features (and κo refers therefore to Kukama/Kukamiria and Omagua). These indigenous forms of speech are, or were until recently, spoken in Brazil (Kukama) and Peru (all three). They have been subsumed under the Tupi-Guarani languages, but as outliers, and also as non-genetic contact languages, among them the two subtypes of mixed languages and creole languages. In my contribution, I focus exclusively on Kukama/κκ, as my best documentation available is for this language, but much of what I say is also true for Omagua.

In the recent past, Master Pieter himself (Muysken 2012: 249–251) wrote a few pages about Kukama in an overview article on contact between indigenous languages in South America, which he ended with a hope for more fieldwork leading to an investigation of “the precise grammatical features, lexical roots, and morphological properties of this language”. A few years later his wish was fulfilled, with Vallejos’ detailed grammatical description (2016) based on her dissertation, and a dictionary (Vallejos & Murayari 2015).

In this contribution, I will discuss Kukama from a contact-language perspective. Is it a contact language, and if so, is it a mixed language or a creole or something else? Being spoken far from the influence from West African Niger-Congo languages and far from Asian Austronesian languages, one could expect Kukama, if it is indeed a creole, to be quite different from other creoles. Also, the time depth of the genesis of identified creoles in the literature is less than 500 years, but it is likely that the contact influence of Kukama is older (Michael 2014), and that may mean that creole features in this language are more difficult to identify, as creoles, once crystallized, develop like any other natural language.

Most traditional classifications of Tupi and Tupi-Guarani (TG) languages include Kukama among the TG languages. TG languages form one branch of the Tupi stock. There are some 75 languages in the Tupi stock, and around 50 of them belong to the TG branch. One of the more recent classifications, however, no longer includes Kukama and the two sister languages (Rodrigues & Cabral 2012) among the TG languages, as they are relegated to the category of non-genetic languages (cf. Thomason & Kaufman 1988), because of its contact-induced changes, to be discussed below. According to Cabral (1995), the language would be a radical contact language based on the Tupinamba language, once a widespread *lingua franca* in Brazil, or on a closely related language, and therefore a new language and not a descendant of Tupinamba or a member of the TG family.

In Section 2, I will summarize some of the literature on Kukama as a contact language: it has been called a creole, and also a contact language of an unspecified type, and also the result of imperfect shift. In Section 3, I will discuss some

classifications of the Tupi-Guarani subgroup, and the position of Kukama within this group. In Section 4, I discuss some recent research on the typology of creole languages, and I will relate this to Kukama in connection with two typological characterizations of creoles. A concluding Section 5 ends the paper. Thus, we try to give an answer to the question of whether Kukama is a non-genetic language, more specifically a creole, or a regular member of the TG family.

## 2. Kukama as a contact language

Aryon Rodrigues (1984–1985: 43–44) has studied the sound changes in TG languages and provided a classification of TG languages. The author noted casually that KK looked like an originally TG language, but one spoken by people who had shifted to it, suggesting influences from other languages.

Wolf Dietrich (1990: 115) was perhaps the first to point out that Kukama was heavily affected by language contact, when he wrote, in a survey of structural features of TG languages, that Kukama (and also Guayaki) “may be members of the Tupi-Guarani language family, but these languages are to such an extent mixed up with features atypical of the Tupi-Guarani languages that they must have been taken over by people who were originally speakers of languages belonging to other language stocks, or they must have been extensively influenced by these speakers.” Dietrich did not specify what kind of contact language Kukama could be, as his purpose was language classification and comparison of grammatical features. His description resembles what happened in the case of creolization, as creoles are languages with the lexicon of one language (the lexifier) and a grammatical system that is considerably different structurally from the lexifier, with generally easily identifiable grammaticalized elements from the lexifier which are otherwise not present in the lexifier.

Ana Suelly Cabral was the first to argue in detail for Kukama as a contact language in her dissertation (Cabral 1995). She argues convincingly for a contact origin of the language: she considers Tupinamba to be the source language of the lexicon, but on the other hand, there are only a few grammatical morphemes found in Kukama that are also found in Tupinamba or in Tupi-Guarani languages generally. The actual proportion is difficult to quantify but see O’Hagan (2011) for reconstructions of KK and Omagua, and Vallejos (2016: 613) for a list of some 30 grammatical morphemes and function words that can be linked with Proto-TG roots. Syntactically, Kukama is also rather different. On the other hand, the bulk of the lexicon is clearly Tupi-Guaraní, and many of the grammatical markers are clearly of Tupi-Guaraní origin, often relatable to open-class words. Her detailed discussion leaves no doubt about the special position of Kukama within the Tupi-Guarani subgroup.

However, she is more ambiguous in her dissertation as to what kind of contact language it would be. Let us first characterize three types of contact languages. Some consider all three of them mixed, but here we distinguish creoles from two types of mixed languages (see also Bakker 2020).

*Creoles* can be characterized as languages with a basic lexicon from one language, rarely two, but only some of the grammatical structures are inherited from their lexifiers (the source of the lexical roots). Typically, inflectional morphology and gender distinctions are not inherited from the lexifiers, with very few exceptions. The grammatical structures are not traceable to any single language, or set of languages, even though the languages spoken by the creators of the creoles (substrate languages), or spoken beside the creole (adstrate languages), asserted influence and contributed some structural traits as well. The bulk of the new grammatical distinctions are created in order to build a full-fledged language out of a reduced, perhaps pidginized, form of the lexifier with severely downsized lexical and grammatical properties. All of the existing identified creoles are highly complex but almost all are typically analytic languages. The complexities are never found in morphological distinctions. In short, creoles have a lexicon of one language, but a spontaneously developed, mostly analytic grammar, where identifiable lexical items grammaticalized into function words. Creoles could be seen as an extreme form of language shift, but without the lexifier being a model or target. Creoles are, in my view, not mixed languages, because a mixture requires two or more sources, and the other, non-lexical, sources, can usually not be identified, or the influence is not pervasive.

There are several types of *mixed languages* (Bakker 2017, 2020; Meakins 2016), and two of these are potentially relevant for Kukama. One type has been called “**intertwined**” (Bakker & Muysken 1994; Bakker 1997). These languages combine the lexicon of one language with the grammatical system of another. The grammatical system here refers to morphology, syntax, and phonology, sometimes also lexical semantics. This type is clearly different from the creole type, in that the bound morphemes and some of the free grammatical morphemes (e.g. adpositions, demonstratives) are identical to those of an identifiable language in intertwined languages, but that is not the case in creoles. Only if the grammaticalizing language is an analytic or isolating language, is there potential confusion between an intertwined language and a creole. In creoles, basically all morphemes – lexical and grammatical – come from the same language, but in intertwined languages, the lexical and grammatical morphemes typically derive from two different sources – and therefore they can be called mixed.

The second type subsumed under mixed languages are sometimes called **converted** languages, and they are the result of a process sometimes called **metatypy**. In converted languages, native materials are used to copy structures, for instance

prepositions from the source language are made into case markers when the model language has a set of case markers. Like in creoles, all the morphemes are from the same source languages, and like intertwined languages, they can be morphologically complex, but here the grammatical and semantic structures have a clear model. The distinction between the three types is clear. We can summarize the characteristics as in Table 1.

**Table 1.** Mixed languages and creoles: Similarities and differences  
(L = Lexicon language, G = Grammar language)

	Creoles	Intertwined mixed languages	Converted mixed languages
Etymology of bulk of lexical roots	One language L	One language L	One language L
Etymology of bulk of grammatical morphemes	One language L	One language G	One language L
Origin of grammatical constructions	Mostly spontaneous creations through grammaticalization of L lexical roots, some inherited from L or inspired by other languages	Easily identifiable as from language G, including the forms from G	Identifiable as copies modelled on language G, but the forms are from L
Origin of free grammatical morphemes	Almost all from L, some taken over directly, most derived/grammaticalized from nominal or verbal roots	From L or G, or both, including the forms	From language G, but the forms are from L

The question is then: does Kukama fit one of these three categories, and if not, could it be another, perhaps newly identified, and perhaps unique, type of contact language? Cabral (1995) excluded the possibility of Kukama being either of the two types of mixed language: “Kokama does not look like a language (in this case a Tupi-Guarani language) that had undergone massive grammatical borrowing, because Kokama grammar does not come from one or two specific language sources.” (Cabral 1995: 308).

Could it then be a creole? According to Cabral, the grammatical system is not from an identifiable source, i.e. not traceable to Tupi-Guarani, and that would point to the possibility of Kukama being a creole language, when she mentions “evidence of imperfect learning of the word structure of the language source and a non-Tupi-Guarani grammar, which in turn does not come from any language source in particular” (Cabral 1995: 308).

In addition, the language shows, according to Cabral, similarities with identified creoles: “Kokama displays many linguistic features that are characteristic of languages that originated in multilingual contexts by means of a language shift that occurred within a small period of time, such as abrupt creoles” (Cabral 1995: 307).

All identified creoles are close to the type of isolating languages, and that could be another argument for her to consider the language to be a creole: “The fact that Kokama is an isolating language, while the known Amazonian languages spoken in the Upper Amazon area at the time in question all were inflectional languages, including Tupinamba, favors the view that although they shared this typological characteristic, they had different degrees of complexity in their structures, together with different functional and positional types of affixes that could allow morphological overlap” (Cabral 1995: 308). In short, the evidence in 1995 pointed in the direction of Kukama as a creole language.

A specific typological profile for creoles was earlier suggested by McWhorter (1998, 2001, 2005). In recent years, massive evidence has been put forward that shows that previously identified creoles, despite their enormous grammatical and lexical diversity, differ systematically from samples of non-creoles (Szmrecsanyi & Kortmann 2009; Bakker et al. 2013/2011; Daval-Markussen 2014; the chapters in Bakker et al. 2017; Clements 2019). One of the criticisms of these studies was that the selected languages were biased in the direction of European-lexifier creoles, and to creoles with Niger-Congo substrates and Western European lexifiers. Kukama thus offers an interesting test case: does Kukama comply with recurring structural characteristics of creoles from other parts of the world? In other words, does this language share the proposed characteristic traits of creoles?

Since Cabral’s study, further publications have become available on Kukama. Especially the impressive grammatical description by Vallejos (2016), and her articles, contributed to the documentation. In addition, important progress has been made in recent years on the linguistic and social history of not only Kukama but also Omagua, especially by Lev Michael and Zachary O’Hagan. Their use of results from research on contact languages, creoles and language contact effects is somewhat limited, and a language contact perspective on Kukama is therefore most welcome. In Michael and O’Hagan (2016), they discuss texts in Omagua from the 17th and 18th centuries.

Michael (2014) agreed with Cabral’s assessment that Kukama-Omagua was a contact language, but he disagreed with Cabral about the socio-historical scenario. Cabral related the genesis of Kukama to the mixed compounds (“reducciones”) into which Jesuit missionaries had forced indigenous groups before the expulsion of the Jesuits in 1767, which at the time was a reasonable assumption. The Jesuits had been active in the region from the 1630s (Michael & O’Hagan 2016). Michael (2014),

however, proved Cabral wrong at this point, and he argued convincingly for an earlier, possibly pre-Columbian origin of the contact features of Kukama-Omagua. The earliest sources of Omagua indeed show structures that are quite similar to what is found in the modern languages. In short, there is agreement that Kukama has undergone sufficient external influence in its grammar to be considered a contact language.

### 3. Kukama and the Tupi-Guarani languages

Is Kukama a Tupí-Guaraní language? Or is Kukama (Brazil, Peru) a creole language? This dichotomy implies that it cannot be both, and that is perhaps a controversial idea. For some authors like Mufwene, creole languages should be classified with their lexifiers. If we adopt such a view, then Kukama would be both TG and a creole. For me, however, a creole is not just a continuation of its lexifier, but a non-genetic language, as only the lexicon (partially) has become part of the creole. Thus, there is an assumption among those creolists who share my view, that creole languages do not belong to the same language family as the one they are obviously lexically related to, like Haitian has its lexicon from French, but unlike its lexifier, it is not a Romance language (cf. Thomason & Kaufman 1988).

Let us first look at how specialists in Tupi-Guarani languages have dealt with Kukama as a member of the family. Here we ignore the wider connections within the Tupi stock and limit ourselves to the well-established Tupi-Guarani branch or subfamily.

The Tupi stock (or family) contains some 75 languages in ten branches. One of these subgroups is the Tupi-Guarani (TG) family with eight branches and some 50 languages. One of the branches of Tupi-Guarani, called Branch III, is the Tupí-Guaraní branch to which Kukama (most often) belongs, and which has between four and six languages – most of them extinct. This classification with numbered branches goes back to the great Tupi specialist Aryon Rodrigues (1984/5) and is based on phonological developments. Cheryl Jensen (1998) followed Rodrigues' classification with eight branches, of which Branch III contains the following languages:

1. Tupinambá
2. Língua Geral Amazônica (Nheengatu)
3. Língua Geral Paulista
4. Cocama, Cocamilla, Omagua

Tupinamba is the once widespread but now extinct language of coastal Brazil, that had a wide impact on the other indigenous languages of Brazil in the form of loanwords, on Brazilian toponymy, on names of flora and fauna, and on (the lexicon of) other indigenous languages. It also influenced Brazilian Portuguese, and via that, a number of other languages, including English. It is sometimes called “Old Tupi”, and it is quite well documented in sources from the 16th century onwards. Two varieties of Tupinamba became languages of wider communication in early Brazil. *Língua Geral Amazônica* was the lingua franca in the northern part of Brazil, which started out from the mouth of the Amazon River, and is now known as Nheengatu, which is spoken today as a native language by a number of Amerindian groups in northwestern Brazil (Da Cruz 2011). *Língua Geral Paulista* started out from the São Paulo region and it also moved westwards into the interior, and it is documented in historical sources (for background see Lee 2005), but it is otherwise extinct, and has no descendant languages. The fourth language grouping is Kukama with its two sister languages.

Branch III of the Tupí-Guaraní family contained the following four languages according to Rodrigues & Cabral (2012: 498):

1. Língua Geral Amazônica (Nheengatu)
2. + Língua Geral Paulista
3. + Tupí (Tupí Antigo)
4. + Tupinambá (Língua brasílica, Tupí antigo)

Tupinamba was, for unknown reasons, split into two (nrs. 3 and 4), both called Old Tupi, and note that the subgroup with the Kukama-related languages has disappeared, and cannot be found elsewhere among the Tupi or Tupi-Guarani languages. The reason can be inferred from a footnote about Kukama (p. 566), where it is stated that it is “a language which emerged from [a] contact between speakers of a conservative Tupí-Guaraní language close to Tupinambá and speakers of non-Tupí languages”. Apparently, the authors moved it out of this Amerindian group, into a group of non-genetic languages, most likely in the form of a creole. In other publications, the authors had discussed the “non-genetic classification” of Kukama (Cabral 1997) and “abrupt creolization” in Kukama (Rodrigues & Cabral 2003). In other words, they have removed Kukama from the TG family because of its non-transmission of the grammatical system, like in creoles and mixed languages. The non-genetic nature of Kukama is also discussed in Cabral (2000, 2007, 2011).

Other classifications were based on comparison of vocabulary. Rodrigues (1958) had earlier made a classification of TG based on lexical similarities. Tupinamba and Nheengatu belong to the “a” branch labelled a1, whereas Kukama and Omagua

belong to a branch b, with Kukama b1 and Omagua b2, all within TG. As no further explanation is given, we cannot assess further implications.

Lemle (1971) made a classification of ten TG languages based on a 200+ word list, and on the basis of the lexical items and the phonological differences between them, she found a solid and close connection between Tupinamba and Kukama (p. 128) within the family, which share a branch. They are connected with another branch comprising Guaraní, Guarayo/Guarayu and Siriono.

Schleicher's (1998) classification was based on phonological changes, but unfortunately Kukama was not included in his sample of Tupi-Guarani languages. Also, a more recent computational study confirmed the close lexical link between Tupinamba and Kukama. Müller, Wichmann et al. (2013) had developed a method that would be able to classify the languages of the world on the basis of the most stable 40 words from the Swadesh word list. The results show a clear branch of Kukama and Omagua, connected with Nheengatu (creolized Tupinamba). Tupinamba is not included, so this branch contains all and only Branch III languages, except for two, which are found under the Tupi family. This branch is connected with 30 other Tupi-Guarani languages, with limited replication of other branches as well (Müller et al. 2013).

Walker et al. (2012) also used the 40-word Swadesh list. They have one branch with 25 of 33 TG languages that would descend from an ancestor common to Tupinamba and Nheengatu (in one branch) and Omagua and Kukama in another. If we can rely on their dating, the most recent common ancestor of these four languages and 21 others would have been spoken some 2250 years ago (Walker et al. 2012: 2).

Another recent and quite elaborate computational classification also casts doubts on Tupinamba as the direct ancestor of Kukama. Kukama and Tupinamba may share their most recent common ancestor with half a dozen other TG languages from three other branches, beyond branch IV. In this recent study, based on 574 word meanings of 19 TG languages (Michael et al. 2015: 204), Omagua and Kukama appeared closely connected, and part of a branch "diasporic languages" with Guaranian and Southern TG languages. There are some contradictions between this classification and earlier ones (see their trees on p. 208), but Tupinamba remains Kukama's and Omagua's closest connection lexically.

Thus far, all classifications except Dietrich were based on lexical comparisons. What about grammatical properties? How extensive are the grammatical differences between Kukama-Omagua, and Tupinamba, which is the language with which Kukama is most closely related among the sufficiently documented languages? A summary of the main differences is given in Table 2. Not all of these properties are accepted today, as more data have become available about the languages. These



features are just given in order to give an impression of the considerable typological differences between Tupinamba as a representative of the closest relative of Kukama and Omagua as perceived by Cabral and summarized by Michael (2014).<sup>1</sup>

As the lexical affinity of Kukama with Tupi-Guarani, and with Tupinamba in particular, is beyond any doubt, Michael (2014) accepts and provides “compelling evidence that Kokama resulted from radical contact-induced restructuring of Tupinambá grammar, combined with the retention of the majority of the Tupinambá lexicon” (Michael 2014: 314).

This echoes some earlier attempts at classification that noted the dichotomy between the clearly TG lexical affiliation and its completely different grammatical structures. Dietrich (1990: 115) had already noted this when he remarked that “Tupinamba and Cocama do not share characteristics which allow them to be put into the same subgroup (Lemle 1971: 128; Rodrigues 1984–85: 39)”.

In a loss-retention-reconstitution scenario of creolization, Kukama shows all the symptoms of these processes: the lexicon of TG was mostly retained but it has lost almost all of its structural features, the language is perceived to be relatively isolating (but see below) and it has reconstituted grammatical properties, such as a new alignment pattern, a number of derivational affixes and a clearly distinct typological profile. Is Kukama a creole?

---

1. The following properties from this table are contested. The status of KO as isolating is doubtful. Vallejos (2016: 82, 84) considers KK an isolating language, or at least isolating compared to other Amazonian languages. Person is expressed with personal pronouns rather than in the verb, but otherwise the description contains almost no sentences with no clitics or bound morphemes, including words with four morphemes, indicating a more agglutinative typology. Vallejos (2016: xiv–xvi) lists 38 clitics indicated with “=” and 22 affixes, as indicated with “-” in her specified list of abbreviations. Michael (2014: 314) mentions Cabral’s (1995: 124–125) discussion of four Tupinamba case markers absent in Kukama. Cabral mentions a partitive, two locatives and a fourth “nominal case”, of which only the latter is found in Kukama, but in a fossilized form. Alignment in KK is summarized by O’Hagan (p.c.) as follows: “Most pronouns [in KK] can express subjects and objects, i.e., there are not distinct forms for these different syntactic relations. The third person male speech pronouns show a split intransitive system, where =*ura* may express a transitive object or a stative intransitive subject (see pp. 159–162 of Vallejos’ 2016 grammar).” For the modality markers, see O’Hagan 2011: 95ff. O’Hagan et al. (2013, cuadro 6) show that four or five nominalizers have been preserved in KO. Omagua and KK retain the privative as *-ima*, cognate with Tupinambá *-e?im* (Michael & O’Hagan 2016). There are several types of reduplication (see Vallejos 2016: 74–75 and elsewhere).

**Table 2.** Grammatical properties of Kukama-Omagua and its close relative, Tupinamba, after Michael (2014) based on Cabral (1995)

	Tupinamba	Kukama-Omagua
<i>Holistic typology</i>	Polysynthetic	Isolating
<i>Person marking</i>	Crossreference/relational prefixes	No subject marking on verb or possession on noun
<i>Morphology</i>	Prefixing person inflection. Derivation mostly suffixing but also prefixing (e.g. causative)	No person inflection. Derivation exclusively suffixing
<i>Internal complexity of words</i>	Complex, multimorphemic words	Some complex words from Tupinamba have been reanalyzed and are unanalyzable units in KO
<i>Case marking</i>	Four cases	No case-marking
<i>Alignment</i>	Active alignment with affixation	Nominative-accusative pronouns
<i>Modality markers</i>	Modal suffixes	No TG modal suffixes preserved.
<i>Causative</i>	Two causative prefixes	One causative suffix
<i>Nominalization</i>	Seven nominalizers	One or two (Vallejos)
<i>Diminutives</i>	Diminutive suffixes	None preserved (but another one was innovated)
<i>Privative</i>	Privative suffixes	None
<i>Pluractional reduplication</i>	Pluractionality reduplication	None
<i>Noun incorporation</i>	Present	Absent

#### 4. Kukama and the typology of creoles

In order to establish whether Kukama is a creole, we can compare its structural properties with those found in established creoles. Vallejos (2016) characterizes Kukama KK as a language with comparatively little morphology (“it exhibits few bound morphemes”, p. 84) and she writes “On morphology, Cabral says that Kukama lacks inflectional morphology. Twelve suffixes have been documented in the Peruvian varieties, although only one could be characterized as inflectional and three are difficult to classify as either inflectional or derivational, while the others are clearly derivational” (p. 27). In the relevant section (3.2.1), three suffixes, all of them aspectual, are considered to be “towards the inflectional end of the continuum” (p. 87).

Two propositions for structural profiles of creoles are considered here, one as formulated by McWhorter (2001, 2005) and one based on results by Bakker et al. (2013).

McWhorter (2001) claimed that three typological properties of languages would be sufficient to set all and only creoles apart from the world's non-creoles. Roughly, one could say that he characterized creoles as languages with no inflection, no grammatical tone and no irregular derivation. Reacting to criticism, and in response to potential counterexamples to the initial formulation, he modified the definition later (2011) to:

I. Absence of inflectional affixation

(“Prototypical creoles have (1) little or no inflectional affixation of any kind and (2) no unbound inflectional markers of (a) contextual inflectional categories at all, and (b) inherent inflectional categories of paradigmatic typology”; McWhorter 2011: 96)

II. Absence of tone

(“Prototypical creoles have little or no distinction of monosyllabic lexical items or morphosyntactic distinctions via tone or register, and do not have vocalic inventories of typologically unusual proliferation”; McWhorter 2011: 103)

III. Derivation

(“Prototypical creoles have little or no noncompositional combination of non-reduplicative derivational morphemes with roots”; McWhorter 2011: 107).

Are these three observations true for Kukama? As for contextual inflection (i.e. inflection that connects one constituent to another, e.g. accusative case markers linking to verbal functions, in contrast to inherent inflection, e.g. locative cases), this may be true. There is no person inflection in the verb in *κκ*, and no contextual inflection. McWhorter's prediction to a large extent fits the first one. There is no mention of tone in Vallejos, so McWhorter's second criterion fits perfectly. Going through Vallejos' (2016) grammar, it is striking that the meanings of the roots with derivational endings are very predictable, and thus Kukama fits McWhorter's third criterion as well, and thus it would be a creole language.

There is, however, a caveat. All prototypical creoles are towards the analytic end of the scale. With few exceptions (see Farquharson 2007 for some), creoles display very little morphological marking, both compared to their lexifiers and to a sample of the world's languages. Both Cabral (1995: 119) and Vallejos (2016: 84) characterize the language as isolating, and thus the language would also fit the overall typological profile of creoles. Its proposed lexifier Tupinamba is morphologically rich, especially with prefixes, but “Kokama, on the other hand is an isolating type of language. It lacks inflectional morphology, and its derivational affixes

(3 valence-changing affixes, 3 nominalizers, 1 intensive, and 1 morphological negative) are all suffixes.” (Cabral 1995: 118). And Vallejos writes: “From a typological perspective, KK can be characterized as an isolating language because words tend to be comprised of one morpheme, and morphemes can be identified with particular meanings.” (Vallejos 2016: 82).

But is Kukama really as isolating and analytic as creoles are? The following three sentences are quite representative for what we find in Vallejos, who uses the symbol “=” to indicate boundaries between enclitics and their hosts (Vallejos 2016: 229, 422, 123):

- (1) tima chipi-yara-ra, utsu=taka ni=erura ra=mia  
 NEG price-HAVE-COND go.FUT=UNCERTAIN 1PL.INCL=bring 3SG.MS=HYP  
 ‘If it were not pricey, we might have brought it.’
- (2) yawara=pura=tu=kana uri=tsuri=ai  
 dog=FOC=AUG=PL.MS come=PAST=already  
 ‘The big dogs (tigers) have come already’
- (3) yaepe=tsui ajan animaru=pura=tu=anu ipu-ka  
 there=ABL DEM animal=FOC=AUG=PL.FS sound-REI  
 ‘After that, these big animals make sounds again.’

Such sentences are not very reminiscent of what is, or what is considered, typical for creoles. We find bound or cliticized morphemes, which are supposedly uncommon in creoles. There are enclitics, for example, tense, mood and conditional, which are rare in creoles. Almost all creoles have Tense-Mood-Aspect markers immediately preceding the verb, but we don’t find anything like this preverbally in Kukama, where we have only postverbal or sentence-final clitics and sentence-initial particles. Also, constituent ordering in Kukama is quite different from that in creoles. In Kukama, depending on the presence of tense or aspect marking, one can get SVO, SOV, OSV, OVS ordering patterns, rather than a fairly fixed order as in earlier identified creoles (most often SVO, sometimes SOV or VSO). We find an inclusive/exclusive distinction in Kukama, and also evidential markers and gender, which we never find in creoles (cf. McWhorter 2018). As for gender, it must be said that the gender markers in Kukama constitute an indexical category of the sex of the speaker, and not grammatical gender or referential gender. Yet, neither type is found in creoles. A language with these dozens of enclitics, and with postnominal and postverbal modifiers (particles, clitics) rather than prenominal and preverbal markers, could it indeed be a creole?

Whether Kukama could be a creole, is a fundamental question, which goes far beyond the status of Kukama itself. The answer is connected to the possibility of identifying creoles by their structural properties alone, in other words, the existence

of a creole typological profile. And even a cursory glance at the example sentences (1–3) would indicate that Kukama is quite unlike any known creole. If it would still be possible to classify Kukama as a creole (it did pass the “McWhorter” test, as we saw), despite the huge typological differences between prototypical creoles and Kukama, we may come closer to a characterization of creoles that is more independent from the typological influences from, say Western European, West African, Austronesian and Arabic languages.

We can apply a second test for creoleness to Kukama. Earlier work by myself and others has shown that, if one uses a reasonably large set of typological features, phylogenetic software divides the languages of the world in two groups, creoles and non-creoles (see Szmrecsanyi & Kortman 2019 for English creoles; Bakker et al. 2011, 2013 for global overviews; Muysken 2015 for Surinamese creoles; and Bakker et al. 2017 for Arabic creoles, French creoles, and Iberian-lexifier creoles). Blasi et al. (2017: Figure 1) also show that creoles have a special profile by using an efficient data mining algorithm as their method, using different datasets than the previous studies. However, in all of these studies, it was not clear which of the many features (between 41 and 96 structural features) were mainly responsible for these splits. Daval-Markussen (2014) attempted to identify the minimal number of features that would still result in a split between creoles and non-creoles of the world, by comparing that minimal set of features with the data on non-creole languages available in *WALS*, the *World Atlas of Language Structures* (Haspelmath et al. 2005; Dryer et al. 2013), for which these features were known. He came up with sets of three, four and five features that either had minimal leakage (say, one of 100 non-creoles ending up among the creoles), or none at all. Surprisingly enough, almost none of the features identified, are traditionally associated with creole languages. Whereas SVO order, preverbal TMA marking and serial verbs (see e.g. Seuren 1998: 292–293) would probably be named as properties typical of creoles, the five most prominent properties (as a set, not individually) of creoles versus non-creoles appeared to be (the numbers refer to the online version of *WALS*, Dryer et al. 2013):

- 38A, value 2: There is an indefinite article derived from the numeral “one”;
- 55A, value 1: There are no numeral classifiers;
- 69A, value 5: There is no tense inflection;
- 112A, value 2: Negation is expressed with a particle (not e.g. a verbal morpheme or an auxiliary);
- 117A, value 5: There is a verb “to have” to express possession (not a construction like ‘it is to me’).

It appears that four features 38A, 69A, 112A, 117A lead to a network in a cluster with all creoles on three connected branches, but with a handful of non-creoles as well. Two features (38A and 69A) will include also some 5% of the 464 *WALS*

languages for which the features are known. Three features (38A, 69A and 55A) are enough to have all creoles on one branch, but it includes one non-creole, Rapanui, out of 154 non-creoles. In all cases, the creoles were the 18 included in Holm & Patrick (2007), with seven different lexifiers from all parts of the world, and hence a pretty good sample, but obviously not the full set of known creoles.

Bakker (2014) took features 38A, 69A, 112A, 117A and found that four or five non-European contact languages proposed to be creoles, had the same feature values as those identified for creole languages. Nativized Chinuk Wawa (North America), Hiri Motu (Papua New Guinea), Nheengatu (South America), and possibly also Plains Indian Sign Language (North America), were all included in the cluster of creoles. Yilan Creole Japanese (Taiwan) did not fit, as it has negative suffixes rather than particles. Thus, this method accurately included four of these non-European creoles with the identified creoles.

How does Kukama behave with regards to these four or five properties? Let us check what we find in Kukama with respect to these features. Negation (112A) is expressed with a particle in Kukama, just like in creoles (see also Examples (1) and (6)):

- (4) *juria tima mena-yara=tsuriai*  
 Julia NEG husband-possess=PAST  
 ‘Julia did not have a husband’ (Vallejos 2016: 339)

In Tupinamba (Barbosa Lemos 1956: 87) negation is expressed with a particle *nda-* and a verbal suffix *-i*, strongly suggesting that this particle is an innovation. O’Hagan (2011: 110) related the negative particle to Proto-TG *e?im*, suggesting a preserved form, as “*tima* precedes its negated predicate” in Kukama (p. 110), but in Omagua and Kukumiria it follows the syntactic element over which it has scope. The *t-* seems to be derived from a third person absolute marker, according to O’Hagan.

What about predicative possession? Kukama has a verb glossed as “to have” (117A), as in (1) and (2) and also (5):

- (5) *iminan=tsui=ka Kukama=kana tima pitsa-yara=tsuriai*  
 long.ago-ABL=LOC kukama=PL NEG fish.net-possess=PAST  
 ‘(From) a long time ago, the Kukama people did not use to have fishnets.’  
 (Vallejos 2016: 339)

Its possible lexifier Tupinamba has a different verb *ekó* “to have” (Barbosa Lemos 1956: 68–69), but that was not continued into Kukama. The morpheme *yara* is derived from a Tupinamba noun *\*jár-a* “owner” (Jensen 1998: 507, cited in Vallejos 2016: 335). Often, the morpheme *yara* is glossed as “owner” by Vallejos. O’Hagan (p.c.) considers *-yara* a nominal suffix and therefore not a verb. Note, however, that it is followed by a tense clitic.

Kukama does not have tense inflection (69A). Vallejos (2016: 388–399) describes tense clitics in a detailed way, and as clitics rather than inflectional markers. As they are attached to the final constituents, which are verbs or object arguments, they are not inflectional markers on the verb but clitics. In the preceding Examples (4) and (5) (and perhaps (2)), we can observe a remote past tense clitic =*tsuriai*. The tense clitics are not attached to the verb, but to the last constituent in the clause. There are also other tense clitics, as in:

- (6) *tima mari epe eyu=utsu.*  
 NEG thing 2PL eat=FUT/AND  
 ‘Nothing you will eat.’ or ‘You will go to eat nothing.’ (Vallejos 2016: 396)
- (7) *tsa tsukuta ipira itsim=ui.*  
 1SG.FS clean fish slime=PAST  
 ‘I clean/took away the slime of the fish.’ (Vallejos 2016: 232)

The future tense marker is clearly derived from the verb *utsu* ‘to go’, ultimately Tupinamba *o-só* ‘he goes’, something reminiscent of processes encountered in pidgin and creole languages (see Bakker & Van der Voort (2017) and Bakker et al. (2018) for examples of pidgins that include fossilized person markers in their verbs). In addition, there are a couple of aspectual clitics as in *yapana=ri* (run=PROGRESSIVE) ‘be running’ and *eyu=pa* (eat=COMPLETIVE) ‘to eat up’ (ultimately from Tupinamba *u-pa* ‘he/she/it finishes’) and *tsatsawa-ka* ‘we crossed (multiple persons/times)’ Tupinamba also has tense and aspectual suffixes, but they have different shapes and meanings. The *ko* aspect markers are derived from Tupinamba verbs: =*pa* from *paβ* ‘to finish’, regressive aspect =*ka* from *katu* ‘to be good’.

There is a 5-way tense distinction expressed via non-obligatory VP-final enclitics, =*ui* for immediate past, =*ikua* for intermediate past, =*tsuri* for remote past, =*utsu* for immediate future and some probability, and =*á* for remote future and less probability.

Thus, we have contradictory information. On the one hand, the grammarians claim that Kukama does not have morphology or tense inflection, or only a little (“*kk* can be characterized as an isolating language”, Vallejos 2016: 82). On morphology, Vallejos says, like Cabral, that Kukama lacks inflectional morphology. Vallejos writes: “*kk* uses order of constituents, specific grammatical words, and particles, rather than inflection” (Vallejos 2016: 82) and “tense and modality are expressed through enclitics and particles” and “most grammatical categories that are cross-linguistically encoded by inflectional morphology (e.g., number, tense, and mood) in *kk* are conveyed by means of categorically unrestricted elements (optionally attachable morphemes) and syntactic strategies (auxiliaries).” (Vallejos 2016: 609). The optionality and the clearly traceable, and hence recent cliticization could be the reason for these seemingly contradictory statements.

It is also a bit more difficult to establish whether Kukama has an indefinite article or not (38A). Vallejos (2016: 113) contains a brief discussion of article-like use of demonstratives in Kukama. There are indeed a couple of examples in which the numeral “one” (*wepe*) is used in a way that reveals a function that has nothing to do with a numeral, but it rather behaves like an indefinite article or a marker of “specific unknown”:

- (8) *raepe rana umi [ikian wepe yatiri westa yauki-ari]-n kai=kana.*  
 there 3PL.MS see this one group party make-PROG-NZR monkey-PL  
 ‘There they see a group of party-making monkeys.’ (Vallejos 2016: 200)
- (9) *tsa mimira-ra-ka=tsuri, wepe napitsara.*  
 1SG.FS SON.FEMALE.EGO-VZR-REI=PAST one man  
 ‘I had a baby again, a boy.’ (Vallejos 2016: 207)

Nevertheless, in most cases where Spanish or English would use an indefinite article, no *wepe* is present. Less documentation is available for Omagua, but on the basis of the examples in O’Hagan (2011), one can conclude that *wipi* “one” has developed into an indefinite article e.g. ex. 3.2., 3.8., 3.9, 4.3, 8.1, whereas it is not yet found in the Old Omagua catechism of the mid-1750s.

Finally, there are no numeral classifiers in Kukama (55A). Thus, Kukama clearly has the same value as creoles for three features (55A, 112A, 117A). The negative particle and the verb “to have” were grammaticalized (from a negative directive and a noun “owner” respectively), not inherited. As for 38A, one could say that there is an incipient indefinite article, but not a frequent one as found in creoles, and that it became an indefinite article in Omagua. As for tense marking (69A), if one has a watertight distinction between inflectional clitics and inflectional affixes, one can take an informed decision. On the other hand, one could infer that the tense clitics are recent grammaticalizations, and then they would have been particles until recently. In fact, Faust (1972), in her grammar of Peruvian Kukama, considered the enclitic future *-utsu* to be a particle (e.g. p. 132), but the past tense marker *tsuri(a)* a suffix (p. 159).

## 5. Conclusion

If we consider Kukama from a viewpoint of retention, loss and reconstitution, then we have to conclude that Kukama is a creole. The lexicon is overwhelmingly from TG/Tupinamba, and the development of the grammatical system is clearly recent, i.e. mostly reconstructable and grammaticalized into the language in the past few centuries or so, and overwhelmingly traceable to Tupinamba lexical morphemes, parallel to what one observes in creoles.



If we take McWhorter's diagnostic for creoles as a point of departure, Kukama is probably a creole. If we apply Daval-Markussen's five-feature test, Kukama has all of the features like creoles, if one accepts *yara* "owner" as having become a verb "to have", as it is glossed regularly in Vallejos (2016). On the other hand, Kukama has many properties as yet unheard of in creoles.

The inclusion of Kukama among the set of creoles would open the door to a broader typological range for creole structural properties than hitherto identified. Non-European languages like Nagamese (Northeast India), Nubi (Africa) and Chinuk Wawa do pass the aforementioned creole tests, like Kukama. The deviant properties, such as the many clitics, could be due to the antiquity of the process, dating back perhaps 1500 years, a new wave of grammaticalization independent of the creolization process, and/or areal influence from neighbouring languages.

Thus, all three sets of criteria point to a creolization process in the history of *ko*. Not enough is known about the social history, but we can point to fact that *ko* is spoken geographically far removed from its lexifier (Eriksen & Galucio 2014), similar to what we observe with creoles. The linguistic result is different from more prototypical creoles (cf. McWhorter 2005), in that the languages of its creators in the contact situation are typologically quite different from other ecologies, and the contact that led to creolization could have taken place centuries earlier than in the case of more established cases of creolization.

From the side of creole studies, we must also move away from the idea that creoles are and remain analytic languages. Creoles are the only languages where separate words of one consonant length are accepted: they are considered to be TMA particles, personal pronouns or possessive pronouns in creoles like Sranan, Cape Verdean and Haitian. These meaningful elements would be considered to be bound morphemes by all people who have described non-creoles. This would move at least some creoles considerably up in morphological complexity, and more resemble a language like Kukama. The lack or rarity of irregularities in creoles and in Kukama definitely points to the fact that they are young languages from the viewpoint of their grammar. Previously identified creoles have a time depth in the range of 150 years to five centuries, and *ko* perhaps a few centuries more.

## Acknowledgements

I am indebted to Zachary O'Hagan and the anonymous reviewers who spotted errors, gaps and misinterpretations in an earlier version. They bear no responsibility for the result. I am also grateful to the editors for their confidence in a publishable version. Finally, I like to thank Stéphane Goyette, Bart Jacobs, John McWhorter and Mikael Parkvall for the email discussion we had about the creole status of Kukama in 2013 and 2014, and audiences at my presentations on Kukama in Bremen and Washington D.C. in 2015 and 2016.

## Abbreviations used

Abbreviations used in the text:

KK	Kukama and Kukamiria
KO	Kukama/Kukamiria and Omagua
TG	Tupi-Guaraní branch within the Tupi language family.

Abbreviations used in glosses:

ABL	ablative	LOC	locative
AND	andative	MS	male speaker
AUG	augmentative	NEG	negative particle
COND	conditional subordinator	NZR	nominalizer
DEM	demonstrative	PAS(T)	past tense
FOC	focus	PL	plural
FS	female speaker	PROG	progressive
FUT	future	REI	reiterative aspect
HYP	hypothetical modality	VZR	verbalizer.
INCL	inclusive		

## References

- Bakker, P. 1997. "A Language of our Own". *The Genesis of Michif – The Mixed Cree-French Language of the Canadian Métis*. Oxford: OUP.
- Bakker, P. 2014. Creoles and typology: Problems of sampling and definition. *Journal of Pidgin and Creole Languages* 29(2): 437–455. <https://doi.org/10.1075/jpcl.29.2.09bak>
- Bakker, P. 2017. Typology of mixed languages. In *The Cambridge Handbook of Linguistic Typology*, A. Y. Aikhenvald & R. M. W. Dixon (eds), 217–253. Cambridge: CUP. <https://doi.org/10.1017/9781316135716.008>
- Bakker, P. 2020. Contact and mixed languages. In *The Handbook of Language Contact*, R. Hickey (ed.), 201–220. Oxford: Wiley & Sons.
- Bakker, P. & Muysken, P. C. 1994. Mixed languages. In *Pidgins and Creoles. An Introduction* [Creole Language Library 15], J. T. G. Arends, P. C. Muysken & N. S. H. Smith (eds), 41–52. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15.08bak>
- Bakker, P. & van der Voort, H. 2017. Polysynthesis and language contact. In *The Oxford Handbook of Polysynthesis*, M. Fortescue, M. Mithun & N. Evans (eds), 408–427. Oxford: OUP. <https://doi.org/10.1093/oxfordhb/9780199683208.013.23>
- Bakker, P., Daval-Markussen, A., Parkvall, M. & Plag, I. 2013. Creoles are typologically distinct from non-creoles. In *Creole Languages and Linguistic Typology*, [Benjamins Current Topics 57] P. Bhatt & T. Veenstra (eds), 9–45. Amsterdam: John Benjamins. <https://doi.org/10.1075/bct.57.02bak> (Published originally in *Journal of Pidgin and Creole Languages* 26(1): 5–42, 2011. <https://doi.org/10.1075/jpcl.26.1.02bak>)
- Bakker, P., Borchsenius, F., Levisen, C., & Sippola, E. (eds). 2017. *Creole Studies – Phylogenetic Approaches*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.211>

- Bakker, P., Gretenkort, T., Parkvall, M. 2018. Dr. Crevaux' Wayana-Carib Pidgin of the Guyanas: A grammatical sketch. *Amérindia: Revue d'Ethnolinguistique Amérindienne* 40: 169–220.
- Barbosa Lemos, A. 1956. *Curso do tupi antigo: Gramática, exercícios, textos*. Rio de Janeiro: Livraria São José.
- Blasi, D. E., Michaelis, S. M. & Haspelmath, M. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1: 723–729.
- Cabral, A. S. de A. C. 1995. Contact-induced Language Change in the Western Amazon: The Non-genetic Origin of the Kokama Language. PhD dissertation, University of Pittsburgh.
- Cabral, A. S. de A. C. 1997. Evidências morfológicas para a não-classificação genética do Kokama. *Boletim da ABRALIN* 21: 152–164.
- Cabral, A. S. de A. C. 2000. En qué sentido el kokáma no es una lengua tupí-guaraní. In *Actas I Congreso de lenguas indígenas de Sudamérica*, II, L. Miranda (ed.), 237–251. Lima: Universidad Ricardo Palma.
- Cabral, A. S. de A. C. 2007. New observations on the structure of Kokáma/Omágwa. In *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area*, L. W. Wetzels, (ed.), 365–379. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), Universiteit Leiden.
- Cabral, A. S. de A. C. 2011. Different histories, different results: The origin and development of two Amazonian languages. *Papia* 21: 9–22.
- Clements, J. C. 2019. Speech communities, language varieties, and typology: Acquisitional history and pidgin and creole typology. *Journal of Pidgin and Creole Languages* 34(2): 377–390. <https://doi.org/10.1075/jpcl.00043.cle>
- Da Cruz, A. 2011. Fonologia e gramática do Nheengatú: A língua geral falada pelos povos Baré, Warekena e Baniwa. PhD dissertation, Vrije Universiteit Amsterdam.
- Daval-Markussen, A. 2014. First steps towards a typological profile of creoles. *Acta Linguistica Hafniensia* 45(2): 274–295. <https://doi.org/10.1080/03740463.2014.880606>
- Dietrich, W. 1990. *More Evidence for an Internal Classification of Tupí-Guaraní Languages*. Berlin: Gebrüder Mann.
- Dryer, M. & Haspelmath, M. (eds). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Eriksen, L. & Galucio, A. V. 2014. The Tupian expansion. In *The Native languages of South America: Origins, Development, Typology*, P. C. Muysken & L. M. O'Connor (eds), 177–199. Cambridge: CUP.
- Farquharson, J. T. 2007. Creole morphology revisited. In *Deconstructing Creole* [Typological Studies in Language 73], U. Ansaldo, S. Matthews & L. Lim (eds), 21–37. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.73>
- Faust, N. W. 1972. *Gramática cocama: Lecciones para el aprendizaje del idioma cocama* [Lingüística Peruana 6]. Yarinacocha: Ministerio de Educación – Instituto Lingüístico de Verano.
- Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B. (eds). 2005. *The World Atlas of Language Structures*. Oxford: OUP.
- Holm, J. A. & Patrick, P. L. (eds). 2007. *Comparative Creole Syntax*. London: Battlebridge.
- Jensen, C. 1998. Comparative Tupí-Guaraní morphosyntax. In *Handbook of Amazonian Languages*, Vol. 4, D. Derbyshire & G. Pullum (eds), 489–618. Berlin: Mouton de Gruyter.
- Lee, M. K. 2005. *Conversing in Colony: The Brasília and the Vulgar in Portuguese America, 1500–1759*. PhD dissertation, Johns Hopkins University.

- Lemle, M. 1971. Internal classification of the Tupi-Guarani linguistic family. In *Tupi Studies I*, D. Bendor-Samuel (ed.), 107–129. Norman OK: Summer Institute of Linguistics.
- McWhorter, J. H. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74(4): 788–818. <https://doi.org/10.2307/417003>
- McWhorter, John. 2001. The world's simplest grammars are Creole grammars. *Linguistic Typology* 5(2): 125–66. <https://doi.org/10.1515/lity.2001.001>
- McWhorter, J. H. 2005. *Defining Creole*. Oxford: OUP.
- McWhorter, J. H. 2011. Tying up loose ends: The creole prototype after all. *Diachronica* 28: 82–117. <https://doi.org/10.1075/dia.28.1.04mcw>
- McWhorter, J. H. 2018. *The Creole Debate*. Cambridge: CUP.
- Meakins, F. 2016. Mixed languages. In M. Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: OUP. <https://doi.org/10.1093/acrefore/9780199384655.013.151>
- Michael, L. 2014. On the pre-Columbian origin of Proto-Omagua-Kokama. *Journal of Language Contact* 7: 309–344. <https://doi.org/10.1163/19552629-00702004>
- Michael, L. & O'Hagan, Z. 2016. A linguistic analysis of Old Omagua ecclesiastical texts. *Cadernos de Etnolingüística. Cadernos de Etnolingüística: Série Monografias* 4.
- Michael, L., Chousou-Polydouri, N., Bartolomei, K., Donnelly, E., Wauters, V., Meira, S., & O'Hagan, Z. 2015. A Bayesian phylogenetic classification of Tupi-Guarani. *LIAMES* 15(2): 193–221.
- Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Holman, E. W., Sauppe, S., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, U., Mailhammer, R., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A. P. & Valenzuela, P. 2013. *ASJP World Language Trees of Lexical Similarity: Version 4* (October 2013).
- Muysken, P. C. 2012. Contacts between indigenous languages in South America. In *The Indigenous Languages of South America: A Comprehensive Guide*, L. Campbell & V. Grondona (eds), 235–258. Berlin: De Gruyter.
- Muysken, P. C. 2015. Conclusion: Feature distribution in the West Africa-Surinam Trans-Atlantic Sprachbund. In *Surviving the Middle Passage. The West Africa-Surinam Sprachbund*, P. C. Muysken & N. S. H. Smith, in collaboration with R. D. Borges (eds), 393–408. Berlin: De Gruyter.
- O'Hagan, Z. 2011. Proto-Omagua-Kokama: Grammatical Sketch and Prehistory. BA thesis, University of California, Berkeley. <<http://escholarship.org/uc/item/1vx978rq>> (24 April 2020).
- O'Hagan, Z., Michael, L. & Vallejos, R. 2013. Hacia la reconstrucción morfológica del proto-omagua-kokama. Conference on Indigenous languages of Latin America VI, October 25, Austin. <[http://linguistics.berkeley.edu/~zjohagan/pdfflinks/ohagan&michael&vallejos\\_cilla\\_vi\\_pok\\_morphology\\_v8.pdf](http://linguistics.berkeley.edu/~zjohagan/pdfflinks/ohagan&michael&vallejos_cilla_vi_pok_morphology_v8.pdf)> (24 April 2020).
- Rodrigues, A. D. 1958. Classification of Tupi-Guarani. *International Journal of American Linguistics* 24(3): 231–234.
- Rodrigues, A. D. 1984/5. Relações internas na família lingüística Tupi-Guarani. *Revista de Antropologia* 27–28: 33–53.
- Rodrigues, A. D. & Cabral, A. S. de A. C. 2003. Evidências de crioulização abrupta em Kokáma? *PAPIA: Revista Brasileira de Estudos Crioulos e Similares* 13: 180–186.
- Rodrigues, A. D. & Cabral, A. S. de A. C. 2012. Tupian. In *The Indigenous Languages of South America: A Comprehensive Guide*, L. Campbell & V. Grondona (eds), 495–574. Berlin: De Gruyter.
- Schleicher, C. O. 1998. Comparative and Internal Reconstruction of Proto-Tupi-Guarani. PhD dissertation, University of Wisconsin, Madison.

- Seuren, P. A. M. 1998. *Western Linguistics. An Historical Introduction*. Oxford: Blackwell.
- Szmrecsanyi, B. & Kortmann, B. 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119: 1643–1663.  
<https://doi.org/10.1016/j.lingua.2007.09.016>
- Thomason, S. G. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- Vallejos, R. 2016. *A Grammar of Kukama-Kukamiria: A Language from the Amazon*. Leiden: Brill.
- Vallejos, R. & Murayari, R. M. 2015. *Diccionario Kukama-Kukamiria Castellano*. Lima: Programa de Formación de Maestros Bilingües de la Amazonía Peruana-FORMABIAP y AIDSESEP.
- Walker, R. S., Wichmann, S., Mailund, T. & Atkisson, C. J. 2012. Cultural phylogenetics of the Tupi language family in Lowland South America. *PLoS ONE* 7(4): e35025.  
<https://doi.org/10.1371/journal.pone.0035025>

# Are creoles a special type of language?

## Methodological issues in new approaches to an old question

Silvia Kouwenberg and John Victor Singler

University of the West Indies / New York University

This paper takes as its starting point Muysken's (1988a: 300) view of a 'creole' language as "just a language." With this statement, Muysken rejects the idea of a creole prototype. Over the past 20 years, that idea has seen several new proponents. We provide a brief overview, before turning our attention to the work of Bakker et al. (2011; *et seq.*). That work relies on statistical and phylogenetic computational procedures which are not well understood by many linguists in the field of pidgin and creole studies. As a result, Bakker et al.'s methodologies have gone virtually unchallenged, although several critiques of their data – and by implication, of their conclusions – have been published. In this paper, we focus on the methodologies, providing detailed explanations of the statistical and phylogenetic computational procedures. We show that the data set used in the statistical procedure, a multiple regression analysis, fails to come anywhere near the minimum number of datapoints required for such an analysis to be meaningful. We show further that the apparent sophistication of the phylogenetic approach, which relies on a combination of computational packages to produce output in the form of reticulate networks, cannot remedy the flaws arising from questionable assumptions about linguistic features, pervasive errors in the data which are fed into the computation (including errors which appear to be systematic), the problematic treatment of gaps in the data, and the overinterpretation of output patterns. Contrary to Bakker et al.'s claims, Muysken's (1988a) statement has not been disproven by these technologies.

**Keywords:** creole typology, creole exceptionalism, phylogenetics, statistical analysis, reliability of data

## 1. Introduction

Notions of *aliqueness*, *exceptionalism*, and *simplicity* have figured prominently and controversially in discussions of the place of creole languages<sup>1</sup> in linguistic typology since the nineteenth century, when these concepts were considered naturally allied: outside of then-standard assumptions about the genetic relatedness of languages, creoles were accorded exceptional status, while simultaneously being dismissed as simple compared to their lexifiers, and alike in their simplicity. These views were challenged in various ways in the course of the second half of the twentieth century, and especially during the 1980s when the field of creole studies can be said to have come into its own. As one of the emergent field's leading thinkers, Pieter Muysken engaged with the issue of creole languages' exceptionality. Thus, his (1988a) article 'Are creoles a special type of language?' addresses the question of its title from the perspectives of alikeness, simpleness, and mixedness, to conclude that "[t]he very notion of a 'creole' language from the linguistic point of view tends to disappear if one looks closely; what we have is just a language" (300).

Despite the 'creole language as just a language' view, Muysken has argued that creole languages come about in circumstances involving contact between speakers of different languages which differentiate them from other language contact outcomes, such as language shift, code-switching, borrowing, and mixed languages (e.g., Muysken 1988b, 2008) – exceptional circumstances, therefore. In other words, Muysken's work exemplifies a position which allows for exceptionalism in the historical context of creole languages' emergence, while he does not support *a priori* assumptions about their alikeness and simplicity.

A different answer to Muysken's question has emerged in some of the creolist work of the past two or three decades, in which the notions *aliqueness*, *exceptionalism*, and *simplicity* have once again become entangled. In the work of McWhorter (1998, 2001), Parkvall (2008), and Bakker et al. (2011, et seq.),<sup>2</sup> creoles are accorded typologically exceptional status. McWhorter posits that creoles are simpler than other languages; similarly, Parkvall places creoles at the lower end of the complexity scale that he constructs. Bakker et al. argue for the alikeness and exceptionalism

---

1. We use the term *creole* to refer to varieties traditionally designated creoles as well as expanded pidgins, on the basis that the two are neither structurally nor functionally distinct. (See Kouwenberg & Singler 2011: 283–286 for detailed discussion.) Bakker et al. (2011, et seq) use *creole* similarly, designating only a handful of more restricted languages as *pidgin*.

2. Also included in the body of work which we here refer to as Bakker et al. are single-authored and jointly authored publications by the participating authors Daval-Markussen and Bakker, up to the most recent publication Daval-Markussen and Bakker (2017).

of creole languages but, unlike McWhorter and Parkvall,<sup>3</sup> they do not consider simplicity to be the basis for the special status of creoles.

In the following, we will briefly survey these approaches before focusing on the work of Bakker et al., who, in 2011, et seq., argue that the application of new and cutting-edge technology can finally settle the question of creole languages' exceptional status among the world's languages. The approaches they use are indeed new to creole typology: statistical analysis and especially phylogenetic computation are not among the tools of the average student of these languages. Perhaps because of this, this body of work has been challenged predominantly for its reliance on flawed data. However, other problems become apparent when the methodologies are probed, as we will do.

## 2. The resurgence of creole exceptionalism

No longer encumbered by their predecessors' racist biases towards creole languages and their speakers, some modern-day creolists have re-engaged with the ideas of creole exceptionalism and grammatical simplicity. Here, we will briefly consider three prominent proponents.

John McWhorter focuses on morphophonological form and assumes that surface marking of paradigmatic contrasts points towards complexity. His claim is that any language which combines the three characteristics of minimal usage of inflectional morphology, little or no use of tone encoding lexical or syntactic contrasts, and generally semantically transparent derivation, is 'creole' (1998). He argues that this combination of properties results from the recent emergence of creole languages: their youthfulness means that morphophonological opacity and complexity has not yet accrued to the extent seen in older languages (2001). The upshot, as expressed in the title of his 2001 work, is that "the world's simplest grammars are creole grammars." McWhorter's claims have been contested on several grounds. Several scholars have challenged the empirical legitimacy of the three criteria that McWhorter advances for his creole prototype, with counter-examples to his statements about tone in creole languages readily available (e.g., Good 2004 on Saramaccan; Kouwenberg 2004 on Papiamentu). A more fundamental objection with regard to McWhorter's treatment of tone is that, as Ansaldo and Matthews (2001) point out, he "assumes that the acquisition of tone is somehow a complex task for learners, a point that needs to be demonstrated" (317). McWhorter's claim that creole grammars are especially simple rests on the measures of complexity

---

3. Note that the "et al." of Bakker et al. includes Parkvall. On the question of creole simplicity, then, Parkvall is part of the (2011, et seq.) team that rejects the arguments put forward in Parkvall (2008).



that he puts forward, but Kusters and Muysken (2001), for example, find fault with the stated notion of complexity as being “too vague and ...intended to cover too much of a grammar to be of much use. It does not tell whether a language will be difficult for an L1 learner or an L2 learner, neither does it evaluate language on anything more than an intuitive level” (185). The essays in Ansaldo, Matthews & Lim (2007) set forth the problems with McWhorter’s framework from a wide range of perspectives.

Parkvall (2008) reconsiders the idea of the creole prototype with the help of features drawn from the then-versions of the online typological databases *World Atlas of Language Structures* (WALS, Dryer & Haspelmath 2013) and *Atlas of Pidgin and Creole Language Structures* (APiCS, Michaelis et al. 2013), which he uses as a measure of a language’s place on a complexity scale. Given the nature of the features he uses, Parkvall’s complexity scale largely encodes the extent to which a language relies on morphological expression. He concludes that creole languages are not randomly distributed over this complexity scale, but instead cluster at the ‘simple’ end, where they are alike and are different from other languages of the world in their lack of morphological treatment of a range of grammatical distinctions. Kouwenberg (2010) presents a detailed critique of Parkvall’s notion of complexity and the problems contained in assessing complexity on the basis of his feature set. She also critiques his scoring of individual languages, and provides a feature-by-feature reassessment of Parkvall’s scoring of Papiamentu, demonstrating pervasive error.

Bakker, Daval-Markussen, Parkvall, and Plag (2011) follow Parkvall (2008) in making use of typological databases, specifically Holm & Patrick (2007), as well as Parkvall’s WALS/APiCS database.<sup>4</sup> While they follow Parkvall in asserting that “creoles as a group all belong to the languages in the less complex part of the spectrum” (p. 8), they appear to depart from him – and from McWhorter – in taking creole simplicity as a basis for the claim of creole likeness and creole exceptionalism.

Like Parkvall (2008), Bakker et al. (2011) take the position that the features used in these databases represent objective measures of linguistic types – thus inheriting from Parkvall (2008) the problems of bias detailed in Kouwenberg (2010). What is new is their use of computational programmes which have gained some traction in the field; while mainly adopted to facilitate new approaches to linguistic phylogeny, computations have also been applied to questions of typology, which is how Bakker et al. use them. It is on this basis that they claim to demonstrate conclusively “that creoles form an identifiable and distinct subgroup among the world’s languages” (2011: 9). This forms the clearest challenge to Muysken’s “just a language” view to date, and it constitutes the focus of the remainder of this article.

---

4. They make use of a third data set as well, that of Hancock (1987). We address their use of this data set in 5.1 below.

### 3. Languages, features, and ‘creoleness’ in Comparative Creole Syntax

Bakker et al. (2011) draw on Holm and Patrick (2007) to argue, using statistical analysis, that the presence of creole features in a language is independent of its lexifier, its geographical location, or the manner of its emergence – in other words: it is entirely due to its being ‘creole’. They then turn to phylogenetic computational software, which is applied not only to the Holm and Patrick data, but to the larger database derived from Parkvall (2008). We discuss these applications in turn.

#### 3.1 The creoleness of features in Comparative Creole Syntax

Holm & Patrick’s (2007) edited volume is entitled *Comparative Creole Syntax: Parallel Outlines of 18 Creole Grammars* (henceforth, CCS). The languages included in CCS are those of Table 1. Scholars knowledgeable about individual creoles responded to a template of 97 morphosyntactic and syntactic features. While two-thirds of the creoles are Atlantic, Holm and Patrick’s survey is notable for having creoles with seven different lexifier languages, including Arabic and Assamese.

According to Holm (2007: vi): “The syntactic features chosen for examination are generally those which distinguish the Atlantic Creoles (those of the Caribbean and West Africa) from their lexical source languages.” We shall see below that this has ramifications for Bakker et al.’s treatment of the CCS features as ‘creole’.

Based on authors’ responses to the 97 features, the editors provided summaries of the occurrence of the features in each creole; a feature considered to have been attested is marked as +, absent as 0, occurring only rarely as R, and where information is lacking, ‘?’ is used. (*ibid.*, p. xi).<sup>5</sup> In using the CCS data, Bakker et al. state: “In order to reduce the distinctions to binary oppositions, we merged <R> with the category of <+>. Question marks were maintained” (p. 16). “Maintaining” the question marks meant that they remained outside the binary opposition (Aymeric Daval-Markussen p.c).

---

5. Patrick observes that reducing languages into 97 occurrences of +/R or 0 “inevitably raised difficulties and required compromises” (2007: xi). He adds: “Though authors held ultimate responsibility for chapter texts, the editors took responsibility for making sure that summary ratings were awarded in the same way across the volume” (pp. xi–xii). However, because feature ratings are included in the chapters, readers have assumed, quite reasonably, that the authors were responsible for them. For instance, Bakker (2014a), refers to “the scores provided by the specialist authors” and claims to have “relied on the native speaker views and expert insights of the authors” in adopting the CCS scoring values (p. 440). The situation is discomfiting for authors who disagree with the ratings given to features in their chapter.

**Table 1.** The 18 creoles in Holm and Patrick (2007), from Bakker et al. (2011: 22). Attributes (notably “Age” and “Type”) provided by Bakker et al. The “feature” columns are explained below

Creole	Lexifier	Area	Age	Type	CCS features (out of 97)	Creole features (out of 69)
Angolar	Portuguese	Atlantic	1550	Maroon	65	63
Berbice Dutch	Dutch	Atlantic	1600	Plantation	72	57
Cape Verdean	Portuguese	Atlantic	1500	Trade	71	62
Dominican	French	Atlantic	1700	Plantation	67	55
Guinea Bissau	Portuguese	Atlantic	1500	Fort	73	63
Haitian	French	Atlantic	1600	Plantation	73	55
Jamaican	English	Atlantic	1650	Plantation	73	63
Korlai	Portuguese	Indian Ocean	1500	Fort	54	48
Krio	English	Atlantic	1780	Plantation	81	59
Nagamese	Assamese	Indian Ocean	1800	Trade	57	48
Ndyuka	English	Atlantic	1700	Maroon	77	60
Negerhollands	Dutch	Atlantic	1650	Plantation	66	59
Nubi	Arabic	African	1850	Trade	54	55
Palenquero	Spanish	Atlantic	1600	Maroon	53	55
Papiamentu	Portuguese	Atlantic	1600	Trade	63	58
Seychellois	French	Indian Ocean	1770	Plantation	71	60
Tok Pisin	English	Pacific	1850	Trade	67	58
Zamboangueño	Spanish	Pacific	1800	Trade	53	51

The “CCS features” column in Table 1 represents the number of instances (out of 97) where a given creole was judged to have displayed the CCS feature, i.e., was given a value of either + or R for it by the editors. Bakker et al. assert: “The degree of creoleness of the languages in the CCS sample is gauged by a rather simple measure, namely the number of typical creole features present in each of the 18 languages in the sample” (2011: 17). Nevertheless, while seventy-six of the 97 features show up in half or more of the creoles, in the remaining 21 cases, a majority of the creoles in the sample do not display the feature, suggesting that not all features can be considered “typical” of creole languages. However, the bigger issue is that, by turning the CCS feature list into a set of 97 binary values (with each of the 97 utterly equivalent in weight), Bakker et al. are imposing on the set of CCS feature ratings a burden it was not constructed to support. A look at two excerpts from the feature list makes this clear.

The first excerpt is of Features 15–9 (prenominal adjective) and 15–10 (postnominal adjective). According to Bakker et al.'s use of the ratings for the features dealing with Adj-N order, a language that has both prenominal and postnominal adjectives (+ for both features) shows greater creoleness than a language that just has one of the two and, further, a language that has one shows greater creoleness than a language that does not have NP-internal adjectives. This illustrates the “roll call” nature of Bakker et al.'s use of the CCS list.

The second example involves pronominal case marking (Features 17.1–17.6 for 1st–3rd person sg and pl). Apart from the question as to why pronominal case marking should count for so much (6/97, or 6.2% of the cells), an issue arises from the interpretation of the feature values. It was constructed by Holm and Patrick with the expectation that not having the feature is the creole characteristic. Thus, a + signals what a creole is expected not to do. Since Bakker et al. treat all + values as equivalent, languages that do not display case marking, e.g. Haitian Creole, Jamaican Patwa, Angolar, show up as maximally non-creole (0/6). In contrast, languages that display case-marking across the board are scored as maximally creole (6/6).

To be sure, Bakker et al. do recognize some of the problems with a simple use of the feature values of the CCS. Consequently, they adjust the scores in two ways, yielding a new configuration that they refer to as “creole features.”<sup>6</sup> First, if six or fewer creoles in the sample of 18 showed a particular feature, they take it that it was the absence of the feature, not its presence, that typified creoles. To ensure that all the pluses signaled creole behavior, the values for any features of this type were reversed. For pronoun case, we consider the reversal appropriate. However, in other instances, the “flipping” of the scores is counterproductive. Consider the following three features for the twelve CCS Atlantic creoles:

**Table 2.** Values for selected features in the 12 Atlantic creoles in CCS

	8.3 Use of ‘for’ as a (quasi-) modal	12.2 Preverbal markers before nouns	12.6 Comparison with ‘pass’	Total
Angolar	–	–	–	0
Berbice Dutch	+	–	–	1
Cape Verdean	–	–	–	0
Guinea Bissau	–	–	–	0
Dominican	+	+	+	3
Haitian	+	+	+	3
Jamaican	+	–	–	1

(continued)

6. In the terminology of Bakker et al., “CCS features” refers to the values assigned in Holm and Patrick, while “creole features” refers to the adjusted version described in the text.

Table 2. (continued)

	8.3 Use of 'for' as a (quasi-) modal	12.2 Preverbal markers before nouns	12.6 Comparison with 'pass'	Total
Krio	+	–	+	2
Ndyuka	–	+	+	2
Negerhollands	+	–	–	1
Palenquero	–	–	–	0
Papiamentu	–	–	–	0
<b>Total +</b>	<b>6 (50%)</b>	<b>3 (25%)</b>	<b>4 (33%)</b>	

The Dominican and Haitian Creoles show all three of the features, and Krio and Ndyuka show two. These are arguably creole features but ones that are limited to 'basilectal' creoles (cf. Winford 2008). In Bakker et al.'s adjusted scoring system, the creoles that lack these features get plusses, while the creoles who do show the features lose theirs. If basilectal creoles are seen as the most fully creole, then Bakker et al.'s scoring represents a reversal of what is appropriate.

When the number of creoles showing a particular feature falls in the range from seven to eleven, Bakker et al. remove the feature. There are 28 such features, 29% of the total number. Just like the indiscriminate flipping described above, this removal of features with indeterminate scores is deeply problematic. The greatest problem in this regard involves Atlantic creoles. If a feature is found in most but not all Atlantic creoles but found in few if any non-Atlantic creoles, then Bakker et al. have removed it. The set of eleven features in Table 3a and the scores for those eleven features by region and lexifier language in Table 3b demonstrate the problem.

**Table 3a.** Eleven of the 28 features whose CCS score ranged from 7 to 11, and were therefore removed from the set of "creole features"

Feature	Description
3.4	Progressive with adjective = inchoative
8.4	'For' introducing a tensed clause
12.1	Preverbal markers before adjectives
12.4	Predicate clefting: adjectives or adjectival verbs
13.1	Equative copula
14.4	Serial 'say' meaning 'that'
14.5	Serial 'pass' meaning 'more than'
15.3	Definite article
16.6	Possessive pronouns as emphatic possessive adjective
18.2	'And' joining sentence parts: different from clause-internal 'and'
20.2	Sentence-final -o

**Table 3b.** Number of the features in Table 3a that show up in given creoles, by region and lexifier language

Region	Lexifier	Language	n (out of 11)
Atlantic	English	Krio	10
		Ndyuka	10
		Jamaican	9
	French	Dominican	8
		Haitian	8
	Dutch	Negerhollands	8
		Berbice Dutch	7
	Portuguese	Angolar	7
		Guinea Bissau	7
		Cape Verdean	5
		Papiamentu	5
Spanish	Palenquero	1	
African	Arabic	Nubi	2
Indian Ocean	French	Seychellois	7
	Portuguese	Korlai	3
	Assamese	Nagamese	1
Pacific	English	Tok Pisin	4
	Spanish	Zamboanguño	0

Although the eleven features in Table 3a are but a subset of all features, the results in Table 3b show a sensitivity within the Atlantic region to the European lexifier (English-lexifier > French-lexifier > Dutch-lexifier > Portuguese-lexifier), with the Spanish-lexifier creole in the region not patterning with the other Atlantic creoles.<sup>7</sup> Bakker et al.'s Indian Ocean group is not a coherent entity. Seychellois is an "Atlantic" creole in the sense that its lexifier is Western European and its primary substrate languages are Niger-Congo. Nagamese is spoken – at its closest – 1900 km from the Indian Ocean. The disparate scores for the three Indian Ocean languages in Table 3b reflect the disparateness of the group.

Quite apart from the inappropriateness of using Holm and Patrick's set of 97 features as the basis for quantitative analysis and quite apart from the adjustments to the set that obscure rather than heighten differences between creoles, there is the remarkable misreporting of the CCS results in subsequent work by Bakker and his associates. Bakker (2014a: 450) cites Daval-Markussen (2013) as saying that there are

7. The bias in CCS features towards Atlantic creoles is not accidental. Recall Holm's explanation that "[t]he syntactic features chosen for examination are generally those which distinguish the Atlantic Creoles (those of the Caribbean and West Africa) from their lexical source languages" (2007: vi).

no exceptions among the CCS creoles to the statement “no tense-aspect inflection.” The chapters in CCS show that Berbice Dutch (p. 27), Cape Verdean (55), Korlai (155), Nagamese (239), and Palenquero (277) – fully five of CCS’s eighteen languages – display tense-aspect inflection. (1) from Cape Verdean illustrates this point.

(1) Cape Verdean

*To ki bu ciga, tudu djenti dja ba-ba.*  
when 2s arrive, all people already go-ANT

‘When you arrived, all the people had already gone.’

(Example 9a in Baptista, Mello & Suzuki 2007: 55)

### 3.2 A multiple regression analysis

Bakker et al. (2011) tested to see which sociohistorical factors might favor or disfavor the degree of creoleness of particular languages among the 18 found in CCS. To do so, they “carried out a multiple linear regression analysis with the number of features as the dependent variable and lexifier, area, type and age as predictor variables” (23). They used the values presented in Table 1 above. They carried this out twice, once with the 97 CCS features, i.e., the presence (+/R) and absence marks (0) derived from Holm and Patrick, and once with the adjustments described above, i.e. “flipping” the values when the score was between 0 and 6 for a feature, and removing the feature when the score ranged from 7 to 11 for a feature, resulting in 69 features.

In doing the initial analysis with the CCS features, Bakker et al. point out that Nubi, as the only creole from the region designated “Africa” (read: non-Atlantic Africa), had to be removed from consideration, for methodological reasons: singularities are not permitted, i.e., no independent (predictor) variable can be represented by a single data point. (Equally, it was the only creole in the CCS sample with Arabic as its lexifier language.) When they carry out the second analysis, this time with their 69 creole features rather than the 97 CCS ones, they once again point out that they have removed Nubi. At no point do they indicate that they have removed Nagamese, which would be required insofar as it is the only creole in the set with Assamese as its lexifier language. Thus, if the analysis is carried out correctly, the two creoles with non Indo-European lexifiers are removed, introducing into the exercise the European lexifier bias common to the field.

Bakker et al. justify their use of “multiple regression as a statistical technique because it is especially well suited to test the influence of many variables at a time (as in this case), namely by calculating the effect of one variable while holding all other variables constant” (p. 23). If both Nagamese and Nubi have been removed, their regression analysis involves 16 data points, i.e. the 16 remaining languages. The four variants yield twelve variables:

Lexifier: Dutch, English, French, Spanish, Portuguese (5 variables)  
 Region: Atlantic, Indian Ocean, Pacific (3 variables)  
 Type: fort/trade, plantation, maroon (3 variables)  
 Age: a continuous variable (1 variable)

Bakker et al. report that their results, whether they use the CCS features or the creole features, are the same: “there is no significant relation between the degree of creoleness ... and the kind of lexifier, the kind of socio-historical situation, the area or the age of a given creole” (p. 17).

In terms of statistics, the assertion that there is ‘no significant relation’ constitutes a claim that no variable exists that falsifies the null hypothesis, i.e. no non-random patterns are to be found in the data. For Bakker et al., this means that there is “no clustering on the basis of lexifier, geography, substrate or historical connections”; instead, “it is likely that creoles themselves form a group” (p. 15). This in turn leads them to say that the results provide evidence for the hypothesis that “universal patterns play a role in creole genesis.”

However, Bakker et al.’s regression analysis runs afoul of the principle of statistical power, a concept that Ellis (2010: 52) defines as describing “the probability that a test will correctly identify a genuine effect. Technically, the power of a test is defined as the probability that it will reject a false null hypothesis.” Murphy observes that “if N is sufficiently small, any result will be judged ‘nonsignificant’ ” (2002: 121). That is precisely what has happened here: the sample size is so utterly small (16 data points for twelve variables) that it is devoid of statistical power. In other words, Bakker et al.’s regression analysis shows nothing. As one of our reviewers points out, Bakker et al. “confuse absence of evidence with evidence of absence.”

### 3.3 Extending the CCS feature set to 12 non-creoles<sup>8,9</sup>

Bakker et al. apply the CCS feature set to 12 non-creoles, listed in Table 4. They used the “original” 97 features with the “original” values. Thus, in the case of pronouns and case marking, this meant eliminating the obvious correction to the feature matrix for the 18 creoles. Now, once again, inflecting pronouns for case is counted as

---

8. An amended version of the coding used in Bakker et al. (2011) is found at <<https://phylogenetic-creole-studies.blogspot.com>>.

9. We were in the final stages of editing this paper when we obtained a copy of Fon Sing (2017). There is overlap between our work and his in 3.3.1 and in the discussion of tense-mood-aspect in 3.3.3. In both sections, our narratives are distinct from his, but we are like him in finding cause for concern in Bakker et al.’s methodology and coding. We refer the reader to Fon Sing’s article, and we thank Peter Bakker for providing us with a copy.



a creole property. Consequently, Haitian Creole and Jamaican Patwa, for example, are deemed maximally non-creole in this regard.

The twelve languages are listed below in Table 4. Bakker et al. report selecting languages that are “isolating and/or low in complexity” (31), reflecting the assumptions regarding creoles’ “simple” character of Parkvall (2008), from which the assessment of complexity derives.<sup>10</sup> They include three Niger-Congo languages, the Kwa language Akan, the Mande language Bambara, and the Bantu language Kimbundu.

In this section we call attention to some of the problems that weaken Bakker et al.’s application of the CCS features to the non-creoles. They are (1) features that are by definition inapplicable to the non-creoles, (2) the high rate of missing data, specifically for the three Niger-Congo languages among the twelve non-creoles, and (3) the high rate of errors in coding. Coding errors have been noted for Mandarin and English, two of the 12 non-creoles (see Fon Sing & Leoue 2012; Fon Sing 2017 for Mandarin), and we will comment further on errors for English. However, our primary concern with erroneous coding is not language-specific, as we shall show.

### 3.3.1 *Inapplicable features*

For the 18 CCS creoles, virtually every cell received a meaningful value. This was not the case for the 12 non-creoles. To begin with, as Fon Sing (2017) notes, in five instances Holm and Patrick’s definition of the CCS feature made explicit reference to the creole’s superstrate language or its lexifier:

- 8.5 subordinator from superstrate ‘that’
- 11.1 passive construction (“Not all Creoles have a passive construction parallel to that in their lexifier, e.g. “That house was built last year” Holm 2007: viii)
- 12.7 comparison as in superstrate
- 15.3 definite article (from superstrate deictic)
- 17.7 reflexive pronouns: distinct form (“Some Creoles have reflexive pronouns that differ from those in their lexifier languages” Holm 2007: x).

Appropriately, Fon Sing asks, “Which lexifiers/superstrates did Bakker et al. (2011) take for the non-Creole languages of their study?” In Table 4, we address Fon Sing’s point and show the feature values that Bakker et al. assign to these five features for the 12 non-creoles. We use “n.a.” to indicate those cells where Bakker et al. indicated that the feature was “not applicable” for the language in question. Surely the feature values ought to be “n.a.” for all twelve non-creoles. However, “n.a.” is only used for

---

10. Nevertheless, the 12 include several languages with complex agglutinative morphologies: the Bantu language Kimbundu, the Northern Dravidian language Brahui, and Kolyma Yukaghir.

the three Niger-Congo languages and English (and, even for these languages, not all the time). The values assigned to the other languages seem random.<sup>11</sup>

**Table 4.** The 12 non-creoles scored for those CCS features that make explicit reference to the superstrate or lexifier

Family	n	Language	Subgroup	Region	8.5	11.1	12.7	15.3	17.7
Niger Congo	3	Akan	Kwa	Ghana	n.a.	0	+	n.a.	n.a.
		Bambara	Mande	Mali	n.a.	+	?	n.a.	n.a.
		Kimbundu	Bantu	Angola	n.a.	+	+	n.a.	n.a.
African non-Niger-Congo	2	Koyra Chiini	Songhay, Nilo-Saharan	Mali	0	0	0	+	+
		Mina	Chadic, Afro-Asiatic	Cameroon	?	0	0	+	+
non African	7	Ainu	Isolate	Japan	?	+	0	0	0
		Brahui	Dravidian	Pakistan	+	+	?	0	+
		English	Indo-European	UK, US, elsewhere	n.a.	+	+	n.a.	n.a.
		(standard) Indonesian	Austronesian	Indonesia	0	+	0	0	+
		Kolyma Yukaghir	Isolate	Russia	0	0	0	0	0
		Mandarin	Sinitic, Sino-Tibetan	China	?	+	?	0	+
Pirahã	Mura, Amerind	Brazil	0	0	0	0	0		

### 3.3.2 Missing data

The 97 features for the 18 CCS creoles received a value – be it +, R, 0 – virtually all of the time. There are only 9 cells with a question mark (9/1746 = 0.5%). This is not equally true for the 12 non-creoles. In particular, as Table 5 shows, Bakker et al. have assigned a question mark to virtually one-fourth of the cells for the three Niger-Congo languages. In that Niger-Congo languages provided the bulk of substratal input into the Atlantic creoles, the widespread absence of feature values for Akan, Bambara, and Kimbundu sets up the real possibility that significant substratal contributions to creole grammars have been overlooked; the inclusion of these three languages among the list of twelve non-creoles meanwhile creates the false impression that substrate languages have been examined and found not to correlate with creole grammars, not even Atlantic creole grammars.

11. It is conceivable that, if a language does not have definite articles (15.3), the appropriate score for that cell is 0.

Table 5. Missing data for the 97 CCS features in the 18 creoles and the 12 non-creoles

	Missing data	n	%
18 creoles	9	1746	0.5%
Niger-Congo			
<b>Akan</b>	15	94	16.0%
<b>Bambara</b>	25	94	26.6%
<b>Kimbundu</b>	27	94	28.7%
<b>Total</b>	67	282	23.8%
African non-Niger-Congo (Koyra Chiini, Mina)	11	194	5.7%
Non-African (Ainu, Brahui, English, Indonesian, Kolyma Yukaghir, Mandarin, Pirahã)	22	676	3.3%

The lacunae are surprising, given that Akan and Bambara are well-studied languages. Kimbundu is perhaps less well-studied, but grammars of the language exist.<sup>12</sup>

Large data gaps are problematic in phylogenetic procedures, as we will discuss in some detail below, and decisions have to be taken about their treatment by the computations. Even before we turn to phylogenetic procedures, we can demonstrate the impact of Bakker et al.'s failure to obtain the necessary linguistic data. We will illustrate the issue using Holm & Patrick's ninth category, "dependent clauses." It consists of the following six features:

- 9.1 Subordinate clauses (non-embedded)
- 9.2 Subordinate clauses (embedded)
- 9.3 Relative clauses (relative pronoun = subject)
- 9.4 Relative clauses (relative pronoun = direct object)
- 9.5 Relative clauses (relative pronoun = object of a preposition)
- 9.6 Relative clauses (zero relative pronoun)

Table 6 shows what Bakker et al. found – or, more accurately, what they didn't find – in examining relevant grammars of the three Niger Congo languages compared with what we found.<sup>13</sup> When we examined these grammars, we were able to fill every one of these cells (and to correct Bakker et al.'s erroneous value for 9.1 for Kimbundu).

12. We have removed the features that Bakker et al. designated "not applicable" from the total number of features, hence *n* does not reach 97 for the non-creoles.

13. Bakker et al. relied on Christaller (1875) for Akan, Binger (1886) for Bambara, and Maia (1964) for Kimbundu. We similarly relied on Christaller (1875) for Akan, and used Bird and Kanté (1976) for Bambara, and Chatelain (1888–1889, 1894) for Kimbundu. Christaller (1875) is now available digitally. A search for "subordinate" leads to § 141 of the work, which begins "Conjunctions that connect *subordinate sentences* with *principal sentences* ..." (p. 90; italics in the original). The section provides numerous examples of non-embedded (feature 9.1) and embedded

Table 6. Treatment of CCS features 9.1–9.6 by Bakker et al.

	Bakker et al.'s findings			Our findings		
	Akan	Bambara	Kimbundu	Akan	Bambara	Kimbundu
9.1	?	?	0	1	1	1
9.2	?	?	1	1	1	1
9.3	?	?	?	1	1	1
9.4	?	?	?	1	1	1
9.5	?	?	?	1	1	1
9.6	?	?	1	0	0	1

### 3.3.3 *Errors in coding*

In discussing individual languages that they know well, various authors have faulted Bakker et al. (2011) for coding errors. We return below to this general issue. Our immediate concern is Bakker et al.'s overall coding for the 12 non-creoles for the 97 CCS features.

The first 27 CCS features involve tense-mood-aspect. Category 1 is “unmarked verbs,” and Category 2 is “anterior (or past) tense.” This sets up the pair expressed by 1.2 and 2.1, which pertain to the interpretation of the bare verb and the verb accompanied by an overt tense marker, respectively.

- 1.2 unmarked verb                      statives with past reference  
 2.1 anterior (or past) tense        statives with past reference

All 18 creoles display 1.2, and 17 of them display 2.1. The overlap is unsurprising; overt marking of tense on statives can be variable in a language. Non-statives show a parallel pair:

- 1.3 unmarked verb                      non-statives with past reference  
 2.2 anterior (or past) tense        non-statives with (past-before-)past reference.

Here the number of creoles deemed to have the features is 16 for 1.3, and 17 for 2.2.

Among the 12 non-creoles, Bakker et al. present the Nilo-Saharan language Koyra Chiini as being like the creoles in variably marking past tense for verbs with stative reference. However, Bakker et al. report the remaining 11 non-creoles as

---

(9.2) subordinate clauses. Christaller's discussion of relative pronouns (under the heading ‘relative pronouns’) is straightforward. He states that Akan relative clauses are marked by the relative clause marker followed by the appropriately case-marked pronoun, whether nominative, objective, possessive, or locative (90–91). Some of the locative examples have the head as object of a postposition. The information in Christaller concerning dependent clauses could hardly be more transparent, yet Bakker et al. assign a question mark to every one of the six features.

doing neither. For non-statives with past reference, their coding is similar: They code Koyra Chiini as using the unmarked verb for past non-statives, and Indonesian for using the anterior/past tense to mark past non-statives. All the other languages do neither.

These results point to a powerful pattern in Bakker et al: according to their coding, unlike the 18 creoles, the 12 non-creoles predominantly do not have tense-mood-aspect systems. Table 7 shows the number of languages in the dataset that are coded as having a particular feature.<sup>14</sup>

Table 7. The 27 CCS features pertaining to tense, mood, and aspect

	Creoles (out of 18)	Non-creoles (out of 12)
1. Unmarked verbs		
1.1 Statives with non-past reference	18	1
1.2 Statives with past reference	17	1
1.3 Non-statives with non-past reference	16	1
1.4 Non-statives with past reference	14	0
Anterior (or past) tense		
2.1 Statives with past reference	18	1
2.2 Non-statives with (past-before-)past reference	17	1
2.3 Anterior (or past) = counterfactual	17	0
2.4 Anterior (or past) with adjective	11	1
2.5 Anterior (or past) with locative	12	1
3. Progressive aspect		
3.1 Indicating progressive	17	3
3.2 Indicating future	13	0
3.3 Anterior plus progressive	16	0
3.4 Progressive with adjective = inchoative	10	0
4. Habitual marker		
4.1 Zero marker for the habitual	15	0
4.2 Progressive marker for habitual	12	2
4.3 Marker for habitual only	8	1
4.4 Anterior plus habitual	16	1
Completive aspect		
5.1 Completive only (before/after V)	11	2
5.2 Completive plus adjective	8	1
5.3 Anterior (or other preverbal markers) plus completive	12	0

14. The terms used for the 27 features in Table 7 come directly from Holm & Patrick (2007). A definition for each is found in that work.

Table 7. (*continued*)

	Creoles (out of 18)	Non-creoles (out of 12)
Irrealis mode		
6.1 Future (= progressive marker)	12	0
6.2 Anterior plus irrealis = conditional	13	0
6.3 Anterior plus irrealis = future in the past	10	0
6.4 Anterior plus irrealis = future perfect	0	0
7. Other combinations of verbal markers		
7.1 Irrealis plus progressive	13	0
7.2 Anterior plus irrealis plus progressive	11	0
7.3 Other auxiliary-like elements	18	3

The authors are quite definite about their judgment that the non-creoles mostly lack these features: apart from the odd presence mark, there are only four question marks (three of them for Indonesian). For these 27 features, creoles display the feature on average 73% of the time (355/486), while non-creoles are coded as doing so only 6% of the time (20/324). In that the features were explicitly constructed by Holm & Patrick to capture creole phenomena, it is to be expected that creoles show the features at a high rate. However, given the nature of the features, their virtual absence from among non-creoles is not plausible.

Holm and Patrick's feature descriptions do not limit the application of the features to the preverbal particles which have come to be seen as the hallmark of creole Tense-Mood-Aspect systems (cf. Singler 1990). Rather, the features pertain to the interpretation of whatever marker a creole language employs to encode the relevant distinction. Take for instance the features 3.1–3.4. The suffixed imperfective marker of Berbice Dutch, which does not fit the prototypical expectation of a preverbal particle, suffices to earn the language a 'present' mark for features 3.1, 3.3, and 3.4. Since a progressive marker exists in English (Holm references it in illustrating feature 3.2 with "I'm leaving tomorrow") it is clear that 'absent' coding cannot be accurate. Yet, English receives '0' for every feature in Table 7. This is true also for the three Niger-Congo languages as well as Brahui and Pirahã. Thus, fully half of the 12 non-creoles receive a '0' for every single feature in Table 7. Of the other six non-creoles, Indonesian receives a '+' for seven of the features and a question mark for three more, while the remaining five languages receive an occasional '?' or '+'.

Fon Sing and Leoue (2012) and Fon Sing (2017) point out extensive coding errors for Mandarin. This problem extends to the coding of WALS/APiCS features discussed in Section 5 below. DeGraff, Bass and Berwick (2013) for Haitian Creole, Kouwenberg (2010) for Papiamentu, Aboh (2016) for Ewegbe (Fon) and

Yoruba – all call attention to extensive coding errors. To the list of languages with a high rate of coding error, we now add English.<sup>15</sup> We find Bakker et al.’s assignment of a feature value for English to be in error a quarter of the time (24/94).<sup>16</sup> In the preponderance of cases (21/24), the error is that, while the feature exists in English, Bakker et al. have assigned it a ‘0’. Above, we noted this for the Tense-Mood-Aspect features. In Table 8, we provide four additional examples. In the first case, the English example is the one cited by Holm (2007) while explaining feature 16.5. For the other three, we provide examples of the relevant features from the text of Bakker et al. (2011).

**Table 8.** Examples from among the 21 CCS features that occur commonly in English but which Bakker et al. have asserted do not occur in English

CCS feature	Illustration
16.5 Possessive Pronoun: Distinct	possessive pronouns of a distinct form like “ <b>mine</b> ” (Holm 2007: x)
8.1 Zero infinitive marker	It goes without saying that the addition of, say, Zulu, Georgian or Apache would make the creoles <b>stand</b> out even more. (Bakker et al. 2011: 32)
9.4 Relative clauses (relative pronoun = direct object)	In this section we discuss in detail <b>those samples and data that we employ in more than one study.</b> (15)
9.5 Relative clauses (relative pronoun = object of a preposition)	The first part of our study to be discussed is a test in evolution, <b>for which the SplitsTree software was developed originally.</b> (18)

### 3.4 Summary

The set of features in Holm and Patrick was not constructed to be the basis for statistical analysis. Bakker et al. use it to carry out a multiple regression analysis, but the paucity of data points is so extreme that a meaningful conclusion is not possible. Below, we consider Bakker et al.’s use of the CCS data set for phylogenetic computations. In 3.3, we presented three flaws in Bakker et al.’s application of the

15. Bakker (2014a) underreports these criticisms, summarizing as “[s]ome data about Mandarin and English are incorrect”, claiming that only 11 errors have been detected, and that “the error rate is still around one per thousand, which is acceptable” (p. 439).

16. In Bakker et al. (2011), there are a greater number of errors in the English scoring, but Daval-Markussen (2013) corrects five of them. However, in two other instances, Daval-Markussen moves in the opposite direction, replacing Bakker et al.’s correct + with an incorrect 0.

CCS features to 12 non-creoles. As we showed there, their database suffers from critical weaknesses, including high percentages of incorrect feature values for individual languages (we have added English to the list of such languages), extended data gaps, and significant imbalances in data available for creole and non-creole languages – flaws which are passed on to the computational procedure. But first, we must explain what that entails.

#### 4. Phylogenetic computations and their applications in linguistics

We will begin by explaining the purpose and nature of phylogenetic programmes before turning to their applications in linguistics, primarily in linguistic phylogeny, but also, to a lesser extent, in linguistic typology. This lays the groundwork for our consideration of their application in creole linguistics, where Bakker et.al. (2011, et seq.) have used phylogenetic computational programmes to model the distance between creole languages and their lexifiers, and to model what they claim to be typological networks, both among creole languages, and as compared to non-creole languages. We will see that the firmness of Bakker et al.'s conclusions regarding the typological exceptionality of creole languages stands in stark contrast to the tentativeness seen in the work of other linguists who are trying to make phylogenetic computations work. We will raise questions regarding their selection of features, their treatment of feature specifications, and the accuracy of their data. We will see that the failure of their methodology to reproduce known typological groupings signals larger issues of validity.

##### 4.1 Phylogenetic programmes: Their purpose and nature and issues in their application

Phylogenetic programmes are used in the study of the evolution of biological species; their purpose is to estimate the relationships between species, leading to increasingly detailed phylogenetic trees or cladograms and assessments of evolutionary rates of speciation. This exercise is made possible by the fact that modern biodiversity results from mechanisms that have produced variation where originally there was none, and that evolutionary distance correlates with variation. Despite the global success of the approach, there are many debates regarding the assumptions about relationships between species that feed into the methodologies, and the roles and responsibilities of the biologist in the computation. We begin by setting out the methodologies and the basic assumptions of the field, and then review some of the issues, which have implications also for the application of the methodologies in linguistics.



#### 4.1.1 Introduction

Vandamme (2009: 16–17) names some of the mechanisms that have produced biodiversity: “mutations, duplication of genes, reorganization of genomes, and genetic exchanges such as recombination, reassortment, and lateral gene transfer” and points out that mutations are “most often used to infer relationships between species” (17). It could be said that mutations are the most reliable measure of evolutionary distance.

Increasingly, molecular data are used in the composition of phylogenetic trees, either replacing or being used alongside the Darwinian approach which is traditionally based on structural (morphological) features.<sup>17</sup> It is important to note though that phylogenetic computations accept coded input and are blind to the nature of the input, which could just as easily code structural features as genome sequence data, or a combination of both. It is this blindness of the computations to the nature of the input that has made it possible for the methodologies to be recruited for purposes other than the study of biodiversity.

Phylogenetic methodologies, whether computer-based or traditional, rely on the notion ‘character’. Wiley and Lieberman (2011: 108) have this to say about that notion:

...when a systematist states that some particular feature of an organism is a character he or she is making a hypothesis that this feature is a property of that organism. For purposes of analysis and comparison with other organisms, the particular form of the character is usually termed the character state.

The situation where character states are identical because of inheritance from a shared ancestor is known as homology, and it is assumed that this is the default situation. This explains why closely related species show little variation and, for many characters, share the same character states. Synapomorphy, the presence of shared inherited character states, constitutes “the only valid basis for establishing monophyly” (i.e., membership of a group with a common ancestor) (Wheeler & Giribet 2016: 337). Although we expect to see increased variation in character states between more distantly related species, the reverse is not necessarily true: where different species display the same character states, this may have come about a-historically, either through truly independent development, or because a developmental system was inherited from a shared ancestor which allowed the similarity in character states to form independently in both species (*ibid.*, p. 111). The term homoplasy refers to situations where identical character states do not result from shared inheritance.

---

17. Wheeler and Giribet (2016) discuss the potential impact of morphological analysis made possible by new non-invasive imaging technologies.

In the pre-computational era, biologists traditionally arrived at hypotheses about phylogeny in a three-stage process: first, a classificatory stage in which character states are described, then a typological stage in which order is imposed on those data, and finally a phylogenetic stage (Hennig 1966: 10–11). In the computational era, thanks to the ability of the computations to produce order from seemingly random data, it seems as if the second stage could be skipped – after all, the computations take the output of the classificatory stage as input. Once the character state data for the relevant taxa have been fed into the computation, the naive observer may feel that all that remains is for the computation to identify the most parsimonious tree;<sup>18</sup> roughly, this is the shortest tree, the most economical representation of relationships between taxa in terms of the number of changes in character states that connect them, as represented by the branches, also known as ‘edges’, in the diagram.

However, the literature cautions that outputs produced by the computations – typically hundreds or thousands of tree diagrams – constitute hypotheses, subject to further tests to discover their congruence with other data; “if different data sets give us similar trees this gives us confidence that both reflect the same underlying cause, namely they reflect the same evolutionary history” (Page & Holmes 1998: 214. See also Wiley & Lieberman 2011: 122; Vandamme 2009: 23). The biologist who evaluates the output must be familiar with the different data sets and their outputs, and must also weigh carefully the impact of missing data (especially likely where fossil data are used; more on this below), consider the length of branches in relation to assumptions about the rate of evolutionary change, and query whether the branching sequence is compatible with known evolutionary pathways. Then there is the possibility of contamination by homoplasy to take into account, which may have come about by a range of different processes, especially where great time depth is involved. The recognition of and coding for homoplasy in the initial data matrix is critical. Thus, the typological stage is very much present in the coding of character states and in the evaluation of the output of the computation.

#### 4.1.2 *The treatment of gaps in the data*

Where the number of taxa being considered does not exceed a dozen or so, and the data matrix is not overly large, manual checking of the outputs, applying a traditional analysis, may be possible to select the most felicitous output, one which is not only parsimonious but also well supported by known groupings, and free of suspect groupings. However, in most instances, “complexities of analysis are usually greater than the ability of an investigator to consider all of the possible phylogenetic

---

18. One reviewer pointed out that many contemporary methods do not concern themselves with parsimony. Be that as it may, much of the literature we surveyed relied on it, and so did the linguistic applications – including that of Bakker et al.

hypotheses that might be inferred from the data” (Wiley & Lieberman 2011: 155). Since the computations produce thousands of possible trees, further checks must be built into the computational techniques. This can be done by using optimality criteria, which allow additional computations to rank trees on the basis of “an explicit function that relates data and tree”, such as maximum parsimony (Page & Holmes 1998: 174f). Also, the computation can be made to carry out any number of additional random searches,<sup>19</sup> apply procedures that rearrange tree topologies by “branch-swapping”, apply techniques that weight a selected number of characters more heavily than others, and so on (Wiley & Lieberman 2011: 172ff). All this means that no single programme is responsible for the outputs we see in publications: combinations of computational techniques, which sometimes go through millions of runs, are used to arrive at these results. At worst, this means that tree-building programmes “are so assumption laden that they point towards a black box approach to systematics” (Williams & Ebach 2014: 170).

This is perhaps nowhere clearer than in the treatment of incomplete data, a situation which arises where fossil taxa are included in the computation. Kearney & Clark (2003: 264–265, following Wilkinson 1995) explain that a distinction should be made between redundant and non-redundant fragmentary taxa, as illustrated in (2) by taxa B and C, respectively. The few known character states for redundant taxon B replicate those of A; as a result, it cannot have an effect on determining the relationships of other taxa. In contrast, non-redundant taxon C, fragmentary as its record may be, represents a different combination of its known character states to those seen in the complete data sets for A and D. This means that it may assist in throwing light on the relationship between A and D.

(2) Fragmentary taxa	(after Figure 2, Kearney & Clark 2003: 265)
A	0 0 0 1 1 2 0 1 0 0 1 1 1 2 2 1
B	? ? ? ? ? 2 ? ? 0 ? ? ? 1 ? ? ?
C	? ? ? ? ? 0 ? ? 1 ? ? ? 0 ? ? ?
D	0 0 1 1 2 0 1 0 0 0 1 2 1 1 1 2

It follows that while B can be safely removed from the computation, C has the potential to enhance the computation. However, depending on the partial data matrix available as compared to other taxa, the inclusion of non-redundant fossil taxa may

---

19. One concern is that the algorithms are often so computationally intensive that they require weeks, even months to run. Additional algorithms have been designed to carry out heuristic search procedures which speed up the search for trees that best meet the optimality criteria, most typically resulting in a consensus tree, which captures the best supported splits in the landscape of most optimal trees (e.g., Dunn 2015: 201f). This represents a ‘consensus’ as determined by the computation, not by the biologist.

also create situations where the choice between alternative trees cannot be resolved. Consider the following additional fragmentary taxa to those in (2):

(2')	E	?	?	?	?	?	2	?	?	0	?	?	?	1	?	?	2
	F	1	?	?	?	1	?	?	?	0	?	?	?	1	?	?	?

E contributes to a situation where the number of most parsimonious trees is increased because of character conflict: it would be classified as sister to A on the basis of its 6th character state, but as sister to D on the basis of its last character state. Thus, its inclusion causes ambiguity in the output. F produces a so-called “wildcard” effect: on the basis of the 5th, 9th and 13th characters, it is a potential sister to either A or D, and the non-redundant information in the data matrix (the character state for the 1st character), does not allow for resolution, since it matches neither A nor D. In other words, “the effects of incomplete taxa and concomitant missing character data are not general, but matrix-specific” (Kearney 2002: 370).

Granted that fragmentary data can be useful, and that we are normally dealing with data sets so much larger than that in (2)–(2)' that manual checking becomes impossible, how then are missing data for taxa that must be included to be treated by the computation, and what effect will these data gaps have in the output? One option is the exclusion of characters for which fossil taxa cannot be scored; in (2), this would exclude all but the 6th, 9th and 13th characters. The issue with this solution is that it is likely to produce more ambiguous output: data resolve relationships, and excluding characters for all taxa makes it harder to resolve relationships for taxa for which (near-)complete datasets are available in principle. In fact, some research found that even highly incomplete taxa (95% missing data) can be accurately analysed, provided the overall number of characters is sufficiently large (for instance, 2,000) but not if the number of characters in the analysis is small (for instance, a few hundred); moreover, much depends on the method used (Wiens 2003; Wiens & Moen 2008).

Given that all characters are to be included, much depends on the specific settings that are employed. The option to allow the computation to randomly substitute actual character states for each '?' until it has identified the output that meets the optimality criterion (such as parsimony) gives the computation the role of the systematist and produces not only a tree as output, but does so on the basis of a hypothesized complete data matrix for the pertinent fossil taxa – a highly problematic result. As an alternative, data gaps may be coded as a different character state, treating '?' as a character state in its own right; however, this too assigns an interpretation to the missing data. We agree with Kearny & Clark (2003) that '?' ought to be treated as truly unknown, therefore incapable of adding changes in character states: “question marks are simply indicators of the unknown and they do not build trees” (272). However, others come to different conclusions (e.g., Ogden & Rosenberg 2007).

In sum, although the computation merely ranks results according to some optimality criterion, its final output reflects the assumptions of the biologist who provided the input and selected the parameter settings, and who retains a crucial role in the interpretation and validation of the output.

#### 4.1.3 *Interpreting trees*

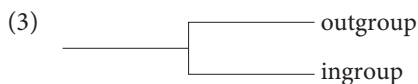
The visible result of the procedure consists of trees or cladograms. As is customary in evolutionary biology, a rooted tree postulates an origin, a postulated ancestral form; branches emanate from the root, and every internal node (i.e., every bifurcation) postulates a speciation event. (Unrooted trees are considered below.) It is worth noting that the root as well as the internal nodes are sets of character states postulated by the computation – which may or may not correspond to known ancestral species.

It is important to recognize that the computations that produce rooted trees do not inherently favour any one character state as ancestral, and will potentially therefore attempt to construct trees using any character state as the historical antecedent of others. If the evolutionary pathway is known, the initial coding of character states can take this information into account, a process known as “polarization”. The different character states for a given character are coded using either digits (0,1,...)<sup>20</sup> or letters (a,b,...). The coding can impose order by indicating polarization, for instance as  $0>1>2$  /  $a>b>c$ , indicating the historical sequence of descent. However, doing so is not always possible or desirable, and some have argued that characters should be allowed to “speak for themselves” (Williams & Ebach 2014: 174, ascribing this position to Colin Patterson, e.g. 1988).

An alternative to initial character state polarization is to include one or several known and well-defined “outgroups” in the computation, groups related to but not included in the “ingroup” being studied. This is an example of the use of sister group relationships to produce a hierarchy. Where a particular character shows variation within the ingroup, the character state shared with the outgroup(s) is assumed to have been inherited from the ancestral state preceding the event that split the groups. In contrast, any character state present among other members of the ingroup which differs from that of the outgroup(s) is concluded to be an innovation – even if it is the most commonly attested character state among extant members of the ingroup. The computation is expected to separate the outgroup(s) correctly from the ingroup, thus inherently providing the rooting for the resultant output. The resultant tree is expected to have, at minimum, the following structure:

---

20. It should be noted that the use of 0 does not imply absence of a character; rather, 0 codes one of a list of different character states.



Applying the outgroup methodology provides a benchmark for basic tree accuracy, but crucially relies on the correctness of prior information about the relationships between the outgroup(s) and (members of) the ingroup (Wiley & Lieberman 2011: 157).

Yet another alternative is to have the computation produce rootless trees, reticulate networks, that model only relatedness at the level of sister groups – perhaps warranted also by the idea that mutations are random and can therefore move in any direction. Page & Holmes (1998: 207) point out that conflicting signals result in competing splits which may be equally likely; in addition, gene recombination and inter-breeding introduce hybridity which cannot be captured in a tree. An unrooted network representation may be more appropriate at times.

Wiley and Lieberman (2011: 166) emphasize that, although the characters on a rootless tree “have no phylogenetic interpretation”, they can be given such an interpretation by the biologist, who assumes once again the traditional role of the phylogenetic analyst by specifying the starting point, in other words, by “rooting the tree”. By doing so, the length and complexity of the tree is not changed, but the polarity interpretation of the characters and the overall topology (branching order) of the tree are affected. In this way the initial network, which models divergent and convergent development, is changed to a visual representation of evolutionary relationships by rooting. This approach has the advantage of separating the problem of rooting a tree from that of estimating the relations between taxa and their character states. However, it places the burden of responsibility on the analyst – and it cannot be taken for granted that this responsibility is properly discharged. Penny’s (2013) discussion of a case where authors’ choice for the position of the root of the mammalian tree is spectacularly wrong exposes several “classical mistakes about evolutionary trees” (819), even leading him to suggest that “unfortunately evolutionary relationships cannot be done seriously by amateurs” (820).

#### 4.1.4 *Summary*

Summing up, while the need to compute evolutionary relationships over very large data sets requires reliance on computational programmes, most authors recognize that the outputs are taken to constitute hypotheses about relatedness rather than hard and fast facts. The biologist retains the traditional responsibilities of ensuring the quality of the input (making sure that it is not contaminated by homoplasy posturing as homology, deciding how to treat data gaps, selecting appropriate outgroups for benchmarking, and so on), as well as verifying the output by evaluating

suspect groupings, interpreting the robustness of the trees selected by the computation as a “true” representation of phylogeny, and even deciding on a position for the root of the tree which is congruent with known facts about evolution. All this requires intimate familiarity with a growing range of computational techniques, including tree-building computations supplemented with heuristic search techniques for optimal trees, and further computations which build consensus trees or otherwise arrive at an output that is considered to best meet the optimality criteria of the biologist.

## 4.2 Computational techniques for linguists

The publication of the online WALS and APICS databases alongside the explosion in the number of (full or partial) language descriptions and lexicons which are now available have made comparison between hundreds of languages possible. Comparisons on these scales require computational methods which do not come into play when a handful of languages are considered. For this purpose, linguists, often in partnership with mathematicians and/or evolutionary biologists, have extended biologists’ computerized modeling of genetic relationships to the reconstruction of the evolution of languages. These applications are made possible, as pointed out earlier, by the fact that the computations accept any data representing coded character states.

Applied to languages, the characters which are used to build the data matrix are either linguistic forms or linguistic features. Most of the work produces output in the form of evolutionary kinship trees, usually relying on cognate forms, but typological data and distance measures which code the amount of change in cognate forms are also used. Additionally, there have been applications in linguistic typology, using features rather than forms. We will discuss these applications in turn. But we begin by pointing out some of the most obvious differences between biological and linguistic phylogeny. These pertain to the nature of change, and to the nature of characters.

### 4.2.1 *The nature of change*

As we noted earlier, the default situation assumed in biology is that shared character states are homologous, i.e., inherited from a shared ancestor. Furthermore, change is normally species-internal, and proceeds via random mutations which ultimately, over great time depths, produce discrete species which preserve enough of the ancestral genetic code to establish biological relatedness with sister species. Languages, in contrast, are not discrete objects. Language change often proceeds as a result of contact situations which cause transfer of forms and constructions between languages, resulting in pervasive homoplasy. The reconstruction of the

genealogy of languages can be made more complex by areal diffusion, and the difficulties of distinguishing between inherited traits and areally diffused traits (Aikhenvald & Dixon 2001: 4).

Evolutionary change in biology is expected to be unidirectional, although it is also possible for mutations to re-introduce the semblance of an earlier evolutionary stage. As long as homoplastic and homologous character states can be differentiated by the biologist, it is theoretically possible to arrive at a data matrix which perfectly encodes phylogenetic signal, and which yields a completely resolved evolutionary tree. In contrast, linguistic change is not polarized; in principle, change can proceed in any direction (for instance, both ‘up’ and ‘down’ a grammaticalization chain – Fischer, Norde & Perridon 2004). Borrowing (i.e., lateral transfer) of forms or features is not constrained by any characteristics of the languages in contact (Thomason 2001: 63). These differences derive from the fact that language change is deeply affected by social and cultural behaviours:

...social factors are the only ones that need to be considered in assessing stability: linguistic factors (such as overall similarity of the languages in contact) seem to be totally irrelevant. (Thomason 2001: 21)

Thomason (2001: 69–71) makes the point that under the right social circumstances, especially where intense contact takes place, it is possible for languages to undergo wholesale change in vocabulary and/or grammar. Winford (2003: 13) notes that “diffusion of features across languages may be so widespread that the boundaries between the languages become blurred, even for the speakers themselves”. Metatypy (Ross’s 1997 term) refers to situations where languages change linguistic type as a result of such influences. Haspelmath (2004) lists several cases (all from Ross 1997), including that of the Kupwar varieties of Urdu and Kannada, which have undergone metatypy with Kupwar Marathi as the model. (See also Ross 2007.)

While basic vocabulary is usually expected to be relatively impervious to external pressure, Blasi et al. (2016) provide evidence of sound-meaning correspondences attested across multiple languages in a 40-word subset of the Swadesh 100 word list – a finding which has implications for the idea that “there is a small set of ultraconserved words that are particularly useful for establishing ancient genealogical relations” (10281).

The upshot is that unrelated languages may share large amounts of vocabulary, and that even similarities in morphology, syntax and semantics between related languages do not necessarily provide us with phylogenetic signal, as typologies frequently pattern areally rather than genetically – a fact which the publication of WALS has made clearly visible. In sum, homoplasy is ubiquitous and can be very difficult to distinguish from homology.



#### 4.2.2 *The nature of characters*

To begin, we consider linguistic features; in 4.3, we will detail how linguists approach the reconstruction of the evolution of languages using the presence/absence of cognate forms in languages. It is obvious that there is no counterpart to the DNA sequence in linguistics. Although WALS encodes an impressive 144 linguistic features (many more if sub-features are counted), linguists are well aware that this merely documents a tiny bit of the full range of linguistic typology, and moreover, that the exercise is steeped in theoretical assumptions, so that a different approach to the same features may well result in different coding of their presence in languages.

For feature-driven applications of the computational approach, this presents a conundrum. A possible solution is to argue that feature selection is an inherently arbitrary process, and that its outcome may therefore be considered sufficiently representative of the complete range of possible features of natural languages to allow the approach to have validity. This would then also entail that features in phonology, morphosyntax, lexical semantics, and so on, can somehow be treated as being of equal value.

Furthermore, because “each linguistic sign, both concrete and abstract, is organized with respect to all other signs” (read: forms and structures) (Fischer 2007: 323), a change in one sign will have repercussions for other signs. For instance, the idea of arbitrariness becomes harder to accept when we note that a language’s use of phonemic vowel nasalization (WALS feature 10) correlates positively with the number of vowel qualities a language distinguishes (WALS 2), so that the presence of vowel nasalization predicts a 50% likelihood that a language is in the “large” category for WALS 2. In contrast, of the total of 564 languages surveyed, only 184 or 33% percent overall have large vowel quality inventories.

So we have to reckon with the non-arbitrariness of features, the theoretical assumptions that they are laden with, their interdependence and lack of equivalency. All this also means that missing data have repercussions different from those in biology. Recall that the research cited previously shows that significant data gaps need not have a detrimental impact on the results of the computation, provided the number of characters (not to be confused with character states) is sufficiently large – for instance, in the thousands (see 4.1.2 for discussion). Such large numbers of characters are not available for languages, so that data gaps are more likely to produce an indeterminate result. Moreover, the interdependence between forms and functions means that the availability of data for some characters but not others may well skew the data, and strongly bias the output.

We now turn to the different applications of the methodologies in linguistics.

### 4.3 Modelling the evolution of language families: The relevance of prior knowledge and benchmarking

The goal of most linguists' use of computational methodologies is to model the evolutionary history of language families,<sup>21</sup> and an understanding of these applications is needed to be able to evaluate the applications in linguistic typology. The expected output takes the familiar form of a tree, described by Nichols & Warnow (2008) as follows:

[T]he root represents the common ancestor of the daughter languages, the languages are represented by the leaves of the tree, the remaining internal nodes of the tree represent intermediate ancestral languages, and the branching order indicated by the tree represents the diversification of the group of languages into subfamilies. (...) Each language in the input is represented by a path in the tree, representing the language in its different states as it evolved over time (761).

This description articulates clearly the view that the rooted tree does more than represent relationships between sisters: it postulates ancestral states which are presumed to correspond historically to actual languages. As is the case in biology, it is important to approach such output with care and to remember that well-trained experts in the reconstruction of the relevant language families are needed both to code the input data and to evaluate the outputs.

The input data most typically consist of characters which code for the presence of cognate forms in the languages. In other words, the method resembles glottochronology rather than the traditional comparative method. As Nichols and Warnow (2008: 765) point out, what this actually means is that a list of "glosses" is used (for instance, derived from the Swadesh word lists), and that what is coded is whether languages have cognate forms for these glosses. It is expected that several non-cognate forms are found across a group of languages, typically yielding multi-state characters, which, however, can be and often are translated to a set of binary characters, as illustrated in Table 9.

---

21. In fact, the methodology is used to draw inferences not only about language evolution but also about human migration and settlement patterns. It would take us too far afield to consider these issues; suffice it to say that some of the purported findings have been controversial because of conflicting interpretations with implications not only for linguistic prehistory but also for the interpretation of archaeological and palaeontological findings.

**Table 9.** Translating a multi-state character to a set of binary characters, after Dunn (2015: 191–2)

	Character <i>n</i>		Binary coding (presence/absence coding)		
	Meaning <i>n</i>	Multi-state coding	Character <i>n</i> 1	Character <i>n</i> 2	Character <i>n</i> 3
Language A	ciŋ	1	1	0	0
Language B	kit	2	0	1	0
Language C	kət, lpəc	2, 3	0	1	1
Language D	lpət	3	0	0	1

As can be seen here, it is loss / replacement of forms that is coded. Since the computation assumes each character in the data matrix to be reconstructable to an ancestral state, the linguist may need to decide which of the cognate character states carry phylogenetic signal, and which represent lexical transfer, coding for the likelihood that the ancestral form is, say, *n*1 rather than *n*2 or *n*3 (see Dunn 2015: 204). Underlying this model is the observation that “cognates can be gained once but lost multiple times” (Bowerman & Atkinson 2012: 828), and that languages are therefore “highly unlikely to independently gain the same cognate” (ibid.:829).

So how good are the results of this method? Nakhleh et al. (2005), using a database of 336 lexical, phonological, and morphological characters for 24 Indo-European languages, show that different procedures yield significantly different results, and that not all characters are equal:

...equal treatment of characters is probably unwise, with improved results obtained by recognizing that some characters (notably characters derived from inflectional morphology and complex phonological characters) are less likely to evolve in parallel or with back mutation. (p. 1)

However, advances in the methodology have led to more successful applications, as argued by Bowerman (2018), based on Bowerman & Atkinson’s (2012) application to lexical characters. Bowerman also claims that errors in individual characters, including undetected cases of homoplasy (loans), “have little impact”; as “the individual input of any single character is very small, the models are robust to individual errors” (286). It is worth considering Bowerman & Atkinson (2012) in a bit more detail. They coded the presence or absence of cognate data for 189 meanings across a sample of 194 Pama-Nyungan languages, yielding a matrix of 14,613 binary characters;<sup>22</sup> the

22. Bowerman (2018: 286) claims that the study includes more than 5,000 binary characters, suggesting that the study meets the typical requirement for reliable biological phylogenies of thousands of characters (see § 4.2). However, these characters derive from the coding of 189 meanings. In other words, on average, each meaning corresponds to over two dozen different cognate forms across the sample, each coded as a binary character, interdependent with other characters for the same meaning. Clearly, this is not the same as having over 5,000 *independent* characters.

sample included languages which provided data ranging from “good” to “extensive” (824). Their approach was “to model language evolution as the gain ( $0 \rightarrow 1$ ) and loss ( $1 \rightarrow 0$ ) of cognates along the branches of a language family tree or ‘phylogeny’” (828). Rates of borrowing, they argue, are generally quite low (826).

As is the case in biology, the computation was supplemented with additional procedures aimed at producing “a sample distribution of trees we should consider plausible, given our data and assumptions” (*ibid.*). Their discussion makes it clear that they made choices, at various points along the way, based on their views on cognate replacement and rates of change, their approach to estimating probability, and “prior beliefs about tree topology (such as knowledge about the existence of certain language groups or the age of those groups)” (*ibid.*). The result is a tree in which 25 of 28 “previously identified Pama-Nyungan subgroups are supported by our analysis,” which postulates macro-groupings several of which are “well supported”, and in which most isolates “were plausibly placed within the tree as sisters to previously established groups” (830).

One important point emerging from this discussion is the use of benchmarking: as illustrated by Bowerman & Atkinson’s study, trees or networks must at least replicate known (sub)groupings (and correctly identify outgroups, where used). If this benchmark has been met, other findings become more plausible. Dunn (2015) discusses the integration of different forms of prior knowledge into the method, so that the method focuses on producing trees that are further resolved than what was known already (202).

So do we now have “evidence that any of these methods is capable of accurate estimation of linguistic phylogenies” (Nichols & Warnow 2008: 184)? Despite the sanitized trees presented in articles, overlaying the most probable outputs of the computations on each other yields multiple pathways between languages, thus representing the conflict / uncertainty between different ways of arriving at the relationships. Moreover, relationships which are very controversial may never be resolvable by this method, given the intersection with the study of human migration and settlement and cultural contact. As Barbaçon et al. put it (2013: 148), the methods really work well when the data and the assumptions made by the methodology match – in other words, when the method works at least in part as a self-fulfilling prophecy.

#### 4.4 Computational approaches on the basis of linguistic features

Using computations to build trees or networks on the basis of characters representing linguistic features rather than cognate forms, with the aim to cluster languages by similarity, does not, strictly speaking, fall under phylogenetics, as pointed out by Bowerman (2018). Nevertheless, this type of work is united with that surveyed in

the preceding section “by a general approach to data gathering and analysis that stresses quantitative rather than solely qualitative arguments, explicit models of evolution, and specific hypothesis generation and evaluation” (282). What sets it apart is the fact that typological characters “are homoplastic by definition” (Nichols & Warnow 2008: 765). In other words, accidental similarities are expected, and cannot be treated as noise in the signal – they *are* the signal. Distance between languages in such a network represents typological difference; whether this is interpreted in relation to phylogeny depends on the goals of the particular application. Dunn (2015) points to three requirements for such methods to be used to infer ancestral states of a feature: “(i) a genealogical hypothesis; (ii) a model of transitions between states of a feature (e.g. probabilities for transition from ‘presence’ to ‘absence’, and from ‘absence’ to ‘presence’); and (iii) a set of observations of feature states from the tips of the tree” (206).

Dunn et al. (2011) is a study of word order universals, based on a sample of languages representing the four large families of Austronesian, Indo-European, Bantu, and Uto-Aztecan. For pairs of word order features, the study “calculated the likelihood that the observed states of the characters were the result of the two features evolving independently, and compared this to the likelihood that the observed states were the result of coupled evolutionary change” (p. 79). The study’s goal is to ascertain whether word order features such as dominant Verb-Object order correlate with Adposition-Noun orders. The study found no evidence that word order features were strongly dependent in *all* language families. On the other hand, they found lineage-specific dependencies, such as the coupling of Adjective-Noun and Genitive-Noun ordering in Indo-European, and of Adjective-Noun and relative clause-Noun ordering in Austronesian and Uto-Aztecan (p. 81). They claim that the study demonstrates “the power of phylogenetic methods to reveal structural linkages that could not be detected by cross-linguistic sampling” (ibid.).

Danielsen, Dunn and Muysken (2011) attempt an application of the methodology to the Arawakan language family, much of which is not well defined. They point out that “lexical borrowing is frequent in South America” (p. 175), which makes reliance on lexical data problematic. Instead, the study “presents a quantitative analysis of structural similarities among Arawakan languages” (180), “using network methods to visualize the patterns of interrelationship in a way that does not presume a genealogically consistent pattern but rather allows for the possibility of considerable horizontal transfer” (ibid.). They conclude that the resultant network “shows which languages seem to cluster and can be compared to the postulated internal relationships within the phylum...” (ibid.). The typological distance between language pairs was also correlated with geographic distance, and the finding was that “the geographical proximity of a pair of languages does not strongly predict

that they will be especially similar on a structural level” (*ibid.*). In fact, they find that geography predicts only about 7% of typological similarity overall, and only about 18% and 14% within the northern and southern groups, respectively (p. 183f).

The authors produced unrooted network representations rather than trees, with the intention to “capture the effect of conflicting historical signals” (*ibid.*). While the networks allow for clusters to be identified, there is a lack of clear major splits or speciation events (p. 185), evidence, they claim, of relatively recent expansion (*ibid.*) There is also a lack of clear network structure for several languages, and the southern groups do not cluster together as a coherent group – further evidence of “ancient diversity in the north and more recent migration to the south” – multiple migrations, in fact (p. 186). They do not consider their findings to be definitive. However, the idea of using reticulate networks to hypothesize contact events is of interest, as it turns reticulation, which is an indication of uncertainty in the relationships that are being modelled and usually considered a disadvantage, into a strength. However, their conclusions are based on the assumption that the methodology, were greater time-depth possible, would yield outputs which show clear splits and coherent clusters. Given that the phylogenetic applications to older families generally rely on prior knowledge about splits and coherent clusters to benchmark the output of the computation, that assumption is perhaps not warranted; we have yet to see evidence that the computation alone can do the work.

#### 4.5 Summary

While the computational procedures are blind to the nature of the data fed into them, we cannot lose sight of the fact that linguistic forms and features are not like DNA segments. The choice of forms and features for coding is absolutely non-trivial.

Where linguists’ goal is to model the evolution of languages, the computation resembles an automated version of glottochronology, with input data which typically consist of characters which code for the presence of cognate forms in the languages. The various applications described in the literature are highly reliant on experts’ abilities to recognize homoplasy, and to benchmark the outputs with respect to known (sub)groupings. Where the computational method is used to model typological relationships, the goal is not usually to ascertain genetic relatedness, although prior knowledge about relatedness may inform the interpretation of the output and it may be possible to use typological applications to model contact events. In all the applications surveyed, authors have been careful to point to the limitations of the procedures, have identified manual adjustments that were called for, and have supported only limited conclusions.

## 5. Computational methods and creole language typology

Bakker et al.'s (2011) innovation is in the application of computational methodologies to what is purported to be a truly typological question, in which family groupings are disregarded altogether. Bakker et al. apply the phylogenetic computational approach first to pidgin and creole databases, then to a comparison between pidgins and creoles on one hand, and non-pidgins and -creoles on the other.

### 5.1 Computational methods and creole language grouping

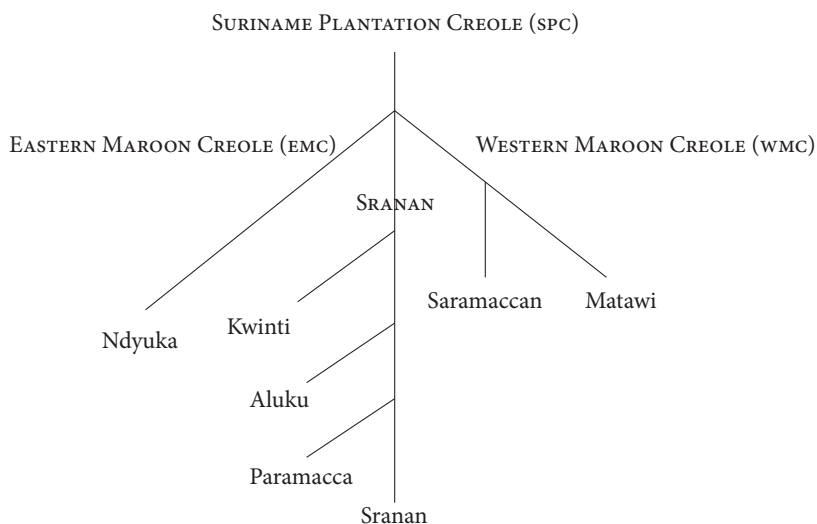
To begin, drawing on Hancock's (1987) database of 33 Atlantic Englishes and English creoles, data for creole languages of a single lexifier, English, were coded and fed into the computation.<sup>23</sup> The stated goal was to establish that the methodology produces accurate results. Using so-called formal features (lexical and phonological), the unrooted tree or network which was produced suggests that the Suriname creoles form a major subgroup, but that the relationships between other English creoles and Englishes is not well defined (p. 19). The same procedure was applied to structural data; although the Suriname group is again distinct, the distance that separates them from the other languages is much reduced, and internal coherence among the remaining languages is further reduced (*ibid.*). Finally, both types of data are combined, but this time (Standard) English is included and the authors have rooted the network, so that it appears as a tree – though not a bifurcating tree: it is notable that all varieties branch off the 'trunk'; the tree does not suggest descent relationships among any of the languages.

One issue with these results is the lack of benchmarking. In 4.3.2, we pointed out that, where trees or networks replicate known (sub)groupings, other findings become plausible. The authors claim to be pursuing a test in language evolution, but claim that "no classification of English-based creoles is yet widely accepted", so that they "are unable to compare the results with a particular model" (p. 20), suggesting the lack of a useful benchmark. However, there are well supported classifications based on known migration and dispersion patterns for several of the languages involved. For instance, in Bakker et al.'s Figures 1–3, Bahamian is not particularly close to the historically related North American varieties Gullah and AAVE (Sea Islands Creole and American Black English, in Hancock's 1987 terminology). Instead, depending on the tree, Bahamian appears closest to the varieties of Providence Island (Figure 1), to St. Kitts and Tobago Creole (Figure 2), and to Trinidad Creole (Figure 3); these are varieties with which it has no historical

---

23. The methodology is explained, although with little detail (Bakker et al. 2011: 18–19).

relations. The Suriname creoles constitute another potential benchmark, since their genealogy is well known. Bakker et al.'s rooted tree (their Figure 3) clusters the Suriname creoles as late developments, seemingly dating their emergence as occurring *after* that of languages that are known to have developed (much) later, such as Jamaican Creole, Guyanese Creole, Krio and other West African varieties, and so forth – clearly an inaccurate result. On the other hand, the relationships within the Suriname cluster of Bakker et al.'s network more or less replicate that of Arends (2017: 8), which represents their evolution and classification as follows:



**Figure 1.** Genealogical tree of the Suriname creoles

These contradictions show that Bakker et al.'s results index not relatedness, but similarity. The place of a language in their tree cannot be taken as anything more than a signal of its acrolectal or basilectal status, reducing the tree to no better than a pan-Creole continuum of English-lexicon varieties, a visual representation of something like Schneider's (1990) cline of creoleness.

Despite Bakker et al.'s assertion that "the results are so robust, [that] we can be quite confident in their validity and reliability" (p. 19), they make no attempt to account for the fact that the network does not represent the evolutionary relationships of the relevant languages; nor do they address the (variable) clustering seen in the diagrams, or the fact that results differ when different data sets are used – a finding which detracts from the so-called robustness of the results.

Next, the methodology is applied to the 18 CCS creoles, using the 97 CCS features (see § 3.1). According to the statistical analysis that was first applied to these data, "no statistically significant effects of age, area, lexifier or type on the number



of features present in a given creole” were found (p. 28). However, as we showed in 3.2, that analysis had so profound an insufficiency of datapoints as to be unusable. In contrast, the network in Figure 7 shows some clustering by lexifier, dominant substrate, and geographic area (*ibid.*), an indication that these factors may be relevant – a possibility briefly considered and rejected by the authors (p. 29). They conclude that “creoles are quite similar to one another, and the similarities are not primarily based on lexifier, area or substrate” (p. 30). We note that no attempt is made to account for the relationships suggested by the network.

Summarizing, having carried out a multiple regression analysis which is utterly invalid, Bakker et al. (2011) next aim to establish that accurate networks could be produced via the application of computational methodology to creole language data, but the outcomes of their efforts are both erratic and plainly inconsistent with known facts

## 5.2 Computational methods and creole typology

Having postulated yet not demonstrated that the computational methodology yields robust results, Bakker et al. (2011) proceed to apply it to the comparison between 18 creoles and 12 non-creoles which was discussed in 3.3. Their Figure 9 (p. 32) presents the network of 18 creoles and 12 non-creoles, and the striking result is that the two groups of languages are neatly separate. Within each group, clustering appears fairly random, especially for the non-creoles, where agglutinative Brahui and Kimbundu cluster with English, for instance. Again, there is no attempt to explain the clustering effects seen here. Nevertheless, the conclusion is drawn that “this shows that creoles form a typological class” (*ibid.*). However, we point here to the grave faults documented in 3.3: non-creole languages appear to have been coded randomly for inapplicable features; there is a high percentage of errors in the assignment of feature values for the 12 non-creoles; and data gaps account for between 16–29% of the data for the Niger-Congo languages.

In light of the nature of the computational procedures, coding errors are not normally expected to skew the results, provided the errors are unsystematic, and that they constitute but a minor presence in the data matrix. In effect, the title of Bakker’s (2016) column “You got Gungbe, but we got the numbers” points to the principle that data errors are insignificant in the larger pool; as Bakker puts it, “[s]ome error rate is unavoidable when you work with more than ten thousand data points” (424) – a point we agree with. However, where data errors are as systematic as we have shown them to be, they will bias the outcome, producing the apparent split between creoles and non-creoles.

This becomes very clear when we turn to the final, even more daring step: the application of the computational procedure to 34 pidgins and creoles and 153 non-pidgin or -creole languages included in the WALS database (Dryer &

Haspelmath 2013). A total of 43 features were coded, this time mostly WALS features rather than creole features.

The result is spectacular: their Figure 10 shows that pidgin and creole languages cluster at one end of a giant, highly reticulate network, separate from 152 of the 153 non-pidgin and -creole languages; only Hmong clusters with pidgins and creoles (p. 34). This result is even more remarkable given that non-pidgin and -creole languages do not appear to cluster in predictable ways. The authors note that “non-creoles do not classify along genetic or areal lines. For example, Basque (isolate, Western Europe), Hindi (Indo-European, India/South Asia), Burushaski (isolate, North Pakistan), and Hunzib (East-Caucasian, Caucasus) cluster” (p. 33).

Unfortunately, the result cannot be considered sound, on several grounds. In the following, we will address the failure of benchmarking for genetic and areal groupings (5.2.1), and a range of issues with the features and the data matrix, including bias in the selection and interpretation of features, systematic errors of coding, and the problematic treatment of data gaps (5.2.2). Bakker et al.’s findings contrast with those of Blasi, Michaelis & Haspelmath (2017), and we will end with a brief comparison.

### 5.2.1 *The failure of benchmarking*

Here, we will consider the fate of genetic and of areal groupings among non-pidgin or -creole languages in Bakker et al. We will see that their findings fail to replicate the typological patterns established by WALS.

Consider the case of Burmese. It appears on the same branch as Yoruba in the network. In contrast, closely related Lahu is at the opposite end of the network, separated by an innumerable number of edges. This being a typological network, this result suggests that Burmese is typologically closer to Yoruba than Lahu. But is it? How do genetic clusters fare generally? The majority of the 153 non-creole languages in the data set are the only representatives of their subgroups. Nevertheless, 27 languages, including Burmese and Lahu, belong to one of 10 subgroups of 2–4 languages. These are our findings:

- For 4 subgroups, there is a semblance of typological unity, in the sense that the languages belonging to them are located in more or less the same area of the network: Romance (French, Spanish), Germanic (English, German, Swedish), Tupi-Guaraní (Guaraní, Urubu-Kapoor), and Bantu (Luvale, Swahili, Zulu);
- For 2 subgroups, the network suggests significant typological separation: Madang (Amele, Kobon, Usan), Algonquian (Cree, Passamaquoddy-Maliseet);
- For 4 subgroups, there is dramatic typological separation: Oceanic (Fijian, Kilivila, Maori, Rapanui), Burmese-Lolo (Burmese, Lahu), Pama-Nyungan (Dyirbal, Martuthunira, Yidiny, Ngiyambaa), Semitic (Arabic, Hebrew); in each case, related languages are found on opposite sides of the network.

The separations suggested here are not congruent with known facts. Take Burmese again. It is true that Burmese and Yoruba show coincidental typological resemblances: both are specified in WALS as having complex tone systems (WALS feature 13), to use a plural word to code nominal plurality (WALS 33), to have third person pronouns which are unrelated to demonstratives (WALS 43), and to have basic subject-verb word order (WALS 81), to name a few features. However, the shared feature values represent only about half of the features for which both languages are specified in WALS. On the other hand, Lahu and Burmese share two-thirds of the features for which both are specified in WALS. If features were weighted equally and were randomly selected from among those in the WALS set, the computation would be expected to generate a network in which Burmese and Lahu are much closer to each other than either of these languages is to Yoruba. In this instance, the output misrepresents the typological reality. So how do we account for the place of these languages in the network? The possibilities which readily come to mind are bias in the feature selection, and pervasive coding errors.

First, the 43 features selected for the computation out of the 144 features in WALS may have been biased in favour of the typological resemblances between Burmese and Yoruba. If so, this would point to a lack of robustness of the output. That the features employed by Bakker et al. were not randomly selected was pointed out by DeGraff, Bass & Berwick (2013), noting that a small number of features force the split in the network between pidgins/creoles and non-pidgin/creole languages.<sup>24</sup> We will return to the selection of features in 5.2.2.

The second possibility is that feature values were entered erroneously. This possibility is not far-fetched. For instance, rescoring Papiamentu for Parkvall's (2008) feature selection, which forms the basis for Bakker et al.'s selection of features, changes the value for 14 features (a third of the total), and results in the total number of features coded '1' to change from 11 out of 43 in Bakker et al.'s data matrix to 21 out of 43 (Kouwenberg 2010). Aboh's (2016) recoding for Yoruba results in a value change in 20 of 43 cells (nearly half). In the discussion of the application of CCS features to non-creoles in 3.4, we argued that even English was not coded

---

24. Bakker (2014b: 188) refers to the 43 features as "an independent selection of features that the typologists compiling the *World Atlas of Language Structures* ... had deemed interesting in a survey of structures". The assumption here, as in Parkvall (2008), is that typologists are disinterested arbiters who randomly select linguistic properties to study. It may be useful to point out that WALS was compiled by sending out a request to linguists to propose chapters for it; if anything, WALS covers a convenience sample of features rather than an arbitrary one. In the compilation of WALS, choices were made regarding the selection of topics, the classifications used in the chapters, and the languages surveyed. We are reminded of Hennig's (1966: 11) point that at every stage of the analysis, including the initial stage of classification, assumptions come into play, and that classification is therefore no less framed by theory than is phylogenetics. This point is as valid in the study of language as it is in biology.

correctly. We believe that these findings are not limited to a mere three languages. We will see below that WALS features were, in many instances, reinterpreted, which also means that WALS values could not simply be adopted – introducing further opportunities for coding errors. It goes without saying that results derived from pervasive data errors lack credibility.

As it turns out, neither the non-arbitrary selection of features nor the widespread entry of erroneous feature values account for the treatment of Burmese and Lahu. In fact, the data appendix made available by Bakker et al.<sup>25</sup> shows that it is Lahu, not Burmese, that ought to pattern with Yoruba:

- for 9 features, Burmese contrasts with Yoruba and Lahu
- for 5 features, Burmese contrasts with Yoruba, and Lahu has ‘?’
- for 4 features only, Lahu contrasts with Yoruba and Burmese
- for 3 features only, Yoruba contrasts with Burmese and Lahu
- for the remaining 22 features, the three languages show no contrast, on the basis that they either all have the same feature value (0 or 1), or are all lacking a feature value (all ‘?’), or that a feature value of 0 or 1 contrasts with ‘?’ for the remaining language(s)

This would predict an output which separates Burmese from a grouping which includes Lahu and Yoruba.

Instead, the baffling result obtained by Bakker et al., which puts Lahu and Burmese at opposite ends of the network and clusters Burmese with Yoruba, is perfectly explained if ‘?’ the signal of an unknown feature value, is counted as a character state:

- 4 “real” contrasts between Lahu and Burmese/Yoruba + 10 instances where Lahu has ‘?’ and Burmese and Yoruba either both have 0 or 1 = 14 contrasts
- 11 features now contrast Burmese with Lahu/Yoruba
- 3 features contrast Yoruba with Burmese/Lahu
- 7 features now show different character states for all 3 languages.

Recall that in principle, missing feature values ought not to build trees (see discussion in 4.1). However, the separation of Lahu from Burmese suggests strongly that the computation treated ‘?’ as a separate character state, allowing it to help build the network.<sup>26</sup> So Bakker et al.’s use of ‘1’ for presence, ‘0’ for absence, ‘?’ for

---

25. We made no attempt to assess the accuracy of the feature values, so this statement should not be taken to mean that feature values as entered for these languages are accurate.

26. When asked, Daval-Markussen responded that “the software was designed so that it can handle these special values” (i.e., ‘?’ and ‘-’), and Bakker that “? are dealt [with] by the program, as far as we know, as unknown, and the different algorithms just ignore them”.

unknown, yields three possible choices, effectively turning each character into a multi-state character instead of a binary choice, by treating ‘?’ not as an unknown, but as a known, third character state. Given that Lahu is less well specified in the WALS database than Burmese, this has had the undesirable consequence that Lahu appeared to the computation as having character states different from those of Burmese for a substantial number of features. At 14 contrasts, Lahu vs. Burmese/Yoruba ends up ahead of all other possible contrasts, thanks to this treatment of ‘?’. As we just saw, a tally of the feature specifications disregarding instances of ‘?’ would have yielded an outcome where Burmese would have been separated from Lahu and Yoruba. Of course, that outcome is, in fact, just as undesirable, given the typologies detailed in WALS.

To illustrate the fate of areal grouping, we will use the 10 Australian languages included in the dataset. They include three Pama-Nyungan languages Dyirbal, Ngiyambaa and Yidiny,<sup>27</sup> alongside 7 languages of diverse genetic affiliation. The network separates them into three clusters: Gooniyandi, Mangarrayi, Ngiyambaa, Kayardild and Wardaman constitute one cluster; Maung, Dyirbal, Tiwi and Nunggubuyu constitute a second cluster; while Yidiny appears by itself. For convenience, we will refer to the groupings as clusters 1, 2, 3.<sup>28</sup> The separation of Yidiny is difficult to explain. Where a feature specification of 0 or 1 is available for Yidiny, 16 features have values which are identical across all three groups; for a further 16 features, Yidiny is specified as ‘?’, which ought to mean that these features cannot serve to distinguish it from other languages (although we just saw that the computation may have done just that); 8 more feature specifications are identical with cluster 1 while contrasting with cluster 2, 1 feature specification is identical with cluster 2 while contrasting with cluster 1, and only 1 feature shows Yidiny contrasting with both clusters 1 and 2. The basis for its separation appears exceedingly shaky. This time, taking ‘?’ values into account does not provide a straightforward explanation. In effect, Dyirbal has 21 ‘?’-specified features; if ‘?’ is counted as a third character state, Dyirbal would have made a better candidate for separation than Yidiny.

Turning to the broader picture presented by WALS, we compared the 136 (mostly multi-state) features which are specified for at least 5 of the 10 Australian languages. This time, we found that the best candidate for separation is Kayardild, which differs from all other languages for 16 features, whereas Yidiny is unique

---

27. There is an eleventh Australian language, Martuthunira, but we were not able to identify it in the network.

28. The three clusters are maximally distant from each other: cluster 1 appears in the bottom part of the network, cluster 2 on the left, 3 on the right. We have not checked Bakker et al.’s feature values for accuracy.

only for 8 features.<sup>29</sup> We also found that dispersion of the three Pama-Nyungan languages over the three different clusters is not well supported: they are identical for 56% of WALS features, which is as good as languages in cluster 2, which share almost 55% of feature values, and much better than languages in cluster 1, which share only 34% of their WALS feature values. Once again, Bakker et al.'s network fails to meet minimum benchmarking standards for typologically supported groupings.

The authors acknowledge that non-creole languages do not pattern along either genetic or areal lines in the network, but appear to take this result to be an affirmation of the validity of the methodology (p. 33); in this context, they also mention an unexpected cluster made up of Basque (isolate, Western Europe), Hindi (Indo-European, India/South Asia), Burushaski (isolate, North Pakistan), and Hunzib (East-Caucasian, Caucasus) (*ibid.*). Considering their overall WALS profile, these four languages share less than a third of their feature specifications. Not only has benchmarking against typologically uniform genetic and areal groupings failed, but the network has grouped together languages which are only minimally similar: cluster 2 of the Australian languages and the cluster made up of Basque, Hindi, Burushaski and Hunzib ought not to have been generated had the methodology done justice to these languages' WALS profiles.

### 5.2.2 *Feature selection and the data matrix*

So far, we have raised issues pertaining to features and data of three kinds: first, the possibility that the selection of features introduced (unintended) bias; second, the possibility that systematic and pervasive errors of coding as well as the treatment of unknown feature values as a separate character state yielded erroneous output; and third, the clustering of typologically dissimilar languages and the lack of benchmarking against known genetic and areal groupings supported by substantially similar WALS profiles.

Let us return to the selection of features. WALS features are classified into different categories, as per Table 10. Some WALS features are expanded into several sub-features. For instance, WALS feature 21A 'Exponence of Selected Inflectional Formatives' is accompanied by 21B 'Exponence of Tense-Aspect-Mood Inflection'. Feature 144 has been expanded into 25 features (144A-Y), all pertaining to the placement of negative morphemes. In the table, we list basic (unexpanded) and expanded numbers of features and we calculate the percentage of features selected by Bakker et al. out of the expanded totals.

---

29. Mangarrayi has a unique feature value for 12 features, Ngiyambaa, Yidiny and Tiwi for 8 each, Nunggubuyu and Maung for 7 each, Gooniyandi for 6, Dyirbal for 5, and Wardaman for 3.

Table 10. Bakker et al.'s feature selection compared to WALS feature categories

WALS feature categories	Number of WALS features		Number of features selected by Bakker et al.	Percentage of (expanded) WALS features that Bakker et al. selected
	basic	expanded		
nominal categories	28	29	14	48%
verbal categories	16	17	8	47%
simple clauses	24	26	11	42%
morphology	10	12	5	42%
nominal syntax	7	8	2	25%
phonology	19	19	1	5%
word order	19	56	0	0%
lexicon	10	13	0	0%
complex sentences	7	7	0	0%
other	2	2	0	0%
sign languages	2	2	0	0%
N/A	–	–	4	N/A
	144	191	45*	

\* Some of Bakker et al.'s features are indicated as capturing several WALS features. We have counted this as addressing all the corresponding WALS features. For instance, their F11 'Gender' is claimed to correspond to WALS 30 Number of genders, 31 Sex-based and Non-sex-based Gender Systems, 32 Systems of Gender Assignment, and was counted for 3 WALS features in 'nominal categories'. In other instances, Bakker et al. have several features corresponding to a single WALS feature. For instance, their F30 Grammaticalized past/non-past and F31 Remoteness distinctions of past both correspond to WALS 66 The past tense, and were counted for 1 WALS feature in 'verbal categories'. The 4 remaining features (F48 Demonstratives marked for number, F49 Demonstratives marked for gender, F50 Demonstratives marked for case, F52 Alienability distinctions) are not derived from WALS.

As can be seen here, the selection of features drawn from WALS by Bakker et al., itself a subset of those selected by Parkvall (2008), is not random. Parkvall (2008) justifies his selection of 53 features (of which 47 correspond to WALS features) in the context of an attempt to develop a measure of language complexity (269ff). For Bakker et al.'s typological purpose, Parkvall's selection, reduced further to 43 features,<sup>30</sup> hardly seems the most appropriate. The (near-) exclusion of phonological features and the complete exclusion of word order features are especially jarring.

30. The reduction was effected by removing four phonological features (Parkvall's F01, F02 – WALS 1, 2 Size of consonant and vowel inventories; Parkvall's F04 – WALS 12 Complexity of syllable structure; F04 – WALS 13 Tone), three nominal categories features (Parkvall's 12, 18, 20 – WALS 30 Number of genders, 41 Distance contrast in demonstratives, 45 Politeness distinctions in pronouns), one nominal syntax feature (his F26 – WALS 59 Possessive classification), and two non-WALS features (Parkvall's F51 Total amount of verbal suppletion, F53 Number of pronominal numbers). No justification for removing these features is provided, but we presume it to be the difficulty of converting them into binary features.

WALS features are for the most part multi-state features. The coding carried out by Parkvall (2008) and adopted by Bakker et al. converted them into binary features. For some features, this was achieved by reconceptualization of the feature. For instance, WALS feature 66 'The Past Tense' has 4 values: Present, no remoteness distinctions; Present, 2–3 remoteness distinctions; Present, 4 or more remoteness distinctions; No past tense. This was converted into two separate features, F30 Past/No Past, and F31 Remoteness of Past. While this conversion appears fairly straightforward, others are less clear or even show that a great deal of liberty was taken in this process. For instance, WALS 37 'Definite Articles' has 5 values: Definite word distinct from demonstrative; Demonstrative word used as definite article; Definite affix; No definite, but indefinite article; No definite or indefinite article. This corresponds to Bakker et al.'s F15 'Definite Articles'. Presumably, the choice is now reduced to 'has / does not have a distinct definite article', but this is not stated; it is unclear whether, for instance, a demonstrative used as definite article is/is not counted as a definite article, or whether a definite affix is/is not counted as a definite article. Bakker et al.'s feature F16 'Indefinite articles' further illustrates the mystery behind some of the coding. That feature is claimed to correspond to WALS 38 'Indefinite articles', which has 5 values: Indefinite word distinct from 'one'; Indefinite word same as 'one'; Indefinite affix; No indefinite, but definite article; No definite or indefinite article. The naïve reader is likely to assume that the binary version would have reduced these options to a choice between having / not having an indefinite article or some other grammaticalized way of marking indefiniteness. Instead, it was reduced to having / not having an indefinite article same as 'one'.

Another example of spurious correspondence is that of WALS 23 'Locus of marking in the clause', which distinguishes between head-marking, dependent-marking, double marking, no marking, and other types. It corresponds to 2 features in Bakker et al., F06 'Overt marking of direct object' and F07 'Double marking of direct object'. How the feature values of WALS translate to those of Bakker et al. is unclear. Languages which appears in WALS with the value 'head-marking' are not expected to mark objects for their grammatical roles – that information is carried on the verb. However, all languages which are either head-marking, dependent-marking, or double-marking receive the value '1' for F06. Only 'no marking' languages receive the value 0. All languages listed as pidgin or creole receive the value 0.

Many of the features used in this exercise do not correspond to APiCS features, which means that ascertaining their value for particular creole languages would have required intimate personal knowledge of the relevant languages, or at least extensive consultation of extant sources and possibly consultation with native speakers and language experts. The detailed discussion of Parkvall's (2008) treatment of Papiamentu in Kouwenberg (2010), which identifies pervasive errors, shows that we cannot give Bakker et al. the benefit of the doubt in this matter. Even



seemingly straightforward features are problematic. WALS features 33 ‘Coding of Nominal Plurality’ and 34 ‘Occurrence of Nominal Plurality’ illustrate. The corresponding APiCS features are 23 ‘Expression of nominal plural meaning’ (no plural, plural prefix, suffix, stem change, tone or stress change, reduplication, plural word preceding or following the noun) and 22 ‘Occurrence of nominal plural marker’ (obligatoriness with human and inanimate nouns). The WALS and APiCS features are jointly captured by Bakker et al.’s F14 ‘Grammaticalized Nominal Plural’. It seems a reasonable assumption that any of the many forms of coding listed in WALS and APiCS would qualify a language for having grammaticalized nominal plural, but this is not made clear. Furthermore, it is unclear whether the information of WALS 34/APiCS 22 regarding the variable occurrence of plural marking is captured. Strangely, among the pidgins and creoles only Sango and Fanakalo receive the value ‘1’ for this feature, Ndyuka and Français Tirailleur are both marked ‘0’, while all other pidgin and creole languages are marked ‘?’; the majority of these languages have a grammaticalized nominal plural marker, as clearly indicated in their APiCS entries.

In sum, the feature selection of Bakker et al. does not represent an appropriate typological range from among that presented by WALS, with word order features altogether absent and phonological features represented by a single feature only. Furthermore, the conversion of multi-state characters into binary state characters appears to have been achieved by either reducing or reconceptualizing the information captured in the features, thus both reducing the typological range and introducing uncertainty about the interpretation of features. The WALS-APiCS-Bakker et al. correspondences are mysterious, including instances where languages are marked 1 while clearly not having a feature (such as object marking), or ‘?’ while clearly having the relevant feature (such as grammaticalized plural).

Finally, let us consider those aspects of the data matrix which may specifically explain how it is that Bakker et al. achieved a separation of pidgin and creole languages from other languages. The following observations are pertinent:

- For 23 features (more than half of the 43 features), not a single pidgin or creole language is coded ‘1’
  - for 17 features, **all** pidgins and creoles are coded ‘0’
  - for 6 more features, **all** pidgins and creoles are either coded ‘0’ or ‘?’
- Hmong patterns with pidgins and creoles on all but three of these features; for those three features (F49, 50, 52), creoles are coded ‘0’, while Hmong is coded ‘?’
- No other non-pidgin or -creole shares this pattern

Recall that the computation assumes that shared character states signal relatedness, and will attempt to account for shared character states by reducing the number of edges between taxa – i.e., languages. Little wonder, then, that pidgins and creoles

cluster together, along with Hmong: they alone share this pattern, and it represents a substantial part of all features. If this were accurate, it would be a significant finding. However, it is safe to say that it is not accurate. The 0 or 0/?-scoring features include such properties as F11 ‘Gender’, F19 ‘Gender marking in pronouns’, F34 ‘Morphological imperative’, F40 ‘Subject marking as both free word and agreement’, F52 ‘Alienability distinctions’ – all features for which we know some creoles which ought not to score 0. For instance, 4 of the 0 or 0/?-scoring features were changed to 1 when rescoreing Papiamentu (based on Kouwenberg 2010). We predict that feeding the corrected scores into the computation would separate languages like Papiamentu from Bakker et al.’s pidgin and creole cluster – since suddenly, they would look like any ‘ordinary’, non-creole language to the computation.<sup>31</sup>

Since systematic errors such as those set out here will bias the outcome of the computation, even a slight correction will change the result. This explains why Bakker et al. have shifted their accounts of both the number of features and the choice of features which distinguish creoles from non-creoles. The number was three in Daval-Markussen (2013) and Bakker (2014b). Bakker (2014b: 189) reported that “all creoles have WALS feature 38, value 2 (Indefinite article = ‘one’), and feature 69, value 5 (No Tense-Aspect Inflection)”. According to Bakker (2014b), combined with creoles’ negative score on the feature for noun classifiers (WALS 55), this sufficed to set all creoles apart from non-creoles. In Daval-Markussen & Bakker (2017: 126–33), the authors acknowledge that both Daval-Markussen (2013) and Bakker (2014b) overlooked that not all creoles have an indefinite article derived from the numeral ‘one’, or make no use of tense-aspect inflection. Instead, they recruit WALS 112, value 2 (Negation is expressed with a negative particle) and WALS 117, value 5 (‘to have’ expresses predicative possession) to complement features 38, 55, and 69. They variously claim that just four features (126) or three to five features (128) suffice to separate pidgin and creole languages from other languages. Although the randomness of this new set is not lost on the authors, they seem not to be troubled by it. We, on the other hand, are deeply troubled by the arbitrary nature of the exercise. Moreover, we need only one language to behave differently for this house of cards to collapse as well. For instance, in contrast to Bakker et al.’s assertion that all pidgins and creoles conform to WALS 112, value 2 – that negation is expressed with a negative particle, we learn from APiCS feature 100 ‘Negative morpheme types’ that there are 2 APiCS languages which exclusively use a

---

31. Daval-Markussen (2013), which forms part of the Bakker et al. body of work, uses a data set for which 14 features that are directly comparable between APiCS and WALS were selected and scored for all 76 APiCS languages and 55 WALS languages. This data set is not included with the sets made available through <<https://phylogenetic-creole-studies.blogspot.com>>. However, since it forms part of the data sets compiled by the same team, the same problems would apply.

negative affix (Ma'a/Mbugu and Sri Lankan Malay, neither of which are included in Bakker et al.), 1 language which exclusively uses a negative auxiliary verb (Gurindji Kriol, also not included), and 4 languages which exclusively use a bipartite negative marker (Santomense, Angolar, Fa d'Ambu, and Media Lengua; only Santomense and Fa d'Ambu are included), along with 53 languages which exclusively use a negative particle and 15 more languages which use a negative particle as one of their options for the expression of negation.<sup>32</sup> While the use of a negative particle is clearly the dominant value, it is not the only one. A similar point applies to WALS 117 / APiCS 78 'Existential verb and transitive possession verb': while the two are the same for 41 APiCS languages, 17 languages differentiate them, 8 show overlap, and 9 lack the relevant form altogether.

Bakker et al.'s findings can be contrasted with those of Blasi, Michaelis & Haspelmath (2017). On the face of it, a comparable methodology and data set were used: 41 features present in both APiCS and WALS were used over 48 APiCS languages and 111 non-creole WALS languages. However, different from Bakker et al., features were selected solely on the basis of their being shared by the two databases and having sufficient coverage; this means that word order features were included, and that features were not reconceptualized, reduced, or split. Their experiment shows that sets of features, combined to form 'creole profiles', can "discriminate efficiently between creoles and non-creoles" (3). However, they also found that for about half of all features, creole languages displayed a phenomenon which they label "ancestry dependence", which means that the feature is associated with a creole's lexifier or substrate group. Many of the remaining features were not classifiable as ancestry dependent because they either did not show enough variation according to lexifier and substrate groups, or too much variation. Ultimately, then, their claim is that "the majority of creole grammars have been transmitted – as in any other natural language – from their ancestry, either from their lexifiers and substrates or through later contact" (5). This leads them to conclude that "[s]ince the ancestry of creole languages usually comes from a few groups, distinguishing them from a balanced sample of non-creole languages is as unsurprising as the fact that a set of languages related through genealogy or area could be distinguished from other unrelated or less related languages" (5). This point is strengthened by the fact that

---

32. It is not clear how Daval-Markussen and Bakker (2017) come to define this feature as 'Negation is expressed with a negative particle'. In their database, their feature F44 (corresponding to WALS 112) is 'Obligatory double negation', reflecting Parkvall's (2008) intention to measure complexity – the presumption being that double negation is more complex than a single occurrence of negation. While Santomense is recognized as having this feature, Fa d'Ambu is not. The dataset as made available by the authors does not include a feature corresponding to WALS 117 Predicative possession.

every profile also misclassified some creoles as non-creoles and vice versa: “some of the languages that were incorrectly identified as creoles were either among the set of ancestral languages or were similar to them (such as English or Yoruba), whereas some of the creoles with typological properties different from the majority of lexifiers and substrates (such as Angolar and Tayo) were classified as non-creoles” (5). The upshot is that so-called creole profiles are profiles of creoles with certain lexifiers and substrate languages and cannot be expected to also identify creoles of different ancestries.

### 5.3 Conclusion

Our primary focus has been on the methodologies of Bakker et al., with an aim to demystify them and, in so doing, lay bare their insurmountable flaws and limitations. Bakker (2014b: 181), based on Bakker et al. (2011), makes the claim that “we know for a fact that creoles are distinct”; he specifically points to “thousands of data points and the systematic comparison of four different databases of creole and non-creole languages”. However, because they are data-driven methodologies, computational approaches can only produce results as good as the data that are fed into them. As we have seen, statistical analyses must meet the minimum expectations for sample size to be a useful exercise; this was not the case in Bakker et al.’s use of a multiple regression analysis. We have also seen that bias in the selection of features and very large numbers of systematic errors each potentially generate misleading outputs when translated into data matrices for computational procedures, as demonstrated by the failure of Bakker et al.’s output to replicate known typological groupings. Instead, the combination of bias, pervasive errors, and uncertainties in the treatment of ‘?’ produce output which is meaningless. The rejection of numerous critiques regarding the validity of the data on the grounds that “[s]ome error rate is unavoidable when you work with more than 10,000 data points” (Bakker 2016: 424) points to a dangerously cavalier attitude towards what is surely the foundation for all linguistic analysis: reliable data.

However, errors can be corrected. Since correction will affect the values (character states) of the very features which have been claimed to separate pidgins and creoles from non-pidgins and -creoles, we do not expect a corrected data matrix to reproduce the separation; as we noted above, with the corrections, creole languages begin to look like any other language – as anticipated by Muysken (1988a).<sup>33</sup>

---

33. One reviewer wanted us to redo the computation with corrected data. There are many reasons for thinking that this is not useful and perhaps even irresponsible. As this article makes very clear, there are many complexities involved in carrying out these computational procedures: the selection of the software and its settings, the additional search procedures intended to identify

We recognize that computational methods and networks have the potential to reveal the classification of creoles based on their ancestry (Blasi, Michaelis & Haspelmath 2017), or can reveal the distribution of typological patterns, as for instance in the carefully conducted study of Danielsen, Dunn & Muysken (2011). The question is whether Bakker et al.'s methodology can likewise be meaningfully employed. In 4.2.2, we touched on the core question whether a typology can be built on the ideas that any phonological, morphological, syntactic or semantic features can be equated and that any selection of these features can represent all typological space. Both premises seem naïve: That features could be treated as equivalent no matter which part of the grammar they come from is typologically dubious, and that results can be replicated regardless of the features employed is unlikely. Moreover, we contend that methodologically sound procedures will fail to produce a typological distinction between creoles and non-creoles. We say that as people who know creoles – because the full range of creoles (not just Atlantic creoles with Germanic or Romance lexifiers and a predominantly Niger-Congo substrate) is too diverse. Specifically, they are diverse in ways that other, non-creole languages are, as Pieter Muysken told us thirty years ago.

## Acknowledgements

We thank the editors for allowing us the opportunity to write an article in honour of a colleague who has consistently set high standards in his scholarly work, ethical values, and collegiality. The originality of Pieter Muysken's insights always seemed to put him ahead of his time. As a result, his ideas have endured, as this article is intended to show. We wish to acknowledge the assistance of Simon Mitchell (UWI, Mona Campus) who introduced Silvia Kouwenberg to the basic concepts and mechanisms of phylogenetic computation, Aymeric Daval-Markussen and Peter Bakker who graciously answered questions about the coding of character states in the Bakker et al. databases, Nathan LaFave and Sean Martin (then at NYU) for their discussion with us of the requirements of multiple regression studies, and the audience at the Society for Pidgin and Creole Linguistics summer 2015 meeting in Graz for their comments on Kouwenberg, Singler & Mitchell (2015). We also thank Norval Smith, who helped us refocus on the significant goals of this paper, and Damián Blasi and an anonymous reviewer, whose comments helped to identify weaknesses in the paper.

---

interpretable output, the computing power required to run thousands of iterations of the procedure, the (manual) interventions needed to ensure that benchmarks based on known groupings are met, and so on. (See 4.1.2 for more detailed discussion.) Even in the unlikely event that we could pull this off, the value of such a procedure is dubious, given the more general questions raised in 4.2.2 regarding the validity of feature sets as representations of typological space, along with the problems of coding and so on.

## References

- Aboh, E. O. 2016. Creole distinctiveness: A dead end. *Journal of Pidgin and Creole Languages* 31.2: 400–418. <https://doi.org/10.1075/jpcl.31.2.07abo>
- Aikhenvald, A. Y. & Dixon, R. M. W. 2001. Introduction. In *Areal Diffusion and Genetic Inheritance*, A. Y. Aikhenvald & R. M. Dixon (eds), 1–26. Oxford: OUP.
- Ansaldo, U. & Matthews, S. J. 2001. Typical creoles and simple languages: The case of Sinitic (Commentary on McWhorter, 2001). *Linguistic Typology* 5(2–3): 311–325.
- Ansaldo, U., Matthews, S. J. & Lim, L. (eds) 2007. *Deconstructing Creole* [Typological Studies in Language 73]. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.73>
- Arends, J. T. G. 2017. *Language and Slavery. A Social and Linguistic History of the Suriname Creoles* [Creole Language Library 52]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.52>
- Bakker, P. 2014a. Creoles and typology. Problems of sampling and definition [Column]. *Journal of Pidgin and Creole Languages* 29(2): 437–455. <https://doi.org/10.1075/jpcl.29.2.09bak>
- Bakker, P. 2014b. Creolistics. Back to square one [Column]. *Journal of Pidgin and Creole Languages* 29(1): 177–194. <https://doi.org/10.1075/jpcl.29.1.08bak>
- Bakker, P. 2016. You got Gungbe, but we got the numbers: Feature pools show that creoles are still typologically distinct (A response to Aboh, 2016). *Journal of Pidgin and Creole Languages* 31(2): 419–435. <https://doi.org/10.1075/jpcl.31.2.08bak>
- Bakker, P., Daval-Markussen, A., Parkvall, M. & Plag, I. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26.1: 5–42. Reprinted 2013, in *Creole Languages and Linguistic Typology*, P. Bhatt & T. Veenstra (eds), 9–45. Amsterdam: John Benjamins. <https://doi.org/10.1075/bct.57.02bak>
- Bakker, P., Borchsenius, F., Levisen, C. & Sippola, E. (eds). 2017. *Creole Studies – Phylogenetic Approaches*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.211>
- Baptista, M., Mello, H. & Suzuki, M. 2007. Kabuverdianu, or Cape Verdean, and Kriyol, or Guinea-Bissau (Creole Portuguese). In Holm & Patrick (eds), 53–82.
- Barbançon, F., Evans, S. N., Nakhleh, L., Ringe, D. & Warnow, T. 2013. An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica* 30(2): 143–170. <https://doi.org/10.1075/dia.30.2.01bar>
- Binger, L.-G. 1886. *Essai sur la langue Bambara, parlée dans le Kaarta et dans le Bélédougou, suivi d'un vocabulaire*. Paris: Maisonneuve frères et Leclerc.
- Bird, C. S. & Kanté, M. 1976. *An ka bamanankan kalan: Intermediate Bambara*. Bloomington IN: Indiana University Linguistics Club.
- Blasi, D. E., Michaelis, S. M. & Haspelmath, M. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1: 723–729.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F. & Christiansen, M. H. 2016. Sound-meaning association biases evidenced across thousands of languages. *PNAS* 113(39): 10818–10823.
- Bowern, C. 2018. Computational phylogenetics. *Annual Review of Linguistics* 4: 281–296.
- Bowern, C. & Atkinson, Q. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88: 817–845.
- Chatelain, H. 1888/1889. *Kimbundu grammar: Grammatica elementar do kimbundu ou lingua de Angola*. Geneva: Typ. de C. Schuchardt. Reprinted 1964, Ridgewood NJ: Gregg Press. <<https://babel.hathitrust.org/>> (26 April 2020).

- Chatelain, H. 1894. *Folk-tales of Angola. Fifty Tales, with Ki-mbundu text, literal English translation, introduction, and notes*. New York NY: Houghton Mifflin, for the American Folk-Lore Society.
- Christaller, J. G. 1875. *A Grammar of the Asante and Fante Language called Tshi [Chwee, Twi], Based on the Akuapem Dialect with Reference to the Other (Akan and Fante) Dialects*. Basel: Basel Evangelical Missionary Society. Republished 1964. Ridgewood NJ: Gregg Press. <<https://catalog.hathitrust.org/>> (26 April 2020).
- Danielsen, S., Dunn, M. J. & Muysken, P. C. 2011. The spread of the Arawakan languages: A view from structural phylogenetics. In *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archaeology, Linguistics, and Ethnohistory*, A. Hornborg & J. D. Hill (eds), 173–195. Boulder CO: University Press of Colorado.
- Daval-Markussen, A. 2013. First step towards a typological profile of creoles. *Acta Linguistica Hafniensia: International Journal of Linguistics* 45(2): 274–295.
- Daval-Markussen, A. & Bakker, P. 2011. A phylogenetics-network approach to the classification of English-based Atlantic creoles. *English World-Wide* 32(2): 115–146. <https://doi.org/10.1075/eww.32.2.01dav>
- Daval-Markussen, A. & Bakker, P. 2017. Creole typology II: Typological features of creoles: From early proposals to phylogenetic approaches and comparisons with non-creoles. In *Creole Studies – Phylogenetic Approaches*, P. Bakker, F. Borchsenius, C. Levisen & E. Sippola (eds), 103–140. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.211.06dav>
- DeGraff, M., Bass, T. & Berwick, R. 2013. Computational phylogenetics, creole languages, and family values. Paper presented at the 19th International Congress of Linguists, Geneva.
- Dunn, M. J. 2015. Language phylogenies. In *The Routledge Handbook of Historical Linguistics*, C. Bownen & B. Evans (eds), 190–211. New York NY: Routledge.
- Dunn, M. J., Greenhill, S. J., Levinson, S. C. & Gray, R. G. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature. Research Letters* 473: 79–82. <https://doi.org/10.1038/nature09923>
- Dryer, M. S. & Haspelmath, M. (eds) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info>> (26 April 2020)
- Ellis, P. D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results*. Cambridge: CUP.
- Fischer, O. 2007. *Morphosyntactic Change. Functional and Formal Perspectives*. Oxford: OUP.
- Fischer, O., Norde, M. & Perridon, H. (eds) 2004. *Up and Down the Cline. The Nature of Grammaticalization* [Typological Studies in Language 59]. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.59>
- Fon Sing, G. 2017. Creoles are not typologically distinct from non-Creoles. *Language Ecology* 1(1): 44–74. <https://doi.org/10.1075/le.1.1.04fon>
- Fon Sing, G. & Leoue, J. 2012. Creoles are not typologically distinct from non-Creoles. Paper presented at the 9th Creolistics Workshop: Contact languages in a global context: Past and present, Aarhus University.
- Good, J. 2004. Split prosody and creole simplicity. The case of Saramaccan. *Journal of Portuguese Linguistics* 3: 11–30.
- Hancock, I. F. 1987. A preliminary classification of the anglophone Atlantic creoles with syntactic data from thirty-three representative dialects. In *Pidgin and Creole Languages. Essays in Memory of John E. Reinecke*, G. Gilbert (ed.), 264–333. Honolulu HI: University of Hawaii Press.

- Haspelmath, M. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? Review article of Aikhenvald & Dixon 2001. *Studies in Language* 28(1): 209–223. <https://doi.org/10.1075/sl.28.1.10has>
- Hennig, W. 1966. *Phylogenetic Systematics*. Urbana IL: University of Illinois Press.
- Holm, J. A. 2007. Introduction. In Holm & Patrick (eds), v–xi.
- Holm, J. A. & Patrick, P. L. (eds) 2007. *Comparative Creole Syntax. Parallel Outlines of 18 Creole Grammars*. London: Battlebridge.
- Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: Mistaken assumptions and conclusions. *Systematic Biology* 51(3): 369–381.
- Kearney, M. & Clark, J. M. 2003. Problems due to missing data in phylogenetic analyses including fossils. A critical review. *Journal of Vertebrate Paleontology* 23(2): 263–274.
- Kouwenberg, S. 2004. The grammatical function of Papiamentu tone. In *Creole Languages and Portuguese*, N. S. H. Smith (ed.). Special issue of *Journal of Portuguese Linguistics* 3(2): 55–69.
- Kouwenberg, S. 2010. Creole studies and linguistic typology: Part 2 [Guest Column]. *Journal of Pidgin and Creole Languages* 25(2): 359–380. <https://doi.org/10.1075/jpcl.25.2.06kou>
- Kouwenberg, S. & Singler, J. V. 2011. Pidgins and creoles. In *The Cambridge Handbook of Sociolinguistics*, R. Mesthrie (ed.), 283–300. Cambridge: CUP.
- Kouwenberg, S., Singler, J. V. & Mitchell, S. 2015. The quality of the output is determined by the quality of the input: Methodological issues in new computational approaches to creole typology. Paper presented at the Society for Pidgin and Creole Linguistics Summer Meeting, Graz.
- Kusters, W. & Muysken, P. C. 2001. The complexities of arguing about complexity. Commentary on McWhorter, 2001. *Linguistic Typology* 5(2–3): 182–185.
- Maia, A. da S. 1964. *Lições de gramática de quimbundo: Português e banto, dialecto omumbuim*. Cucujães: Escola Tipográfica.
- McWhorter, J. H. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74(4): 788–818.
- McWhorter, J. H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3): 125–166.
- Michaelis, S. M., Maurer, P., Haspelmath, M. & Huber, M. (eds) 2013. *The Atlas of Pidgin and Creole Language Structures Online*. Oxford: OUP. <<http://apics-online.info/>> (26 April 2020).
- Murphy, K. 2002. Using power analysis to evaluate and improve research. In *Handbook of Research Methods in Industrial and Organizational Psychology*, S. G. Rogelberg (ed.), 119–137. Oxford: Blackwell.
- Muysken, P. C. 1988a. Are creoles a special type of language? In *Linguistics: The Cambridge Survey, Vol II: Linguistic Theory: Extensions and Implications*, F. J. Newmeyer (ed.), 285–301. Cambridge: CUP.
- Muysken, P. C. 1988b. Lexical restructuring in creole genesis. In *Beiträge zum 4. Essener Kolloquium über "Sprach-kontakt, Sprachwandel, Sprachwechsel, Sprachtod" vom 9.10.-10.10.1987 an der Universität Essen*, N. Boretzky, W. Enninger & T. Stolz (eds), 193–210. Bochum: Brockmeyer.
- Muysken, P. C. 2008. Creole studies and multilingualism. In *Handbook of Pidgin and Creole Studies*, S. Kouwenberg & J. V. Singler (eds), 287–308. Oxford: Wiley-Blackwell.
- Nakhleh, L., Ringe, D. & Warnow, T. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81(2): 382–420.
- Nichols, J. & Warnow, T. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5): 760–820.



- Ogden, T. H. & Rosenberg, M. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Molecular Phylogenetics and Evolution* 42: 817–826.
- Page, R. D. M. & Holmes, E. C. 1998. *Molecular Evolution: A Phylogenetic Approach*. Oxford: Blackwell.
- Parkvall, M. 2008. The simplicity of creoles in a cross-linguistic perspective. In *Language Complexity. Typology, Contact, Change* [Studies in Language Companion Series 94], M. Miestamo, K. Sinnemäki & F. Karlsson (eds), 265–285. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.94>
- Patrick, P. L. 2007. Introduction. In Holm & Patrick (eds), xi–xii.
- Patterson, C. 1988. The impact of evolutionary theories on systematics. In *Prospects in Systematics*, D. L. Hawksworth (ed.), 59–91. Oxford: Clarendon Press.
- Penny, D. 2013. Rewriting evolution – “Been there, done that”. *Genome Biology and Evolution* 5(5): 819–821. <https://doi.org/10.1093/gbe/evto48>.
- Ross, M. 1997. Social networks and kinds of speech-community event. In *Archaeology and Language* R. M. Blench & M. Spriggs (eds), 209–261. London: Routledge.
- Ross, M. 2007. Calquing and metatypy. *Journal of Language Contact* 1(1): 11–143.
- Schneider, E. W. 1990. The cline of creoleness in English-oriented creoles and semi-creoles of the Caribbean. *English World-Wide* 11(1): 79–113. <https://doi.org/10.1075/eww.11.1.07sch>
- Singler, J. V. (ed.). 1990. *Pidgin and Creole Tense/Mood/Aspect Systems* [Creole Language Library 6]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.6>
- Thomason, S. G. 2001. *Language Contact*. Washington DC: Georgetown University Press.
- Vandamme, A.-M. 2009. Basic concepts of molecular evolution. In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edn, P. Lemey, M. Salemi & A.-M. Vandamme (eds), 3–29. Cambridge: Cambridge University Press.
- Wheeler, W. C. & Giribet, G. 2016. Molecular data in systematics: a promise fulfilled, a future beckoning. In *The future of phylogenetic systematics: The legacy of Willi Hennig*, D. Williams, M. Schmitt & Q. Wheeler (eds), 329–343. Cambridge: CUP.
- Wiens, J. J. 2003. Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a missing data problem? *Journal of Vertebrate Paleontology* 23(2): 297–310.
- Wiens, J. J. & Moen, D. S. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* 46(3): 307–314.
- Wiley, E. O. & Lieberman, B. S. 2011. *Phylogenetics. Theory and Practice of Phylogenetic Systematics*, 2nd edn. Hoboken NJ: Wiley-Blackwell.
- Wilkinson, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* 44: 501–514.
- Williams, D. & Ebach, M. C. 2014. Patterson’s curse, molecular homology, and the data matrix. In *The Evolution of Phylogenetic Systematics*, A. Hamilton (ed.), 151–188. Berkeley CA: University of California Press.
- Winford, D. 2003. *An Introduction to Contact Linguistics*. Oxford: Wiley-Blackwell.
- Winford, D. 2008. Atlantic creole syntax. In *Handbook of Pidgin and Creole Studies*, S. Kouwenberg & J. V. Singler (eds), 19–47. Oxford: Wiley-Blackwell.

PART 2

# Linguistic areas



# Separating layers of information

## The anatomy of contact zones

Rik van Gijn

Leiden University

Linguistic areas, or *Sprachbünde*, can be described very broadly as geographical areas where a group of languages have become similar to each other because of prolonged contact between the speakers of the different languages. Numerous linguistic areas have been proposed over the years and all over the globe, and the study of linguistic areas has become an integral part of the more general study of contact linguistics and contact-induced change. Nevertheless, linguistic areas are notoriously hard to define in a consistent and meaningful way, and some scholars have even suggested abandoning the notion altogether. In this paper, I argue against abandoning the notion of linguistic area as such, but I am in favor of changing the approach or the procedure we follow in detecting or confirming linguistic areas. I would like to introduce a conceptual change, moving away from the idea of a linguistic area as a well-circumscribed geographical area towards an idea of contact zones as windows on the past. The procedure I propose is based on what may be called the anatomy of linguistic areas, consisting of a geographical, cultural-historical, communicative, and (structural-)linguistic layer. An approach to a linguistic area can start in any of these layers, but crucially refers to the others as well. One research set-up that is made possible by this approach is to set up a hypothesis-test procedure whereby one or more layers predict the others.

**Keywords:** linguistic areas, language contact, Sprachbund

### 1. Introduction

Since Trubetskoj's (1928) coinage of the term Sprachbund, numerous studies have appeared describing potential *Sprachbünde* or linguistic areas.<sup>1</sup> The linguistic areas suggested by these different studies vary considerably from each other in almost

---

1. I use these terms interchangeably.

all relevant aspects: the number and type of shared features, the degree of overlap between ‘isoglosses’, size, and time depth. It has been pointed out by various authors that this range of variation, and the lack of generalizable principles for defining linguistic areas, is an obviously serious problem for the study of the phenomenon. In fact, some specialists (e.g., Campbell 2006) suggest that we should abandon the practice of defining linguistic areas altogether, or (Stolz 2002, 2006) at least abandon the suggestion of a linguistic area as a neatly describable object in physical space.

Although subscribing to most criticisms brought forward by authors such as Dahl (2001), Campbell (2006), and Stolz (2002, 2006), I would like to argue for a way to approach linguistic-area like situations which hopefully deals with the points of critique in an adequate manner, without losing the possibility of studying contact zones.

In what follows I will briefly summarize criticisms of the way linguistic areas have been approached in the past (Section 2), before moving to a, in my view, more fruitful approach of linguistic areas (Section 3). Section 4 summarizes the points made in this paper.

## 2. Problems with linguistic areas

Approaches to linguistic areas have been heavily criticized. The main problem is that there seems to be no way to come to a cross-situationally applicable definition.<sup>2</sup> A succinct, often cited definition, comes from Thomason (2001: 99), who defines a linguistic area as “a geographical region containing a group of three or more languages that share some structural features as a result of contact.” Four aspects are posited as central to linguistic areas:

1. Linguistic areas can be described as a geographical region.
2. There are more than two languages involved (not in this definition, but often required: these should not be closely related).
3. More than one structural feature should be shared between the languages in question.
4. These shared features are the result of contact.

All of these aspects are problematic in their own right, and have been criticized. I will briefly summarize some of the problems raised for these four points.

---

2. See Campbell (2006: 3–7) for an excellent overview of proposed definitions.

## 2.1 Geography-related problems

Given the fact that shared features due to contact presuppose communicative contact, it seems very reasonable to assume the geographical proximity of the speakers of the different languages that form a linguistic area, especially since we are mostly talking about times predating modern communication technologies. Nevertheless, Campbell (2006) mentions the case of what he calls “vertical” contact, where ancestral languages have influenced extant languages (e.g., Latin influence on Spanish and French). Another case of vertical contact is when two languages are spoken in exactly the same location (e.g., in cases of community-wide bilingualism). Towards the other end, Campbell mentions that sometimes contact arises between non-neighbours, e.g., as the result of long-distance trade or conquest. Note, however, that all of these cases still require geographical space and proximity for contact to take place at some moment in time (which does not necessarily correspond to the synchronic locations of language speakers).

A related problem to do with the geographical dimension is what may be termed the boundary problem: there seems to be no objective way to establish the outer boundaries of a linguistic area when based on the distribution of linguistic features. Linguistic features do not diffuse *en bloc* and different features may have different areal distributions. This means that each language that is part of a putative linguistic area is likely to have a different set of contact-induced features, and some languages may have more than others. Where can we draw the outer boundary of a linguistic area? This also relates to the problem of features, discussed below. Note, moreover, that languages that conform only to a very limited extent to the areal prototype are not necessarily situated at the fringes of a putative linguistic area (Dahl 2001: 1058), so that perhaps sometimes boundaries should be drawn around area-internal islands.

## 2.2 Language-related problems

Thomason’s (2001) definition requires the presence of at least three languages in a linguistic area, because otherwise the notion of linguistic area would be trivialized to meaning any contact situation between two or more languages, and what makes linguistic areas special is exactly the fact that the contact effects go beyond pairs of languages. For Campbell (2006) this is an arbitrary cut-off point, because to him, linguistic areas are not principally distinct from any other language contact situation, including one between two varieties. This, however, seems to throw the baby out with the bathwater in the sense that it loses the connection between geography and contact, to which I will come back below. Of course one can say that a linguistic

area is an unusually high accumulation of bilateral contact situations, but if you do that, the traditional problems of defining a linguistic area resurface (where is the boundary, what language or language pair belongs to the linguistic area, are the same features involved or different ones, etc.?).

Another issue relating to languages not part of Thomason's definition, but often mentioned in definitions of linguistic areas (see e.g., Emeneau 1980: 124; Enfield 2005: 190) is that the languages comprising a linguistic area represent different language families, or at least distantly related ones. This requirement does not seem to be a principled one, given that contact between speakers of different varieties can and does take place between speakers of related languages (see e.g., Bowerman 2013), as is evidenced, for instance, by the Balkan Sprachbund (Sandfeld 1930), the best-known linguistic area, which contains genealogically related languages. Two points can be made in favor of the non-relatedness requirement. First, one can exclude common inheritance as a cause for similarity (see below), and second, as Dahl (2001: 1457) puts it, it is more spectacular if a feature crosses a family boundary, as it is a greater hurdle to overcome than is a boundary between related languages. For these reasons, I will focus on linguistic areas involving non-related languages.

### 2.3 Problems related to features

Several difficulties arise in relation to the shared features of a linguistic area. First, a very basic problem refers to the minimum number of features required to be able to speak of a linguistic area. The only principled answer that can be given, as argued by Campbell (2006) is one, since zero can obviously not be indicative of any contact situation, and any number above 1 is arbitrary. Nevertheless, it is clear that the more the shared features, the stronger the proposal for a linguistic area becomes: a situation where a number of languages share a single feature can be due to chance, whereas this becomes increasingly unlikely when the number of shared features increases. The main problem is that we do not have a default expectation for areas when it comes to the number of shared features.

Another difficulty is the distribution of linguistic features. In an ideal case, the isoglosses of different shared linguistic features coincide, but this is hardly ever the case. Each linguistic feature or pattern has its own history, and these histories do not need to coincide. This raises the question which feature(s) should be taken as diagnostic, and it thus feeds into the boundary problem mentioned above.

Another point related to features is the question what are and what are not good features when it comes to diagnosing contact-induced effects? I come back to this below when discussing the difficulties associated with 'proving' that contact is responsible for the shared features, but here I will mention a perhaps deeper

problem: how refined should a feature be in order to be a telling diagnostic of a history of contact? Surely, for a feature with two possible values, e.g., ‘present’ versus ‘absent,’ there is a large chance factor that might influence the patterns, while using a handful of such general features to characterize a linguistic area seems far removed from the more specific features (see e.g., Joseph 2010 for an overview) of the Balkan Sprachbund. Moreover, when it comes to contact-induced innovations, we know that they don’t usually spread through the entire language immediately, but rather creep in construction by construction or word by word. For example, two languages may be classified as ergative, but display ergativity in different parts of the language, making it less likely that one of the two is modelled on the other.

Finally, Stolz (2006: 39ff), discussing definiteness marking in the European linguistic area, raises the point that one can look at the marking patterns from different angles: marking patterns (whether languages have a marked indefinite or not), hosting word-class (whether definiteness is marked on the head noun or on attributes), morphological status (whether definiteness markers are free or bound), linearization (the order of the definiteness marker and the word it associates with morphosyntactically). Depending on which perspective you take, you will get rather different isoglosses.

## 2.4 Making the case for contact

It is not always easy to make the case for contact as the reason why languages share features. The chief difficulty is that there are other possibly influential factors such as inheritance from a shared ancestor, universally preferred structures, language-internal dependencies, and chance. This is not to say that shared inheritance, universally preferred structures, and language-internal dependencies cannot be areal patterns. Retention of features can be stimulated by areal patterns, universally preferred structures can also spread through contact, and dependency chains can be set in motion and promoted through contact. It is just that in these cases it is harder to isolate the contact effect.

One suggestion that is often made is that shared features are more likely to be due to contact if members of a language family spoken within a linguistic area have a particular characteristic, which the other members of the family lack (or vice versa). In this sense negative data can be informative as well (see also Masica 2001). For this to work, however, one needs to have the fortunate circumstance that language families are represented in sufficient numbers both outside and inside the putative linguistic area. If, on the other hand, the proposed linguistic area contains mainly isolates and/or only one or two members per language family, or if most families are spoken entirely within the linguistic area, this diagnostic criterion is no longer available or not as strongly.



One can resort to arguing that the shared features are typically those that easily diffuse through contact, but in spite of various suggestions for generalizations (e.g., Weinreich 1953; Matras 2007; Parkvall 2008; Wichmann & Holman 2009), there seems to be no clear consensus on this point yet, rendering Thomason & Kaufman (1988: 14) oft-cited claim that “as far as the strictly linguistic possibilities go, any linguistic feature can be transferred from any language to any other language” still valid.

In the end, the only truly reliable evidence for contact-induced change is when the historical process is documented from beginning to end, which is unattainable. One can, however, try to come as close to a complete picture as the circumstances allow. This means looking beyond the linguistic picture (as argued for by Campbell 2006 in his distinction between circumstantialist versus historicist approaches to linguistic areas).

### 3. The layers of a potential contact zone

These and other problems associated with linguistic areas, have led a number of researchers to abandon the concept. Campbell (2006) argues that there is no fundamental distinction between language contact between two groups and a linguistic area. Matras (2011: 157) likewise says that “linguistic areas are simply cases of convergence that catch our attention because of the density of shared isoglosses in a multiplicity of languages”. Stolz (2006: 46) is perhaps most explicit in his judgment:

Thus, one should either strip the term of its unwelcome and much too suggestive connotations or abolish it for good (but it should be kept in the virtual museum of linguistic thought as an example of how difficulties and misunderstandings can be created via terminology).

The essential problem seems to be that linguistic areas are not physical entities waiting for linguists to discover them, but rather mental constructs. Consequently, the question whether or not a particular geographical zone is a linguistic area or not is a misguided one. We should rather focus on the question “What happened?” (Campbell 2006).

With so many fundamental problems, it comes as no surprise that some linguists call for abandoning the term. On the other hand, fundamental problems are part and parcel of just about any linguistic construct, and still they have their usefulness. The question for any construct is what is it useful for. So why do we need linguistic areas?

Apart from serving as an ancillary discipline, like for example, for typology (for sampling purposes), descriptive linguistics (just like the study of related languages is part of the preparation of a field project, so should an areal study), and historical linguistics (to understand the contact-induced developments of members of a family), in my view linguistic areas serve a purpose of their own: they tell us something about the interaction between geography, human behavior, and language. Given this interest, it is clear that a multi-disciplinary approach is required. In this section, I will sketch the outlines of such an approach, which will hopefully get us beyond a number of the problems mentioned above.

Stolz (2006) discusses four ways to delimit linguistic areas:

- Geography: Boundaries determined by natural topography.
- History: Boundaries determined by the territories of groups with known shared cultural-political histories
- Communication: Multilingual communication practices determine the center and extent of a linguistic area
- Language: the distribution of particular linguistic features determines the extent of a linguistic area.

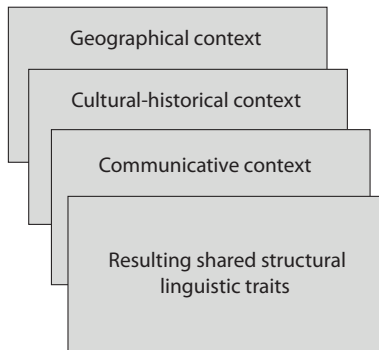
He dismisses the first three as unfeasible. His argument for all three dismissals is that there is no direct connection between the different levels, that is, they may (and are perhaps even likely to) give conflicting signals. For instance, geographical barriers do not necessarily impede inter-group communication, nor do facilitators of language contact (e.g., navigable rivers) necessarily lead to contact between groups. A shared cultural-political history does not necessarily lead to widespread bilingualism, and widespread bilingualism does not automatically lead to areal distribution of features across language boundaries. The only viable route is through the observable linguistic results of contact: the areally shared features. Given the problems associated with such an approach (see above), however, this option cannot lead to any definitive boundaries of linguistic areas.

In short, the first three approaches (geography, culture/history, and communication) are not deterministically related; the latter approach (linguistics) is not historically stable (because of e.g., immigrations and emigrations) and has too many methodological problems associated with it to adequately determine the geographical boundaries of a linguistic area.

## 4. Information layers

How to proceed then? One way to go is to give up the notion of a linguistic area altogether, as Campbell (2006) suggests, the other is to weaken the notion of linguistic area, dissociating it from any ideas about definite boundaries (Stolz 2006), a third option, which seems to be gaining ground (Dahl 2001; Muysken 2008) is areal linguistics, understood as the study of the distribution of features independently of geographical, cultural, or communicative contexts.

In this contribution, I would like to advocate an approach that takes as many of the layers mentioned by Stolz (2006) into account, and that starts out as a fully deterministic hypothesis, whereby one layer (e.g., geography) fully explains another (e.g., cultural contact), which in turn fully explains communicative behavior, and finally the distribution of shared features. It is clear that such a model will not be able to explain many of the facts, but the function of the approach would be to uncover those aspects that cannot be explained by a deterministic hypothesis, after which alternative hypotheses can be explored for unexplained facts. The layered structure is schematized in Figure 1, and can be called the anatomy of a contact zone.



**Figure 1.** The anatomy of a contact zone

In the remainder of this contribution, I will outline some of the potential information sources for each layer of a contact zone shown in Figure 1, which could play a role in such a hypothesis-driven approach to linguistic areas, starting with geography (4.1), then history (4.2), communication (4.3), and language structure (4.4).

### 4.1 Geography

Physical topography influences population movement – especially historically – and therefore contact possibilities between groups. On a basic level, one can distinguish between facilitating and impeding factors. This is in itself not a simple

distinction. Waterways, for instance, can be facilitators (e.g., in the Amazon, rivers are often the principal routes for people movement – see e.g., Arias et al. 2018), but they can also be barriers (e.g., seas, very wide rivers). The list below, which mentions some of the geographical dimensions that have been identified as important for population movement and inter-group contact, is based in particular on an overview paper by Greenhill (2014):

- Physical barriers to human contact, like (non-navigable) water masses, dense jungle (Diller 2008), desert, marshes, mountains, and simply travelling distance (leading to a clinal pattern).
- Physical pathways for human movement and thus facilitators of contact. Interestingly, water can also be a pathway of movement, if it is navigable, given the technologies of the societies. Likewise, plateaus on mountains, passes, or manmade trails (e.g., the Inca trails) would all count as pathways.
- Continental axis. May be of influence in the sense that it is easier to move crops from east to west and vice versa than from north to south, so that continents that have a long east-west axis facilitate expansion, whereas north-south axis continents do not, see also Güldemann & Hammarström (in press).
- Biodiversity. As one gets closer to the equator languages tend to spread out less (Mace & Pagel 1995; Collard & Foley 2002). This may be related to the fact that species and biodiversity increases around the equator (see e.g., Nettle & Romaine 2000), which in turn creates ecological niches in which different populations can survive side by side on a relatively small territory (Nettle 1999), such a situation may favor a loosely organized exchange system (Dixon 1997).

The list is clearly not exhaustive, but it serves its purpose in this paper of showing the potential of geographical information to generate predictions on where linguistic areas are to be expected, namely in areas where there is geographical incentive for contact (where populations are connected by pathways, especially within the same climate zone, and in places where there is less biodiversity).

## 4.2 Socio-cultural history

The role of this information layer is to establish or reconstruct the extent of contact between different groups we know have inhabited the area under study.

The most direct sources for historical information are historical records, which should certainly be exploited if they are available. Of course the interpretation of historical documents can be problematic, as authors have their own agendas, biases, observational powers, selection methods, etc. Nevertheless, historical records are (ideally) eyewitness accounts, while most other sources are indirect. Unfortunately, however, the availability of such documents reporting on contact situations are the

exception rather than the rule – this is what makes linguistic areas for which we have more historical information, like the Balkan Sprachbund (see Lindstedt 2000 for a discussion of socio-history) so valuable.

Ethnological and archaeological information in the sense of the shared use of cultural practices and material culture may indicate traces of contact. Cultural areas have experienced a long-standing interest, in particular in the North American tradition, where cultural areas often formed the basis for a linguistic area hypothesis (see the overview in Mithun 1999: 314–321). Cultural areas, however, are obviously subject to the same types of criticism as linguistic areas are. In fact, cultural practices seem to be even less subject to locality conditions, as they seem to travel more easily from community to community. Nevertheless, couched in a context of further disciplines, ethnological (and archaeological) information can provide powerful insights for contact history (see e.g., Eriksen 2011; Muysken 2012).

A powerful indirect source for the existence of historical contact is population genetics. Past contact between groups can be established by comparing the genetic and the linguistic affiliation. If there is a mismatch between the linguistic and genetic affiliation, it is indicative of contact (note, however, that if there is no mismatch, it does not follow that there was no contact). This pattern may be indicative of contact in two different situations: the situation of language shift and the situation of language maintenance (Thomason & Kaufman 1988). Moreover, maternally transmitted mtDNA and the paternally transmitted Y-chromosome may tell us something about the roles of the different sexes in the contact (Pakendorf 2014).

Again, much more could be said about ways to establish that there has been contact between ethnolinguistic groups, but it is clear that the establishment of contact forms a separate layer of information.

### 4.3 Communicative practices

Contact between different ethnolinguistic groups is common, in some places ubiquitous, but the type of contact that leads to shared grammatical characteristics (as is the case with linguistic areas) is less common. In their seminal monograph, Thomason & Kaufman (1988) claim that the extent to which lasting contact-induced effects are visible in a language is a function of time. Peripheral vocabulary is borrowed rather easily, but – under a scenario of language maintenance – core vocabulary and language structure are borrowed only after sustained intensive contact. In other words, for linguistic areas to arise, a multilingual communication practice must be in place for an extended period of time.

Unfortunately, in the absence of eyewitness accounts (important exceptions are studies like Gumperz & Wilson 1971 on Kupwar village and Aikhenvald 2002 on the Vaupés) it is extremely hard to make the case for sustained multilingual

communication. One way to show that there has been some linguistic contact is by examining the extent to which languages in an area have borrowed words from each other. The added advantage of lexical research is that the semantics of words can give insights into the nature of the contact (e.g., they can be indicative of specialized trade relations) and the extent to which loanwords have been adapted to the native phonology and grammar may tell us something about relative time depth (certainly no absolute numbers).

Another, more programmatic, way forward, proposed by Muysken (2010) is the “scenarios approach” (term due to Muysken 2013), which advocates combining the results from studies of contemporary language contact situations with the results from historical areal linguistics. He proposes to do this by studying language contact situations at varying time depths and at various levels of spatial extension, and catalogue the regularities in the outcomes. Muysken discusses four levels of aggregation (see also Muysken 2008): the bilingual individual, the bilingual community, geographical region (linguistic area), and large parts of the world (macro areas).

The central notion to be compared across these dimensions is the notion of scenario, defined as “the organized fashion in which multilingual speakers, in certain social settings, deal with the various languages in their repertoire” (Muysken 2010: 267). By cataloguing and comparing the patterns found in the linguistic effects in contact situations at various levels of aggregation, we can establish the extent to which insights from one level are applicable to the next. This would open up a possibility to make inferences with respect to communicative situations that have led to particular linguistic areas based on insights from psycholinguistics and sociolinguistics.<sup>3</sup>

#### 4.4 Structural features

From the discussion of problems associated with linguistic areas above, it should have become clear that the establishment of a linguistic area solely on the basis of bundled isoglosses as a criterion is highly problematic. One of the responses to these problems is to let go of the geographically delimited contact areas, and to look at the areal distribution of individual linguistic features (see Nichols 1992; Dahl 2001). In other words, to go from linguistic areas to areal linguistics (Muysken 2008). Areas of higher quantities of shared features would then be more or less the

---

3. See Muysken (2013) for a more concrete proposal for a model based on “bilingual strategies”, and see Kihm (2011) for an application of a scenario-like approach to Pidgin and Creole formation which he connects both to bilingual strategies (untutored second language acquisition) and historical developments comparable to the ones seen in linguistic areas (I thank Tonjes Veenstra for drawing my attention to this paper).

equivalent of the linguistic area, but without fussing over the exact borders and about which languages belong to a linguistic area and which do not.

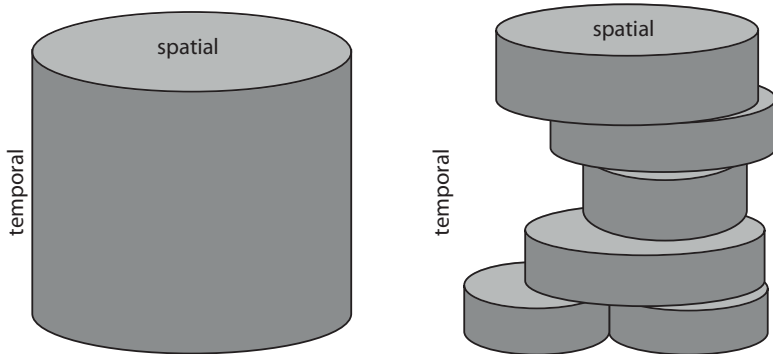
The advantage of such an approach is that the structural-linguistic information layer is studied in its own right, and – if it is carried out globally or for larger areas – it creates a direct index for contact zones in terms of number of shared features, because the same set of features is used for the entire area. However, as an approach to reconstructing the past it is incomplete by itself since it is disconnected from the historical and social dimensions.

Moreover, many of the objections relating to features discussed above are still valid for areal linguistic approaches (see Dahl 2001 for an overview of challenges for areal linguistics). First, we have seen that the problem of determining the boundaries of a linguistic area based on the distribution of linguistic features is problem-ridden with no solution in sight. Therefore, I suggest that if one must set areal boundaries, to do that on the basis of the geographical layer.

Second, it is not clear what constitute “good” features for determining a linguistic area. This is a very difficult problem, to which no straightforward answer is possible. In general terms, however, it is obvious that we need to move away from coarsely defined features in the area of grammar (e.g., “presence versus absence of X”) and move towards a more refined coding of features. The most promising line of research in this respect that I am aware of is multivariate typology (Bickel 2010, 2011), which aims to move away from predefined major comparative categories (subject, ergative, subordination, nominalization, etc.) and towards cataloguing the variables which relate to these categorizations in individual languages. By proceeding in a bottom-up fashion, allowing the set of variables to grow with each language that one investigates, one can show (and quantify) the degree to which e.g., the “subject” in language A is similar to or differs from “subject” in language B. This approach is potentially fruitful for contact research, because (a) it allows for detecting more sensitive contact signals and (b) it offers a solution to the perspective problem (the question of which sub-aspects of features are highlighted has consequences for the isoglosses) raised by Stolz (2006) discussed in Section 2 above.

#### 4.5 Non-isomorphism, bottom-up, and top-down

The links between ecological, socio-economic factors and linguistic outcomes are not necessarily direct ones (cf. e.g., Stolz 2006; Greenhill 2014; Pakendorf 2014), which may lead to conflicting signals (see Figure 2). We know that linguistic areas (or the accumulation of shared features in an area) are the result of human behavior in a particular area over time and we know that that human behavior is hard to predict.



**Figure 2.** Two conceptualizations of a “linguistic area”

Rather than dismissing geography, socio-history, or communicative behavior as possible determinants of the extent of linguistic areas (Stolz 2006), I propose to regard the non-isomorphism of different signals as informative. In other words, to embrace a concept of “linguistic area” as not necessarily cohesive (the one to the right rather than the one to the left in Figure 2).

In this way we can move away from the question of what exactly the boundaries of a linguistic area are, and which language does or does not belong to it, and move towards the question “what happened?” (Campbell 2006).

Bickel and Nichols (2006) propose an approach to (large) linguistic areas based on what they call Predictive Areality Theory (PAT): “We define an area based on a theory of population and language spread and on information from other disciplines; hypothesize that it is a linguistic area; and test the hypothesis by seeking statistically non-accidental signals. (...) For a PAT to work, it must be grounded in what we know about population history from archaeology, genetics, ecology, geography, economics, demography, etc.” In other words: Bickel and Nichols (2006) – see also Bickel (in press) – argue from non-linguistic towards linguistic data. One interesting aspect of their proposal is that, if one separates the linguistic from the non-linguistic component, one can set up a testable hypothesis, where one set of data predicts another.

Another proposal in the literature related to the one introduced in this contribution is the suggestion by Muysken (2008) and Muysken et al. (2015) that linguistic areas can be approached bottom-up or top-down. In a bottom-up approach, one first establishes an area, e.g., on the basis of shared cultural traits, and starts looking for similarities between languages. The top-down approach involves the use of an independent set of features which are checked for whether they show areal bias (the areal typology approach). Using an independent set of features is, in



my opinion, the methodologically superior choice when it comes to establishing linguistic areas, as it avoids bias and opportunism (more area-specific features can be looked for after the establishment of a significant areal pattern). And in fact, Bickel & Nichols (2006), who after all advocate a type of bottom-up approach, show that that does not exclude the use of an independent set of features. The important point raised by Muysken (2008) for the present proposal, however, is that one can go in both directions.

Returning to the layered anatomy of linguistic areas introduced in the beginning of Section 4, we can generalize the insights from Bickel & Nichols (2006) and Muysken (2008), Muysken et al. (2015). Since each layer in Figure 1 constitutes a separate layer, each of them (or combinations of them) can in principle be used as a basis for a linguistic area hypothesis (or put more generally: a hypothesis for a zone of increased language contact), in line with Bickel & Nichols' (2006) proposal. This hypothesis can subsequently be corroborated or rejected on the basis of information from the other layers, based on the simplistic idea that there is a deterministic relation, which is as follows:

Geography → Socio-history → Communication → Shared structural features

In case of non-isomorphism, new hypotheses can be created to explain the deviations from the deterministic view. For example, if a feature spills over the boundaries set by the geography layer, this may point to emigrations out of the zone, and vice versa a “non-cohering” language inside a predicted contact zone may point towards a recent immigration. Evidence from the sociohistorical layer pointing towards contact but without there being much evidence in terms of shared structural features may tell us something about the historical communicative situation, etc. In general terms, the more different types of information are taken into account, the more complete the picture becomes.

One of the consequences of this approach to linguistic areas or, broader, contact zones, is that linguistic areas as a whole are no longer more or less clearly delimited areas in space, but a conglomerate of information layers that overlap, but need not be isomorphic. Moreover, if we start out with a deterministic position, and isolate the elements that do not conform to that hypothesis as explananda, we shift our focus to the question “What happened?” rather than where the boundaries are and which languages do and which languages do not belong to a particular linguistic area.

In fact, even the information within a single layer can yield conflicting signals, related to the temporal dimension of contact zones. Contact may not be a stable factor in a particular area, but rather be more or less strong in different eras. This would lead to e.g., multi-layered genetic, archaeological, ethnological, and linguistic signals, which can contribute to the reconstruction of past contact situations.

Finally, Muysken (2008) mentions the dimension of time as a special problem for linguistic areas. Proposals for linguistic areas range from the relatively small Vaupés area (Aikhenvald 2002) to (almost) continent-sized, deep-time linguistic areas – e.g., Australia (Dixon 2001), Africa (Heine & Leyew 2008), Mainland Southeast Asia (Enfield 2005), Europe (Haspelmath 2001), the Amazon (Dixon & Aikhenvald 1999), and even areas that stretch beyond a single continent, like the Pacific Rim (Bickel & Nichols 2006). Although, as Bickel & Nichols (2006) show, very large and ancient areas can still be approached in a way similar to the one proposed here, it is true that the signals are likely to involve a greater number of conflicts and may therefore be harder to interpret. Moreover, as Muysken (2008) mentions, the communication or contact scenario (i.e., the mechanism with which a contact zone arises) becomes very vague or even uninformative. Nevertheless, I think it is important to try and form hypotheses about contact zones large and small, also because the bigger areas form the background against which smaller, younger, and perhaps more intense contact zones can be detected (Aikhenvald 2011: 15–16), as long as we realize that larger linguistic areas generally offer less clear signals, at least with respect to the mechanisms that give rise to a linguistic area than smaller ones, and are therefore weaker as hypotheses.

## 5. Conclusion

In this paper, I have argued that, in spite of the many difficulties associated with it, the concept of linguistic area has its usefulness in allowing for the study of the interaction between geography, human behavior, and language. I do agree with Stolz (2006) that we should abandon the idea of a well-circumscribed, internally coherent phenomenon, waiting to be discovered by astute linguists. Rather, we need to move away from questions of determining the exact boundaries of a linguistic area and which languages do and which languages do not belong to it, and move towards the question “What happened?” (Campbell 2006).

The approach proposed here involves the separation of different information layers that each tell us something about what happened, so that one can set up an initial hypothesis, which can then be evaluated against the information from the other layers. If we start out from a simplified deterministic idea, we can identify the cases that require explanation, and formulate new, more precise hypotheses. The layered structure can also be used to evaluate existing proposals for contact zones in respect of their completeness.

## Acknowledgements

I would like to thank two anonymous reviewers for their comments and helpful suggestions. Remaining errors are mine.

## Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 818854 – SAPPHIRE) and from the Swiss National Science Foundation (Grant No. CRSIII\_160739 – “Linguistic morphology in time and space”). Their support is gratefully acknowledged.

## References

- Aikhenvald, A. Y. 2002. *Language Contact in Amazonia*. Oxford: OUP.
- Aikhenvald, A. Y. 2011. Areal features and linguistic areas: Contact-induced change and geographical typology. In *Geographical Typology and Linguistic Areas*, [Tokyo University of Foreign Studies 2]. O. Hieda, C. König & H. Nakagawa (eds), 13–40. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/tufs.2>
- Arias, L., Barbieri, C., Barreto, G., Pakendorf, B. & Stoneking, M. 2018. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in northwestern Amazonia. *American Journal of Physical Anthropology* 165(2): 238–255.
- Bickel, B. 2010. Capturing particulars and universals in clause linkage: A multivariate analysis. In *Clause Linking and Clause Hierarchy: Syntax and Pragmatics* [Studies in Language Companion Series 121], I. Brill (ed.), 51–102. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/slcs.121>
- Bickel, B. 2011. Multivariate typology and field linguistics: A case study on detransitivization in Kiranti (Sino-Tibetan). In *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, P. K. Austin, O. Bond, D. Nathan, & L. Marten (eds), 3–13. London: SOAS.
- Bickel, B. 2020. Large and ancient linguistic areas. In *Language Dispersal, Diversification, and Contact: A Global Perspective*, E. I. Crevels, J.-M. Hombert & P. C. Muysken (eds), Oxford: OUP.
- Bickel, B. & Nichols, J. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. In *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society, Special Session on the Languages and Linguistics of Oceania*, Z. Antić, C. B. Chang, C. S. Sandy & M. Toosarvandani (eds), 3–15. Berkeley CA: BLS.
- Bowern, C. 2013. Relatedness as a factor in language contact. *Journal of Language Contact*. 6(2): 411–432.
- Campbell, L. 2006. Areal linguistics: A closer scrutiny. In *Linguistic Areas: Convergence in Historical and Typological Perspective*, Y. Matras, A. McMahon & N. Vincent (eds), 1–31. Houndmills: Palgrave MacMillan.

- Collard, I. F. & Foley, R. A. 2002. Latitudinal patterns and environmental determinants of recent human cultural diversity: Do humans follow biogeographical rules? *Evolutionary Ecology Research* 4: 371–383.
- 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), 1456–1470. Berlin: De Gruyter.
- Diller, A. 2008. Mountains, rivers or seas? Ecology and language history in Southeast Asia. In *SEALSXIV: Papers from the 14th meeting of the Southeast Asian Linguistics Society (2004)*, Vol. 1, W. Khanittanan & P. Sidwell (eds), 93–106. Canberra: Pacific Linguistics.
- Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: CUP.
- Dixon, R. M. W. 2001. The Australian linguistic area. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, A. Y. Aikhenvald & R. M. W. Dixon (eds), 64–104. Oxford: OUP.
- Dixon, R. M. W. & Aikhenvald, A. Y. (eds). 1999. *The Amazonian Languages*. Cambridge: CUP.
- Emeneau, M. B. 1980. *Language and Linguistic Area: Essays by Murray B. Emeneau, Selected and Introduced by Anwar S. Dil* [Language Science and National Development 14]. Stanford CA: Stanford University Press.
- Enfield, N. J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34:181–206.
- Eriksen, L. 2011. Nature and Culture in Prehistoric Amazonia: Using G.I.S. to Reconstruct Ancient Ethnogenetic Processes from Archaeology, Linguistics, Geography, and Ethnohistory. PhD dissertation, Lund University.
- Greenhill, S. 2014. Demographic correlates of language diversity. In *The Routledge Handbook of Historical Linguistics*, C. Bower & B. Evans (eds), 557–578. London: Routledge.
- Güldemann, T. & Hammarström, H. 2020. Geographical axis effects in large-scale linguistic distributions. In *Language Dispersal, Diversification, and Contact: A Global Perspective*, E. I. Crevels, J.-M. Hombert & P. C. Muysken (eds). Oxford: OUP.
- Gumperz, J. J. & R. Wilson. 1971. Convergence and creolization: A case from the Indo Aryan/Dravidian border in India. In *Pidginization and Creolization of Languages*, D. Hymes (ed.), 151–167. Cambridge: CUP.
- 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), 1492–1510. Berlin: de Gruyter.
- Heine, B. & Leyew, Z. 2008. Is Africa a linguistic area? In *A Linguistic Geography of Africa*, B. Heine & D. Nurse (eds), 15–35. Cambridge: CUP.
- Joseph, B. 2010. Language contact in the Balkans. In *The Handbook of Language Contact*, R. Hickey (ed.), 618–633. Oxford: Wiley-Blackwell.
- Lindstedt, J. 2000. Linguistic Balkanization: Contact-induced change by mutual reinforcement. In *Languages in Contact*, D. Gilbers, J. Nerbonne & J. Schaeken (eds), 231–261. Amsterdam: Rodopi.
- Mace, R. & Pagel, M. 1995. A latitudinal gradient in the density of human languages in North America. *Proceedings of the Royal Society B: Biological Sciences* 261(1360): 117–121.
- Masica, C. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In *The Yearbook of South Asian Languages and Linguistics 2001. Tokyo Symposium on South Asian Languages: Contact, Convergence and Typology*, R. Singh, P. Bhaskararao & K. V. Subbarao (eds), 205–267. New Delhi: Sage.

- Matras, Y. 2007. The borrowability of structural categories. In *Grammatical Borrowing in Cross-linguistic Perspective*, Y. Matras & J. Sakel (eds), 31–74. Berlin: Mouton de Gruyter.
- Matras, Y. 2011. Explaining convergence and the formation of linguistic areas. In *Geographical Typology and Linguistic Areas* [Tokyo University of Foreign Studies 2], O. Hieda, C. König & H. Nakagawa (eds), 143–160. Amsterdam: John Benjamins. <https://doi.org/10.1075/tufs.2>
- Mithun, M. 1999. *The Languages of Native North America*. Cambridge: CUP.
- Muysken, P. C. 2008. Introduction: Conceptual and methodological issues in areal linguistics. In *From Linguistic Areas to Areal Linguistics* [Studies in Language Companion Series 90], P. C. Muysken (ed.), 1–24. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.90>
- Muysken, P. C. 2010. Scenarios for language contact. In *The Handbook of Language Contact*, R. Hickey (ed.), 265–281. Oxford: Wiley-Blackwell.
- Muysken, P. C. 2012. Modelling the Quechua-Aymara relationship: Sociolinguistic scenarios and possible archaeological evidence. In *Archaeology and Language in the Andes*, P. Heggarty & D. G. Beresford-Jones (eds), 83–107. Oxford: OUP.
- Muysken, P. C., Hammarström, H., Birchall, J. T. R., van Gijn, R., Krasnoukhova, O. V. & Müller, N. J. 2015. Linguistic areas, bottom-up or top-down? The case of the Guaporé-Mamoré. In *Language Contact and Documentation*, B. Comrie & L. Golluscio (eds), 205–237. Berlin: Mouton de Gruyter.
- Nettle, D. 1999. *Linguistic Diversity*. Oxford: OUP.
- Nettle, D. & Romaine, S. 2000. *Vanishing Voices: The Extinction of the World's Languages*. Oxford: OUP.
- Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago IL: University of Chicago Press.
- Pakendorf, B. 2014. Historical linguistics and molecular anthropology. In *The Routledge Handbook of Historical Linguistics*, C. Bowern & B. Evans (eds), 627–641. London: Routledge.
- Parkvall, M. 2008. Which parts of language are the most stable? *Sprachtypologie und Universalienforschung* 61: 234–250.
- Sandfeld, K. 1930. *Linguistique balkanique: Problemes et resultats*. Paris: E. Champion.
- 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.
- Stolz, T. 2006. All or nothing. In *Linguistic Areas: Convergence in Historical and Typological Perspective*, Y. Matras, A. McMahon & N. Vincent (eds), 1–31. Houndmills: Palgrave MacMillan.
- Thomason, S. G. 2001. *Language Contact. An Introduction*. Edinburgh: EUP.
- Thomason, S. G. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- Trubetskoj, N. S. 1928. Proposition 16. In *Actes de 1er Congrès international de linguistes*, 17–18. Leiden: A.W. Sijthoff's Uitgeversmaatschappij.
- Weinreich, U. 1953. *Languages in Contact: Findings and Problems*. New York NY: Linguistic Circle of New York.
- Wichmann, S. & Holman, E. W. 2009. *Assessing Temporal Stability for Linguistic Typological Features*. Munich: Lincom.

# Areal diffusion of applicatives in the Amazon

Mily Crevels and Hein van der Voort

LUCL, Universiteit Leiden / Museu Paraense Emílio Goeldi

Many genetically unrelated languages of (North)western Amazonia share certain specific characteristics, which may be due to areal diffusion.<sup>1</sup> Especially the eclectic classifier systems encountered across this region have been in focus lately, as well as valency changing strategies, in particular applicatives involving forms resembling *-tA*, which show conspicuous similarities across (North)western and Southern Amazonia. Ongoing research in Southern Amazonia has led to the identification of a number of very similar traits and even some striking formal similarities. Even though the (North)western and Southern Amazon may not constitute a single linguistic area, certain grammatical traits such as classifier systems and valency-changing strategies can be considered as areal features of the Amazon region as a whole, and beyond. In this article we will discuss correspondences between the expressions of valency change in the (North)western and Southern Amazon regions.

**Keywords:** applicatives, valency-changing morphemes, areal diffusion, Amazonia

## 1. Background

This article is about the shared occurrence of certain specific grammatical forms combined with their specific grammatical meaning in languages of the (North)western and Southern Amazon, regions that have an extremely high indigenous linguistic diversity. Although during the past two decades many languages of these regions have been described to a reasonable extent and genealogical relationships have been established or refined, the overall picture is still fragmentary and many questions regarding language relations in these parts of the Amazon remain. Sometimes, certain lexical or morphological correspondences between languages are very strong, but not according to patterns one would expect between genealogically related languages, such as systematic sound correspondences involving basic vocabulary,

---

1. See Eriksen (2011) for a detailed overview of regional divisions of the Amazon Basin.

similar inflectional paradigms and the like. When neighboring languages do not appear to be closely related, coincidence of non-basic vocabulary items often points to lexical borrowing as a result of direct language contact. Furthermore, languages of a certain geographical region may share lexical or structural features that do not seem to be attributable to genealogical relations, but rather to long-term coexistence of their speaker communities, forming a *Sprachbund* or linguistic area. Attempts to establish language relations in the Amazon region have revealed many traces of language contact and led to proposals of several linguistic areas.

The most well-known linguistic area of South America is the Upper Rio Negro region in the Northwest Amazon, where Arawakan, Eastern Tukanoan and Naduhup languages have been in contact and influenced each other's grammatical structures (Sorensen 1967; Aikhenvald 2002; Stenzel 2005; Epps 2007; Stenzel & Gómez-Imbert 2009; Epps & Stenzel 2013). Several other regions have been proposed as linguistic areas, such as the Xingu region, considered as an incipient linguistic area by Seki (1999), where Arawakan, Cariban, Macro-Jê, and other languages are in contact, and the Chaco region, which is directly adjacent to the Amazon region, where Guaycuruan and Matacoan languages have been in contact (Comrie et al. 2010; Messineo 2011; Campbell & Grondona 2012). Areal phenomena have also been reported in the Guaporé-Mamoré river basin in the Southern Amazon (Van der Voort 2005; Crevels & Van der Voort 2008). In spite of the extreme linguistic diversity in this latter region, which is home to at least seven linguistic families/stocks and up to 10 linguistic isolates, several grammatical traits are shared across genealogical boundaries between different subsets of languages. We can mention, for example, certain possessive structures (Krasnoukhova 2012), quotative constructions (Van der Voort 2013, 2016), hierarchical alignment (Crevels 2011), and verbal number (Crevels 2006; Crevels & Van der Voort 2008), which are shared among various unrelated languages of the region. As yet no clear picture has emerged of the Guaporé-Mamoré region as a single linguistic area, but the areal diffusion of certain specific traits is unmistakable (see Muysken et al. 2015: 222). Such traits usually define specific subregions of the Guaporé-Mamoré area, while at the same time they may extend beyond the Guaporé-Mamoré area (see also Van Gijn 2014, 2015).

Some areal traits identified in the Southern Amazon are also found in the (North)western Amazon and beyond. Lexical elements, such as the words for 'bean' and 'sieve', and a word that usually refers to the giant armadillo (*Priodontes maximus*, in Spanish *tatú canasta*, in Portuguese *tatu canastra*), are highly similar and at times identical among many language isolates and families all over the Amazon region (including the Orinoco-Guiana region, see Table 1).<sup>2</sup>

---

2. This table only includes languages with words for 'bean', 'sieve' and 'giant armadillo' that show similarities. An empty slot means that the word in question is not similar at all. A slot with

Table 1. Lexical similarities in the Amazon

Southern Amazon		'Bean'	'Sieve'	'Giant armadillo' (unless indicated otherwise)	Source
Aikanā	ISOLATE	kumāda	manarɛ	marura	personal fieldwork
Kanoé	ISOLATE	kometa	mænærɛ		Bacelar (2004)
Kwaza	ISOLATE	kumada	manarə	xoronā harurai (armadillo)	personal fieldwork
Irantxe	ISOLATE	kumātaʔa		malula/mauwa	Meader (1967)
Itonama	ISOLATE			marora (armadillo sp.)	personal fieldwork
Paresi	ARAWAKAN	kumeta		malola	Rowan (2001)
Moré	CHAPACURAN	komat		manim	Angenot de Lima (2002)
Napeka/ Kitemoka				murorá (armadillo)	Riester (n.d.)
Arikapú	MACRO-JÉAN		manarə		personal fieldwork
Xavante	MACRO-JÉAN, JÊ			wārā wawē	Hall et al. (2004)
Sabanê	NAMBIKWARAN		məna:lə	mulula	De Araujo (2004)
Latundê	NAMBIKWARAN, NORTHERN	kamat	manan-	wolon-	Telles (2002)
Apiaka	TUPIAN, TUPI-GUARANI	kumanda	?	?	Padua (2007)
Tupari	TUPIAN, TUPARI		mārā:rɛ	paroro/maroro-ato	Alves (2004)
<b>(North)western Amazon</b>					
Puinave	ISOLATE	kumana	?		Haynie et al. (2014)
Makuna	TUKANOAN	kūmana			Haynie et al. (2014)
Nadëb	NADUHUP	kamaan	?	?	Haynie et al. (2014)
Arabela	ZAPAROAN			moraja (armadillo)	Rich (1999)
Tariana	ARAWAKAN	kumáda			Aikhenvald (2001)

*(continued)*

a question mark means that we have insufficient data. Not included are languages in which none of the words resemble those in the table or languages for which we have insufficient data for any of those words.



Table 1. (continued)

		‘Bean’	‘Sieve’	‘Giant armadillo’ (unless indicated otherwise)	Source
<b>Orinoco-Guiana</b>					
Tarumã	ISOLATE	?	kumáta*	marura	Sérgio Meira (p.c.)
Lokono	ARAWAKAN		manari	?	Van Baarle et al. (1989)
Palikur		kumat		malula-imá	Nimuendajú (1926)
Wapishana		komaasa	manaru	maroro	Wapishana Language Project (2000)
Tiriyo	CARIBAN	kumata	manare	moora(-imë)	Meira (1999)

\* The similarity of the Tarumã word for ‘sieve’, which in this case is a long tube-shaped braided manioc press (a.k.a. *tipiti*), with the word for ‘bean’ in the other languages is remarkable.

The origin of these and certain other widespread words (see Van der Voort 2005) cannot be traced to a specific language and can be considered as true *Wanderwörter* (see Haynie et al. 2014). The word for ‘bean’ is also attested in Tupinambá (TUPIAN, TUPI-GUARANI) as *komana*, and it may have profited from the expansion of the descendant of Tupinambá, the lingua franca Nheengatú. However, the widespread words for ‘sieve’ and ‘giant armadillo’ do not have corresponding similar forms in Tupinambá or Nheengatú. One could speculate that the word for ‘giant armadillo’ stems from an early settlement period of South America some 10.000 years ago, when the Volkswagen Beetle-size glyptodon was still knocking about, although it is wholly unlikely that it maintained the same form for such a long time.<sup>3</sup> The word for ‘sieve’ has even spread to the Caribbean region, where it was attested in the now extinct creole language Virgin Islands Creole Dutch (Oldendorp 1996 [1767–1768]: 109, item 2107), by which it was borrowed from a local indigenous language.

Also, certain grammatical forms and structures are widespread throughout the Amazon region. Although Crevels & Van der Voort (2008) have reported on a relative absence of nominal classifying constructions in the Guaporé-Mamoré

3. The majority of the words for ‘giant armadillo’ in Table 1 are quite similar and only in some cases they refer to other armadillo species. Furthermore, words for ‘giant armadillo’ often do not show any likely etymological relation to words for other armadillo species.

region, those in the languages that have them show considerable similarities.<sup>4</sup> Even though the relevant languages are genealogically unrelated, the characteristics and structures of their classifier systems are very similar and even the forms of some of the classifiers coincide (Van der Voort 2015, 2018). Furthermore, the classifier systems encountered in the Guaporé-Mamoré region are of the very same multi-functional type as described for the Northwestern Amazon by Grinevald & Seifart (2004) and Seifart & Payne (2007). Some aspects of classifying constructions are conspicuously similar in languages of both regions. One of those aspects concerns the ‘empty root’ construction that involves a semantically unspecified noun formative root, usually of similar form (*e-* or *i-*), that has the capacity to turn a bound classifier into an independent noun with the classifier’s specific semantic contents. Such constructions are encountered in various languages of the Guaporé-Mamoré region, in Panoan languages and Harakmbut (HARAKMBUT-KATUKINAN) in the Southern Amazon, as well as in Miraña (BORAN) and Muinane (WITOTOAN) in the (North)western Amazon. As Krasnoukhova (2012: 263) observes: “[T]hese two regions are indeed two separate ‘epicenters’ of multifunctional classifier systems.” The similarities between these epicenters are intriguing.

Areal linguistics in South America is still very much ‘work in progress’. Certain specific linguistic convergence phenomena found in the Upper Rio Negro linguistic area also appear in adjacent regions (Epps 2006, 2009; Seifart & Payne 2007). Recent work by Krasnoukhova (2012, 2014), Birchall (2014a,b), and Muysken et al. (2014) interpreting the clustering of multiple areal features is suggestive of a division between eastern and western South American linguistic profiles, rather than of a division between Amazonian and Andean languages. As Campbell (2006) stated:

Every ‘linguistic area’, to the extent that the notion has any meaning at all, arises from an accumulation of individual cases of ‘localized diffusion’; it is the investigation of these specific instances of diffusion, and not the pursuit of defining properties for linguistic areas, that will increase our understanding and will explain the historical facts. (Campbell 2006: 18)

In this spirit, the present article proposes to add further data to what is known about the spread of certain valency-changing strategies, in particular applicatives involving a grammatical form identical with, or similar to, *-tA* and occurring typically in a position immediately following the verbal root. Payne (1985) and Wise (1990, 1993, 2002) have noticed the existence of very similar applicative constructions and forms across the Eastern Andean slopes in the Ecuadorian and Peruvian Amazon. Recent new data from the Southern Amazon reveal very similar traits and forms.

---

4. Except for the classifier constructions in Itonama (see Crevels 2012; Van der Voort 2018).

## 2. Applicatives in the Amazon

The applicative is a grammatical construction with or without an overt applicative morpheme that typically results in promoting oblique arguments to object status, usually but not necessarily increasing the valency of an intransitive or transitive verb by one. In languages without an overt applicative morpheme, the applicative construction can be compared to the dative shift, like in English *Mily gives the book to Pieter* versus *Mily gives Pieter the book*, in which the prepositional object *Pieter* is promoted to dative object status.

The semantic role of the applied object (the object added in the applicative construction) tends to be less specified cross-linguistically and may represent several roles, such as beneficiary, concomitant, instrument (Polinsky 2013). In languages that do have an applicative morpheme, the applicative construction is often similar to other grammatical valency-changing operations such as the causative, to the extent that the causative promotes a causer to argument status, thus increasing the valency of the verb. Languages can even use the same morpheme for applicatives and causatives alike (e.g. Tuggy 1988; Comrie 1989; Song 1990; Shibatani & Pardeshi 2002; Peterson 2007). According to Shibatani and Pardeshi (2002: 116–122), this causative/applicative syncretism can be ascribed to the connection between the applicative meanings of comitative, instrumental and benefactive forms with sociative causatives, since in many cases the causer is actively participating in the event. Consider Swift's Caquinte (ARAWAKAN) examples in (1), where *-aka* appears in (1a) as a sociative causative marker, whereas it functions as an indirect causative marker in (1b).

- (1) a. *i-N-kih-aka-apoh-ak-e-ri*  
 3.M-FUT-enter-COM-AL-PFV-FUT-3.M  
 'He will make him enter with him.'
- b. *i-N-kamaraNk-aka-ak-e-ri*      *aisa ora*      *iβeNki kačatyakiri=βeNki*  
 3.M-FUT-vomit-COM-PFV-FUT-3.M also DEM.F piripiri maquisapa-piripiri  
 'He will also make him vomit *piripiri* (plant sp.) to hunt *maquisapas* (monkey sp).'
- (Swift 1988: 103)

Furthermore, the applicative is not necessarily a strictly transitivity operation, since it can also increase the valency of transitive verbs to ditransitive and it is even attested in some languages as a detransitivizing operation.<sup>5</sup> In many of the world's languages it involves the addition of an affix to the verb. In some cases the affix is

---

5. Dixon (2012, p. 294) reports that out of a sample of 82 languages with either canonical applicative or quasi-applicative, in 54 languages (66%), both intransitive and transitive verbs may enter into applicative derivations. Only in 17 languages (21%), applicative is restricted to intransitives, and for 11 (13%) it applies just to transitives.

etymologically related to a preposition, which is why Baker (1988) analyzes the applicative as preposition incorporation.

Many Amazonian languages have grammaticalized applicative constructions. Wise (2002: 338–340) points out that most applicative morphemes in Amazonian languages are suffixes, whereas she considers applicative prefixes the general pattern elsewhere in the world. Note, however, that Hawkins and Gilligan (1988: 224) claim that if a language has valency affixes on the verb, they are suffixed with more than chance frequency.<sup>6</sup>

## 2.1 (North)western Amazon

In a research note, Payne (1985) mentions the existence of transitivity morphemes in languages of the Peba-Yaguan family and the Zaparoan family of the Ecuadorian and northern Peruvian Amazon. In various languages of these families, the form of the morpheme is *-ta* and its transitivity behavior is not straightforward. In Yagua (PEBA-YAGUAN) it appears to indicate that the object of the verb is comitative or instrumental. The same element can be attached to nouns, representing a comitative or instrumental postposition. In Arabela (Zaparoan), *-ta* and its allomorph *-tia* have almost exactly the same functions and distribution (Rich 1975, per Payne 1985). In Andoa (Zaparoan), the morpheme *-ta* functions as an instrumental postposition and as a verbal transitivity morpheme (Peeke 1953, 1959, per Payne 1985). Although Payne does not explicitly identify it as such, *-ta* is obviously an applicative morpheme. Apart from certain other striking phonological and grammatical similarities between Peba-Yaguan and Zaparoan languages, there is hardly any shared lexicon. Payne (1985) speculates whether those similarities, especially concerning *-ta*, may be symptomatic of the Western Amazon as a linguistic area.

Wise (2002) takes up the issue of *-ta* in an article on applicatives and adds data from additional language families. In Shawi (KAWAPANAN, a.k.a. Chayahuita) there is an applicative suffix with two allomorphs, *-të* [ti] and *-ta*, which may verbalize nouns (Examples 2a, b), change impersonal verbs to intransitive verbs (3a, b), transitivity morphemes (4a, b), detransitivize transitive verbs (5a, b), and change transitive verbs to ditransitive (6a, b). Illustrative examples of these different types of expressions are from Hart (1988):<sup>7</sup>

---

6. Based on a 40-language computerized sample of morphological properties collected by Revere Perkins and provided by Joan Bybee (see Bybee 1985), Hawkins and Gilligan (1988) point out that 93% of the languages with valency affixes are suffixing.

7. We have chosen different examples from those already cited twice in Wise (1999, 2002). The presentation of examples in the present article follows the orthography used in the respective sources.

- (2) Shawi (KAWAPANAN; Hart 1988: 269–270)
- a. *nansë*  
bone  
'bone'
- b. bone-APPL-IND-3SG.S  
*nansë-të-r-in*  
'He/she loses weight.'
- (3) a. *të'na-r-in*  
cold-IND-3SG.S  
'It is getting cold.'
- b. *të'na-të-r-in*  
cold-APPL-IND-3SG.S  
'He is getting cold.'
- (4) a. *ama-r-in*  
bathe-IND-3SG.S  
'He/she takes a bath.'
- b. *ama-të-r-in*  
bathe-APPL-IND-3SG.S  
'He/she bathes him/her.'
- (5) a. *pichi-r-in*  
count-IND-3SG.S  
'He counts it.'
- b. *pichi-të-r-in*  
count-APPL-IND-3SG.S  
'He counts.'
- (6) a. *inqui-r-in*  
throw.away-IND-3SG.S  
'He threw it away.'
- b. *inqui-të-r-in*  
throw.away-APPL-IND-3SG.S  
'He threw it away for someone.'

Likewise, Shawi's sister language Shiwilu (a.k.a. Jebero) has an applicative morpheme *-ta* that directly follows the root and can function as a verbalizer (7) and a transitivizer (7, 8), but does not share the other functions of Shawi *-ti/-ta*.<sup>8</sup>

---

8. Valenzuela (2011) also mentions *-tu*, which has a verbalizing/transitivizing function and is glossed as a 'valency modifier'.

- (7) Shiwilu (KAWAPANAN)
- a. *tumu<sup>3</sup>-ta-pa-ti*  
trick-APPL-DUR-NFUT.3SG  
'I am fooling/fooled (someone).'
- b. *tumu<sup>3</sup>-ta-pa-tɛn*  
trick-APPL-DUR-NFUT.1SG>2SG  
'I am fooling you.' (Peperkamp 2005: 84, 38)
- (8) a. *anani-pa-lli*  
distribute-DUR-NFUT.3SG  
'S/he is distributing/sharing.'
- b. *anani-ta-pa-lli*  
distribute-APPL-DUR-NFUT.3SG  
'S/he is distributing/sharing (it).'
- c. *anani-ta-pa-dek-lli*  
distribute-APPL-DUR-3PL.O-NFUT.3SG  
'S/he is distributing (it) to them.' (Valenzuela & Butler 2009)

Wise (2002) also shows how the Zaparoan languages Arabela and Iquito both have the multifunctional applicative morpheme *-ta*. She furthermore mentions the Yaguan applicative, citing examples from Payne (1997), and notes the fact that the Yaguan morpheme has two allomorphs as well, *-ta* and *-tya*.

Other language families from the Peruvian Amazon also have applicative constructions, but the forms are different. Wise (2002: 332) cites examples from Yaminahua (PANOAN), which has different benefactive, malefactive and comitative applicative morphemes. In Yanesha' (ARAWAKAN) there are several applicative morphemes specified for benefactive, malefactive and comitative, as well as a more multifunctional applicative that can be benefactive, malefactive, sociative, change the valency of the verb and in some cases even its meaning. Nomatsiguenga and other Arawakan Campa languages have even more diverse applicative systems with specified morphemes that indicate whether the applied object represents a reason, purpose, instrument, benefactive, comitative, causative or a generic relationship ('with respect to'). According to Mihás (2015: 274), Ashéninka Alto Perené shows 'canonical applicatives', which, using Dixon's (2012: 295) terminology, means that there are alternative ways to express an oblique argument, as well as 'quasi-applicatives', which, in Dixon's (2012: 299) terminology, means that there are no underlying constructions for which the applied argument appears as an oblique, i.e., there is no corresponding non-applicative construction. Consider the following examples, in which (9) represents a canonical applicative construction with the generalized applicative *-ako*, and (10) a quasi-applicative with the presential applicative *-imo*:

- (9) Ashéninka Alto Perené (ARAWAKAN, CAMPA)  
*ari a-shee-ty-ako-t-ak-i-ri*  
 PP 1PL.S-be.afternoon-DIM-APPL-EP-PFV-REAL-REL  
 ‘That was the case that twilight came upon us.’ (Mihás 2015: 279)
- (10) *te no-n-kov-i pi-saik-imo-t-aj-e-na*  
 REAL.NEG 1SG.S-IRR-want-IRR 2.A-be.at-APPL-EP-TERM-IRR-1SG.O  
 ‘I don’t want to live with you.’ (Mihás 2015: 298)

In Ashéninka Ucayali-Pajonal, *-imo* seems to be used in the same way:

- (11) Ashéninka Ucayali-Pajonal (ARAWAKAN, CAMPA)
- a. *i=may-imo-tz-i-ri*  
 3.M.S=sleep-COM-EP-FRS=3M.O  
 ‘He sleeps with him.’
- b. *r=oy-imo-tz-i-ri*  
 3.M.S=eat-COM-EP-FRS=3M.O  
 ‘He eats with him.’ (Cacique Coronado 2016: 55; Toni Pedrós, p.c. 2017)

Note that (*i*)*mo-* is used as a causative prefix in Baure (ARAWAKAN) (12), Tupi-Guarani languages, and other Tupi languages, such as Mundurukú (see Section 2.2).

- (12) Baure (ARAWAKAN; Danielsen 2007: 251)  
*pimokokavi*  
 pi=**imo**-koka-wo=pi  
 2SG=CAUS-laugh-COP=2SG  
 ‘You make me laugh.’

Although the Panoan and Arawakan languages do not share the applicative form *-ta*, the constructions and the multiplicity of the semantic roles involved are reminiscent of the Cahuapanan, Peba-Yaguan and Zaparoan systems. Interestingly, Payne (2001: 487) mentions two lexical/derivational causative affixes in Ashéninka (ARAWAKAN), of which one, *-(t)ag* is a suffix occurring on perhaps 30 verbs, e.g. *iyó-/iyotag-* ‘know, learn/teach’, *pey-/peyag-* ‘disappear, die, convert (INTR)/kill, make disappear, convert’. Payne (2001: 502) emphasizes that Wise (1990) already showed that “all Maipuran languages have some kind of reflex of a *\*-ta* causative (most likely with an aspirated [t] [...]).” Another example would be the causative marker *-(o)tal/-(o)tu* in the critically endangered Arawakan Resigaró language (Seifart 2012).

Wise (2002: 341) notes that there are further similarities with the causative morpheme *-ta* in Witoto (WITOTOAN). Furthermore, the fact that many Arawakan languages have a causative *-da/-ta* and/or an epenthetic, verbalizing, or transitive suffix, which includes a *-t*, points at a wide-spread grammatical feature, like the ones mentioned by Payne (1990). Wise (2002: 341) finishes by asking the

question whether this applicative feature is limited to the Northwestern Amazon, since the Panoan languages, which possibly originate from Bolivia, do not exhibit the form *-ta*.<sup>9</sup>

The Jivaroan family also shows applicatives with broad functions. Overall (2007: 306) reports an applicative suffix *-hu/-tu* in Aguaruna, whose two forms are in complementary, lexically-conditioned distribution, defining two verb classes labeled ‘*hu* verbs’ and ‘*tu* verbs’ by Overall (2007, 2017). The applied object is typically a beneficiary or maleficiary, although it may also be a goal with intransitive verbs of motion.

- (13) Aguaruna (JIVAROAN; Overall 2017)  
*ámina balán iŋkithámkamĩ*  
 ami=na bala=na iŋki-**tu**-hama-ka-mĩ  
 2SG=ACC bullet=ACC put.in-APPL-2.O-PFV-RPST:3:DECL  
 ‘He’s loaded his gun (to shoot) you.’ (Overall 2017: 303)
- (14) *wíi numín tsupíŋkathimi*  
 wíi numi=na tsupi-**hu**-ka-ta-himi-i  
 1SG wood=ACC cut-APPL-PFV-IFUT-1>2:PL-DECL  
 ‘I will cut wood for you (PL).’ (Overall 2007: 320)

Interestingly, the same forms used for the applicative, *-hu* and *-tu*, are also used for the first singular object marker, which also conform to the same verb class pattern. Overall (2007) analyzes these forms as homophonous suffixes, each with its own function.

- (15) *ĩũj́áatak yuháawai*  
 ĩũj́á-**tu**-a-kũ yuha-a-wa-i  
 look.for-1SG.O-IPFV-SIM+3:SS go:PL-IPFV-3-DECL  
 ‘They’re looking for me.’ (Overall 2017: 247)
- (16) *áu tsupíŋkamĩ*  
 au tsupi-**hu**-ka-mĩ  
 DIST cut-1SG.O-PFV-RPST:3:DECL  
 ‘S/he cut me.’ (Overall 2017: 247)

---

9. Wise does not give a reference for the hypothesis of a Bolivian homeland for the Panoan family. Fleck (2013: 21), who rejects this hypothesis, does not mention where it is from either. One reviewer pointed out that it may originate from Lathrap (1970: 80). Under the assumption that the Panoan and Takanan language families are genealogically related, a southern origin is plausible (Pilar Valenzuela p.c.).



When the applicative and the first singular object marker co-occur in a word, the object suffix switches its form (Overall 2017: 247), as can be seen in Examples (17) and (18).

- (17) *mínau númi tsupíntuktahum*  
 mi-nau numi tsupi-**hu-tu**-ka-ta-humi  
 1SG-POSS tree cut-APPL-1SG.O-PFV-IMP-2PL  
 ‘Cut my wood for me.’ (Overall 2017: 247)
- (18) *miníthawai*  
 mini-**tu-hu**-a-wa-i  
 arrive-APPL-1SG.O-IPFV-3-DECL  
 ‘It’s coming towards me.’ (Overall 2007: 82)

According to Kohlberger (2013), Shiwiari, one of Aguaruna’s Jivaroan sister languages, also has two applicative morphemes, /ru/ and /tu/, which, in addition to being used in typical applicative contexts, may appear as object markers and verbalizers. Contrary to Overall, Kohlberger assumes that the identical distribution of /ru/ and /tu/ in all three cases suggests that they are one and the same morpheme.

Furthermore, the Guahiboan language Sikuani of Eastern Colombia has an applicative prefix *to-*, which is applied to both transitive and intransitive verbs (Queixalós 2014: 322). The Tupi-Guarani language Kokama-Kokamilla of Peru has a highly productive causative suffix *-ta*, which is even attached to nouns (Vallejos 2014: 269).

## 2.2 Southern Amazon

Van Gijn (2015: 113) proposes prefixed valency markers as a prime candidate for an areal Guaporé-Mamoré (GM) feature, since in the majority of GM languages valency markers are prefixed, which is surprising in the light of Hawkins and Gilligan’s (1988) observation that valency markers are usually suffixed (see footnote 2). Focusing on valency-augmenting applicatives, however, Van Gijn (2015: 114) shows that these are prefixed in only four GM languages: Leko (ISOLATE), Mosestén (MOSETENAN), Yurakaré (ISOLATE), and Itonama (ISOLATE).

In the remainder of this section we will zero in on the spread of the applicative *-tA*, which is more widespread than has been thought so far and occurs in various language families and isolates of the GM region.

The Kwaza (ISOLATE) language has several transitivizers. The morphemes *-ta*, *-tya* and the fossilized element *.te/.te* occur close to the verb root and display the same multifunctional applicative use as attested in the (North)west Amazon. Applicative *-ta* is in the first place a verbal transitivizer, as in the contrasted Examples (19a,b). The conditions for the (less frequent) occurrence of its allomorph *-tya*, illustrated by Examples (20a,b), are not clear.

- (19) Kwaza (ISOLATE; Van der Voort 2004)
- a. *oya-da-ki*  
go-1SG-DECL  
'I went./I'm going.'
- b. *zuze-wā oya-ta-da-ki*  
José-AO go-TRZ-1SG-DECL  
'I visited José.'
- (20) a. *pěřěyā-da-ki*  
speak-1SG-DECL  
'I spoke.'
- b. *pěřěyā-tya-da-ki*  
speak-TRZ-1SG-DECL  
'I spoke to him.'

It is difficult to identify a consistent specific semantic role of the applied object. It is often related to location or accompaniment, as in Examples (21a,b). Its general meaning can perhaps best be described as 'with respect to', as in (22a,b).

- (21) a. *na-ai-ryĩ*                      *elele a-ki*  
PROX-that-DIR:area very exist-DECL  
'He lives on the same side of the river / in the same place as me.'
- b. *tyaʔatye a-ta-ki*                      *maryāu-dĩ ryĩ*  
close exist-TRZ-DECL Marião-POSS-DIR:area  
'He is living close to Marião.'
- (22) a. *hiya-ki*  
fall-DECL  
'It fell.'
- b. *hiya-ta-ki*  
fall-TRZ-DECL  
'It fell in front of him.'

If the verb is already transitive, the applicative morpheme can put the existing object in focus. This may create a definiteness effect, as in (23a,b):<sup>10</sup>

- (23) a. *we-da-ki*                      *yexwa*  
fear-1SG-DECL jaguar  
'I'm afraid of dogs (in general).'
- b. *we-tya-da-ki*                      *yexwa-wā*  
fear-TRZ1SG-DECL jaguar-AO  
'I'm afraid of (this) dog.'

10. Wise (2002: 340) mentions definiteness phenomena with applicatives in other Amazonian languages. See also Polinsky (2017) for a general discussion of this in terms of 'scope'.

Alternatively, it can add a second object, as, for instance, in (24a,b) below.

The element *-ta* is a light syllable that consists of rather common phonemes. It should therefore be no surprise that it is homophonous with other morphemes, in this case the first person singular object morpheme *-ta*.<sup>11</sup> What distinguishes the applicative from the person marker is that the first behaves as a derivational morpheme and the second as an inflectional morpheme, which is signaled by main word stress placement. The ultimate syllable of the (extended) verb stem usually attracts stress, whereas the verb's inflections are as a rule not stressed, as shown by these contrasted examples:

- (24) a. *waya-'ta-ki*  
bring-TRZ-DECL  
'He is taking it there to him.'
- b. *wa'ya-ta-ki*  
bring I.O-DECL  
'He took me.'
- (25) a. *wai-di-'ta-ki*  
good-CAUS-TRZ-DECL  
'He is good to her.'
- b. *wai-'di-ta-ki*  
good-CAUS-I.O-DECL  
'It does good to me.'

In addition to *-ta* and *-tya*, there are three applicative morphemes of very limited use: *-xote*, *-dwate* and *xete*. These morphemes seem to contain a fossilized element *.te* or *.te*, which may be cognate with *-ta/-tya*.

In addition to the general applicative *-ta/-tya*, there is a semantically more specific applicative morpheme *-ti* that conveys a malefactive/detrimental sense. It can increase the valency of intransitive verbs but not of transitive verbs. Note these examples:

- (26) a. *kui-da-ki*  
drink-1SG-DECL  
'I drink.'
- b. *kui-ti-da-ki*  
drink-MAL-1SG-DECL  
'I drink from him. / I drink his.'

---

11. Recall that the same phenomenon is found in the Jivaroan languages Aguaruna and Shiwiar, where the same forms used for the applicative are also used for the first singular object marker.

- (27) a. *kosa hiri-da-ki*  
 sun steal-1SG-DECL  
 ‘I stole a lantern.’
- b. \**luzeu-wā hiri(-tĩ)-da-ki*  
 Luzeu-AO steal-MAL-1SG-DECL  
 ‘I stole from Luzeu.’
- c. *luzeu-dĩ hĩ kosa hiri-tĩ-da=-i*  
 Luzeu-POSS-NMLZ sun steal-MAL-1SG-DECL  
 ‘I stole Luzeu’s lantern.’
- (28) *si-di-hĩ maʔĩte-wā tyari-he-tsi-hĩ=wara*  
 PRO:1SG-POSS-NMLZ cousin-AO kill-NEG-POT-NMLZ=but  
*si-di-hĩ dyi-wā tyari=asa-tĩ-ta-ki*  
 PRO:1SG-POSS-NMLZ brother-AO kill=leave-MAL-1.O-DECL  
 ‘He did not kill my cousin, but he killed my brother.’

Finally, there is a comitative applicative-like morpheme *-ete* (29, 30) that tends to be optional when occurring together with comitative case marking on the oblique object (31):

- (29) *a-ete-nā-da-ki*  
 exist-COM-FUT-1SG-DECL  
 ‘I will stay together (with her).’
- (30) *kukuihĩ-nāi a-ete-da-hĩ-ki*  
 ill-NMLZ exist-COM-1SG-NMLZ-DECL  
 ‘I was born with an illness.’
- (31) *teteru-dinĩ tomā(-ete)-da-mĩ*  
 Teteru-COM bathe-COM-1SG-VOL  
 ‘I’m going to take a bath (together) with Teteru.’

The Aikanā (ISOLATE) language also appears to have several valency changing suffixes, although their nature and productivity are not yet fully understood. They are usually placed after the subject marker, which tends to immediately follow the (sometimes derived) verb root. There is a frequently occurring transitivity suffix *-pa*:

- (32) Aikanā (ISOLATE; field notes Van der Voort)
- a. *ware-durika-ē*  
 go-DIR:inside-DECL  
 ‘He went inside.’
- b. *ware-durika-pa-ē*  
 go-DIR:inside-TRZ-DECL  
 ‘He took them inside.’

The applicative *-ita* introduces a concomitant object that has to be overtly marked itself:

- (33) a. *dürü-ʔē*  
sit-IMP  
'Sit down!'
- b. *dürü-ita-a-ʔē*  
sit-APPL-1SG.O-IMP  
'Sit close to me!'
- (34) a. *hayi titü-ke-ʔē*  
stone throw-AG-IMP  
'Throw the stone!'
- b. *hayi titü-k-ita-a-ʔē*  
stone throw-AG-APPL-1SG.O-IMP  
'Throw the stone close to me!'
- (35) a. *ite wā-ka-ē*  
here live-1SG-DECL  
'I'm living here.'
- b. (*hīzā-ete*) *wā-ita-ka-ʔe-ē*  
you-ALL live-APPL-1SG-2SG.O-DECL  
'I'm living close to you.'

There is a clear malefactive suffix *-sa* [tsa]:

- (36) a. *wāzāu kū-ka-ē*  
papaya eat-1SG-DECL  
'I ate papaya.'
- b. *wāzāu kū-ka-sa-ē*      *piu-zū*  
papaya eat-1SG-MAL-DECL Piu-POSS/BEN  
'I ate Piu's papaya.'
- (37) *dūdū pu-ʔa-pa-ʔi-ye*      *kikire-ye*  
parrot raise-IMPERS-TRZ-NMLZ-ACC parakeet-ACC  
*hāe~hāe-ke-pe-sa-ʔeye-ē*  
grab~grab-AG-MAL-3PL.O-DECL  
'Pet parrots, parakeets, he grabbed them from the residents.'

The transitivizer and the malefactive suffix can be combined:

- (38) *pau-me-pa-sa-hāyā-ē*  
run-2SG-TRZ-MAL-1PL.O-DECL  
'You abducted us.'

Aikanã also has a causative element *-za*. It occurs usually as part of a causative verbal person marking paradigm, which is often used when stative verbs are used causatively:

- (39) a. *timũ-ẽ*  
sweet-DECL  
'(The coffee) is sweet.'
- b. *kafe timũ-diza-ẽ*  
coffee sweet-1SG.CAUS-DECL  
'I sweetened the coffee.'

The Kanoé (ISOLATE) language has a transitivity morpheme *-to*, which is used for specific verb classes (Bacelar 2004: 187). Unlike the applicative in most other languages discussed here, the Kanoé transitivity marker is not suffixed directly to the verb root but attached at the end of the inflected main verb. The morpheme apparently may introduce a patient, goal, beneficiary and perhaps even other kinds of arguments, such as comitative.

- (40) Kanoé (ISOLATE)
- a. *aj vara-õ-ro õ-e-re*  
1SG speak-1-ASBJ 1-AUX-DECL  
'I am talking.'
- b. *aj vara-õ-ro-pe-to õ-e-re*  
1SG speak-1-ASBJ-2-APPL 1-AUX-DECL  
'I am talking with you.' (Bacelar 2004: 187)
- (41) a. *aj i-pateñu õ-e-re*  
1SG 1-know 1-AUX-DECL  
'I know.'
- b. *aj mi i-pateñu-pe-to õ-e-re*  
1SG 2SG 1-know-2-APPL 1-AUX-DECL  
'I know you.' (Bacelar 2004: 188)
- (42) a. *aj akiki-e-õ-ro õ-e-re*  
1SG shout-ASP-1-ASBJ 1-AUX-DECL  
'I'm screaming.' (Bacelar 2004: 185)
- b. *akiki-mi-ro-õ-to-kũ*  
shout-2-ASBJ-1-APPL-NEG  
'Don't shout at me!' (Bacelar 2004: 225)

A comitative argument, however, requires a construction that includes additional comitative marking, as in (43):

- (43) *aj ña e pe-ō-ja-to-ky ã-to-e-re*  
 1SG POSS.1SG woman lie-1-DIR:DOWN-APPL-COM 1-APPL-AUX-DECL  
 ‘I lie down with my wife.’ (Bacelar 2004: 191)

Bacelar (2004: 195) observes that the transitivizer *-to* is sometimes realized by its allomorph *-ta*, which is rare, but which apparently occurs in free variation.

Telles (2002: 241) notes that in the Northern Nambikwaran language *Latundê* the suffix *-ka*, depending on the verb it attaches to, can take on the roles of beneficiary (44b, 47), goal or source, augmenting the valency of the verb, just as an applicative. It is placed directly after the verb root:

- (44) *Latundê* (NAMBIKWARAN, NORTHERN)  
 a. *wã-wi'taj 'nãn-Ø-'tãn*  
 2-husband cry-3.S-IPFV  
 ‘Your husband is crying.’  
 b. *wã-wi'taj 'nãn-ka-nah-Ø-'tãn*  
 2-husband cry-APPL-2.O-3.S-IPFV  
 ‘Your husband is crying for you.’ (Telles 2002: 321)

There is also a causative suffix *-tu* in *Latundê*. Its description and the few examples and are not conclusive, but seem to indicate an indirect kind of causation:

- (45) a. *i'te ka'lã?-Ø-'ten-ni-'tã*  
 hold stand-3.O-DES-FUT-1SG.IPFV  
 ‘I am going to hang/hold it up.’  
 b. *i'te ka'lã?-tu-Ø-'ten-ni-'tã*  
 hold stand-CAUS-3.O-DES-FUT-1SG.IPFV  
 ‘I will hang it, hold it up.’ (Lit.: ‘I will cause its suspension.’)  
 (Telles 2002: 242)

There is furthermore a “causative/resultative” (Telles 2002: 97) prefix *ã(l)-*, which is described as “agentive” elsewhere (Telles 2002: 238):

- (46) *ãl-mu'mã?-ñã*  
 CAUS-be.tame-STAT  
 ‘It is/was tamed.’ (an animal) (Telles 2002: 97)  
 (47) *ã-'lo?-ka-wi-'tãn*  
 CAUS-sink-BEN-1DU-IPFV  
 ‘We two sank it.’ (Telles 2002: 121)

According to Eberhard (2009: 408), the Northern Nambikwaran language *Mamaidê* has what the author calls an “oblique suffix” *-ka*. This suffix introduces an object that can have various semantic functions, such as benefactive, source,

goal and referent. At times it references any other nominal except the principal arguments of the verb:

- (48) Mamaindê (NAMBIKWARAN, NORTHERN)  
*ta-haiʔka ta-sanin-sq-tu hai*  
 POSS.1-word POSS1-happy-CLF.liquid-FNS he  
*set-ka-Ø-le-a-nan-wa*  
 speak-APPL-3.O-IPST-1.S-PST-DECL  
 ‘I spoke my words, my happy words, to him.’ (in intermediate past time)  
 (Eberhard 2009: 390)
- (49) *set-ka-na-lat<sup>h</sup>a-wa*  
 speak-APPL-2.O-3.S-PRS-DECL  
 ‘He/she is speaking about you.’
- (50) *saninʔ-ka-Ø-Ø-nan-wa*  
 happy-APPL-3.O-3.S-PST-DECL  
 ‘He/she is happy about it.’ (Eberhard 2009: 409)

In Mamaindê, the causative prefix *ta(l)-* is used to signal that the subject of the verb is a third party who voluntarily caused the action to occur (Eberhard 2009: 395):

- (51) *ta-wate-Ø-lat<sup>h</sup>a-Ø-wa*  
 CAUS-disappear-3.O-3.S-PRS-DECL  
 ‘He caused (it) to disappear/be gone.’
- (52) *tal-enkūn-ʔna-Ø-t<sup>h</sup>unna-wa*  
 CAUS-heal-2.O-3.S-FUT2-DECL  
 ‘He will cause you to be healed.’ (Eberhard 2009: 396)

This prefix is probably cognate with Latundê *ã(l)-*.

The Tupian language Karo also possesses two causative prefixes, *ma-* ‘causative’ (54) and *ta-* ‘sociative causative’ (55) (Gabas Jr. 1999: 80). The *ta-* causative adds a comitative object:

- (53) Karo (TUPIAN, RAMARAMA)  
*et owā e=ma-kə-t*  
 2SG mother 2SG=CAUS-walk-IND1  
 ‘Your mother made you walk.’ (Gabas Jr. 1999: 81)
- (54) *wat owā o=ta-kə-t*  
 my mother 1SG=CAUS.SOC-walk-IND1  
 ‘My mother made me walk, walking with me.’ (Gabas Jr. 1999: 83)

Apart from the causative *ma-* and the sociative causative *ta-*, Gabas Jr. (1999) makes no mention of other valency-augmenting morphemes in the language.



The Tupian language Tupari seems to have a common causative prefix *ō(m)-* and a comitative causative prefix, *ete-* or *ite-* (cf. Kwaza), which also seems to display some applicative traits:

- (55) Tupari (TUPIAN; TUPARI)  
*tet-* ‘to go’ *ete-tet-* ‘to let someone go and accompany him/her’  
*wat-* ‘to go’ *ite-wat-* ‘to take with one’  
*wak-* ‘to cry’ *ite-wak-* ‘to cry about someone’  
 (Rodrigues & Caspar 1958: 72)

The comitative causative is usually attached to verbs of motion, although there are some exceptions, as Singerman (2018: 125–130) points out, such as *wak-* in (55), which can be interpreted as ‘to mourn, to miss’.

As Table 2 in Section 3 shows, the Tupian languages do not seem to have a straightforward applicative *-tA* suffix, although some similarities with and between several Tupian languages can be observed in valency marking; as pointed out in Section 2.1, Kokama-Kokamilla has a highly productive causative suffix *-ta* at its disposal.

The Takanan language Cavineña does not make use of an applicative mechanism (Guillaume 2008: 255), but Guillaume does report a suffix *-ta(na)*, which converts transitive verbs into intransitive verbs:

- (56) Cavineña (PANO-TAKANAN, TAKANAN)  
 a. *Roberto=ra kashi ara-ya.*  
 Roberto=ERG sweet.banana eat-IPFV  
 ‘Roberto eats sweet banana.’  
 b. *Kashi ara-tana-ya.*  
 sweet.banana eat-PASS-IPFV  
 ‘Sweet bananas are eaten.’ (Guillaume 2008: 256)

The suffix *-ta(na)* applies almost exclusively to (di)transitive verbs and Guillaume considers it a passive marker “with the original O argument becoming the S argument of the derived verb and the original A argument being obligatorily omitted” (2008: 256).

In a later publication, a distinction is made between two verbal suffixes, *-ta* and *-tana*, which both have core passive meanings but are distinct morphemes. The main differences lie in that *-tana* is fully productive and can also be interpreted semantically as an anticausative marker (57), while the semantics of *-ta* are purely passive (58) (Guillaume 2012: 116–117). In case the semantics of *-tana* is anticausative, the notional O is expressed as S:

- (57) [Ekwe karusune iyakwake utsa-wa=ju] dyuru-tana-chine.  
 1SG.GEN pants new.ABS wash-PRF=DS shorten-PASS1-RPST  
 ‘(The women) washed my new pants and they shrunk.’ (Guillaume 2012: 119)
- (58) A-ta-nuka=dya ekwana.  
 affect-PASS2-ITR=FOC 1PL.ABS  
 ‘We will be attacked (lit. affected).’ (Guillaume 2012: 121)

Emkow (2006: 558–559) reports a marker *-tana* in the Takanan language Araona, which is used for constructions in the ‘middle domain’, such as bodily motion.

- (59) Etna hewichanaibo.  
 ema hewi-tana-ibo  
 1SG.ABS wipe.away-MID-PFV.NVIS  
 ‘I slipped.’ (Emkow 2006: 559)

According to Guillaume (2011: 530–531), cognates of *-ta* are present in all the remaining Takanan languages, where the morpheme is not a passive marker but rather a cross-reference marker for a third person plural S in intransitive clauses and a third person number-neutral A in transitive clauses. These findings led him to propose that a third person plural subject marker in both intransitive and transitive clauses is the source of *-ta* in (pre-)proto-Takanan. Guillaume (2012: 126) concludes that the passive marker *-ta* in Cavineña derives from the same source “through an intermediate stage where it would have referred to an indefinite A argument”.

The Arawakan language Baure has an applicative suffix *-čo* [tʃo], which has various functions. In the first place, it attaches obligatorily to Spanish loan verbs, as in (60):

- (60) Baure (ARAWAKAN)  
*kanach-* < Sp. *ganar* ‘earn, win’ (Danielsen 2012: 323)  
*-awantach-* < Sp. *aguantar* ‘bear’  
*-kasach-* < Sp. *casar* ‘marry’

The applicative is also a productive verbalizer. It can verbalize nouns and some adjectives:

- (61) *havi* ‘paddle’ *-havič-* ‘paddle’ (Danielsen 2007: 243)  
*haki* ‘door’ *-hakič-* ‘close’  
*wer* ‘medicine’ *-weroč-* ‘cure (apply medicine)’

The main function of the applicative, however, is turning intransitive verbs into transitive verbs, as in (62), where *-čo* is rendered as *-či* due to vowel harmony:

- (62) a. *ni=koes*  
 1SG=wake.up  
 ‘I woke up.’  
 b. *piti’ pi=koes-či=ni*  
 2SG 2SG=wake.up-APPL=1SG  
 ‘You woke me up.’ (Danielsen 2007: 242–243)

In later work, Danielsen (2014: 294) analyzes *-čo* as a transitivizer.

The Baure causative prefix *i(mo)-* (63), which is also found in many other Arawakan languages, is similar to the causative in unrelated Tupi-Guarani languages, such as Emerillon *mo-* (64), Guajajara *mu-*, Paraguayan Guarani *mbo-/mo-*, and certain other Tupi languages, such as Mundurukú *mu-/muy* (Gomes 2006: 78). The presential applicative in Ashéninka Alto Perené (ARAWAKAN) is form-wise similar as well, be it that it is a suffix instead of a prefix (65); the same applies to the comitative applicative suffix *-imo* in Ashéninka Ucayali-Pajonal (ARAWAKAN) (cf. 10).

- (63) *pimoekomorikonir*  
*pi=imo-ikomoriko=ni=ro*  
 2SG=CAUS-kill=1SG=3SG.M  
 ‘You make me kill him.’ (Danielsen 2012: 324)
- (64) Emerillon (TUPIAN, TUPI-GUARANI)  
*ere-mo-zaug-a-ōwā-zepe-ʔe-po* *mamā, dʒasor?*  
 2SG.I-CAUS-bathe-REF-little-CONC-INTNS-INT Mum Djasot  
 ‘But did you really wash Mum properly, Djasot?’ (Rose 2008: 442)
- (65) Ashéninka, Alto Perené (ARAWAKAN, CAMPA)  
*Te intyerovimotapakyari.*  
*te i=N-tyerov-imo-t-ap-ak-e=ri*  
 NEG.REAL 3M.A=IRR-kneel.down-APPL-EP-DIR-PRF-IRR=3M.O  
 ‘He didn’t kneel in his presence.’ (Mihas 2010: 129)

The Arawakan language Paresi has a causative suffix *-ki*, which occurs in combination with the causative prefix *a-* and can also have a sociative causative meaning. In addition, there is a transitivizing suffix *-tya*, which is combined with the causative prefix *a-* in the same manner as *-ki*, but which is associated with direct causation (Brandão 2014: 28). Furthermore, *-tya* is also used in combination with the causative prefix *a-* as a verbalizer of nouns and adjectives.

Also, the southeastern Peruvian language Harakmbut (HARAKMBUT-KATUKINAN) has verbal valency-changing morphemes that correspond to the *tV* pattern. Van Linden (fc.) identifies a comitative applicative prefix *ta-* and a sociative causative prefix *to-*. Example (66) shows both prefixes:

- (66) Harakmbut (HARAKMBUT-KATUKINAN)  
*ehua'* 'go' (Tripp 1995: 204)  
*o'tohua'* 'He takes her along.'  
*eta'hua'* 'follow an animal's tracks, follow, persecute'

Finally, several Macro-Jêan languages also have a *-tV*-like valency-changing element.<sup>12</sup> The semantic and morphosyntactic analyses of this element differ considerably between sources. For example, Bororo (MACRO-JÊAN) has a causative element *da*, which is analyzed as a causative auxiliary verb by Nonato (2008). Rikbaktsa (MACRO-JÊAN) has different prefixed and suffixed elements *ti*. Boswood (1974) analyzes the prefix as causative and the suffix as imperative, whereas Silva (2011) treats both the prefix and the suffix as imperative. In some cases, however, Boswood's prefix is analyzed by Silva as an accusative suffix on the preceding noun. In Canela (JÊAN), the postposition *to* follows the object that results from valency change of intransitive verbs (Alves 2014: 192–193). In the case of causative valency change (only of non-agentive verbs), a new subject is introduced and the demoted object is followed by *to*. In case of applicative valency change (only of agentive verbs), an object is introduced and followed by *to*. Martins et al. (2009) have noted the similarities in causative functions and forms between Apinajê, Xikrin, Krahô-Canela, Bororo and Rikbaktsa and assume these are cognate.

### 2.3 Bolivian Andean foothills

Yurakaré, Leko, and Mosestén are the three Bolivian languages that have traditionally been considered foothill languages, even though the Yurakaré have been moving out of the area in the past decades under the pressure of the production of coca by Andean settlers and illegal drug trafficking.

Yurakaré (ISOLATE) has various applicative prefixes, which, with the exception of the comitative applicative, are all combined with special person paradigms (Van Gijn 2006: 149); for example, the benefactive applicative *n-*, malefactive applicative *la-* and the goal applicative *y-* (Van Gijn 2015: 117).<sup>13</sup> Their forms do not suggest a relationship with *-tA*. However, there is a causative element *-che* [tʃe] as well as a middle voice marker *-ta* (Van Gijn 2006: 170ff.). The latter also expresses reflexive, reciprocal and agentless passive. Furthermore, it is often lexicalized:

12. As appears from Van der Voort's fieldnotes, Djeoromitxí (MACRO-JÊAN) may have a transitive suffix *-tu*, although its productivity has not been established.

13. Leko (ISOLATE) also has a benefactive marker *in-* and a malefactive *ih-* that combine with special person paradigms (Van de Kerke 2009: 304), as in Yurakaré. As pointed out by Van Gijn (2015: 117), the benefactive marker, whose vowel fuses with the vowel of the object person prefix, is formally very similar to the Yurakaré benefactive marker *n-*.

- (67) Yurakaré (ISOLATE)  
*bata* ‘leave’                      *bache* ‘send’                      (Van Gijn 2006: 172–174)  
*wëta* ‘stoop’                      *wëche* ‘bend’  
*bëjta* ‘see’                      *bëjche* ‘show’  
*duta* ‘burn’ (INTR)    *duche* ‘burn’ (TR)

Finally, the element *-ta* is also attached to nouns, expressing associative and ‘former’:

- (68) *ti-dala* ‘my head’      *ti-dala-ta* ‘my pillow’                      (Van Gijn 2006: 175)  
*bashti* ‘wife’                      *ti-bashti-ta* ‘my ex-wife’

Mosetén forms together with Tsimané’ (a.k.a. Chimane) the small linguistic family Mosetenan. The language has a total of five applicatives – three applicative suffixes and two prefixes – which have rather specific functions (Sakel 2004: 318–319). We will discuss the suffix *-tya/-tye/-tyi* and the prefix *ti-/tí-*. The suffix *-tya/-tye/-tyi*, which is possibly related to the verb *tyi* ‘give’, often denotes that something is done to another person or thing, including something being taken from another person, and in that sense is reminiscent of a malefactive marker:

- (69) Mosetén (MOSETENAN)  
*Yae je-k-tye-te*                      *atya*      *jiri-ty* *camisa*.  
 1SG take-AM-MAL-3M.O uncle(M) one-M shirt.M  
 ‘I take a shirt from my uncle.’                      (Sakel 2004: 321)
- (70) [...] *keo'tso-tye-tye-'*      *o'sho'*  
 step.on-VSM-MAL-3F.O clothes  
 ‘[...], (but) they stepped on her clothes.’                      (Sakel 2004: 321–322)

The applicative prefix *ti-/tí-* appears predominantly with intransitive verbs that express an emotional state, while adding a condition or reason that provoked the emotional state:

- (71) *Yäe tí-fäk-e-'*                      *mö'*      *nanasi'*  
 1SG APPL-angry-VSM-3F.O 3F.SG girl  
 ‘I was angry because of the girl.’                      (Sakel 2004: 322)

## 2.4 Andean

In this subsection we will review, be it superficially, a few valency markers found in Andean languages that have caught our eye.

Aymara has several valency related suffixes. Coler (2014: 364) points out that the momentaneous suffix *-t'a* in Muylaq' Aymara can also have an applicative meaning, taking intransitive-like verbs and applying them to objects. According to Coler, the suffixation of *-t'a* on the word *sar-ch-i-x* ‘he must have gone’ in (72), basically makes the verb more transitive-like, resulting in a change of meaning from ‘go’ to ‘visit’.

- (72) Muylaq' Aymara, (AYMARAN)  
*Ukat jichax ma wisinax sart'chix.*  
 uka-t(a) jicha-x(a) ma wisina-x(a) sar(a)-t'(a)-ch(i)-i-x(a)  
 that-ABL now-TOP one neighbor-TOP go-MOM-CNJ-3-TOP  
 'And after a neighbor must've visited.' (Coler 2014: 364)

In (73), the suffixation of *-t'a* to *jiwa* 'die' signals in a way that the woman died 'on the man':

- (73) *Ukat jichax warmipax jiw'tajwatayn.*  
 uka-t(a) jicha-x(a) warmi-pa-x(a) jiw(a)-t'a-jwa-tayn(a)  
 that-ABL now-TOP wife-3.POSS-TOP die-MOM-BFR-3.DPST  
 'Then now his wife had died.' (Coler 2014: 364)

In addition, there is a derivational verbalizing factive suffix *-ča* [tʃa] in Aymara, which indicates that that what is referred to by the root is made or realized, as in *uta-ča-* (house-FACT) 'build' (Cerrón-Palomino & Carvajal Carvajal 2009: 200). Finally, Aymara has a causative suffix *-ya*, as in *iki-ya-* (sleep-CAUS) 'cause to sleep', *jacha-ya-* (cry-CAUS) 'cause to cry', *yati-ya-* (know-CAUS) 'inform', etc. (Cerrón-Palomino & Carvajal Carvajal 2009: 200). Again, the similarity with regard to form, function and position with the other languages discussed here is striking.

Central Peruvian Quechua has a factive morpheme *-cha*, like Aymara, which verbalizes nouns, adjectives, and adverbs:

- (74) *Jaqay punta-man ri-yku, chay-pi ch'uñu-cha-mu-yku.*  
 DEM:ADV:DIST point-AL go-1PE DEM:MED-LOC chuño-FACT-CIS-1PE  
 'Over there, we're going to that point, there we'll make the *chuño*.'  
 (Plaza Martínez 2009: 256)

Other valency related morphemes of Quechua that should be mentioned here are the applicative transitivizer *-pa(:)* (75a), the causative suffix *-či* [tʃi] (75b), and the medio-passive/middle marker *-ka(:)* (75c) (Adelaar with Muysken 2004: 229):

- (75) a. *ayku-* 'to laugh' *ayku-pa(:)-* 'to laugh at', 'to smile at'  
*muna-* 'to want' *muna-pa(:)-* 'to desire', 'to long for'  
 b. *wanu-* 'to die' *wanu-či-* 'to kill', 'to cause to die'  
*maqa-* 'to beat' *maqa-či-* 'to make (sb) beat', 'to have (sb) beaten'  
 c. *punu-* 'to sleep' *punu-ka(:)-* 'to fall asleep'  
*laqa-* 'to stick' *laqa-ka(:)-* 'to get stuck to sb'

Chipaya has two verbalizing suffixes, "transformative" *-khi*, which derives from the verb *khi* 'become' (76), and the verbalizer *-ta* (77), which according to Cerrón-Palomino (2009: 64) was a copular verb originally:

- (76) Chipaya (URU-CHIPAYAN)  
*oñi* 'deaf' *oñi-khi-* 'become deaf' (Cerrón-Palomino 2009: 64)  
*chiwi* 'white' *chiwi-khi-* 'become white'

- (77) a. *tii aznu-ki Huwan-zh-ta*  
 this donkey-TOP Juan-GEN-VBLZ  
 ‘This jack (male donkey) is Juan’s.’  
 b. *taa azn-i-ki Huwan-a-ta*  
 this donkey-F-TOP Juana-GEN-VBLZ  
 ‘This jenny (female donkey) is Juana’s.’ (Cerrón-Palomino 2009: 54)

Furthermore, there is an instrumental/comitative suffix *-tan(a)*:

- (78) *Luwis-ki par-zh-tan paku jwat-chi-tra.*  
 Luís-TOP stick-GEN-INS dog hit-PST-DCL  
 ‘Luis hit the dog with a stick.’  
 (79) *Huwanitu-ki Luwisit-zh-tan thon-chi-tra.*  
 Juanito-TOP Luisito-GEN-COM come-PST-DECL  
 ‘Juanito came with Luisito.’ (Cerrón-Palomino 2009: 54–55)

Mapudungun (a.k.a. Mapuche), the only extant Araucanian language today, has four applicatives, of which we will focus on the suffix *-tu*. While *-tu* increases the valency of intransitive verbs or promotes another object of transitive verbs (Smeets 2008: 297), it can also decrease valency as shown by Zúñiga (2009). Moreover, it may function as a verbalizer as well (Fernández Garay 2001: 51; Zúñiga 2013) (80a). Zúñiga (2013) refers to a “heterogeneous” marker *-tu*, since it appears to have a whole array of valency-increasing (80b), valency-neutral (80c), and valency-decreasing properties (80d):

- (80) Mapudungun (ARAUCANIAN)
- |    |                |                            |                   |                            |
|----|----------------|----------------------------|-------------------|----------------------------|
| a. | <i>wekufü</i>  | ‘evil spirit’              | <i>wekufü-tu-</i> | ‘bewitch, cast a spell on’ |
|    | <i>sapatu</i>  | ‘shoe(s)’                  | <i>sapatu-tu-</i> | ‘put on/wear shoes’        |
| b. | <i>aye-</i>    | ‘laugh’                    | <i>aye-tu-</i>    | ‘laugh at’                 |
|    | <i>nümu-</i>   | ‘smell’ (INTR)             | <i>nümu-tu</i>    | ‘smell’ (TR)               |
| c. | <i>kansha-</i> | ‘get tired’                | <i>kansha-tu-</i> | ‘rest’                     |
|    | <i>amu-</i>    | ‘go’                       | <i>amu-tu-</i>    | ‘go away, leave’           |
| d. | <i>imül-</i>   | ‘roll’ (TR)                | <i>imül-tu-</i>   | ‘roll for fun’ (INTR)      |
|    | <i>kayfü-</i>  | ‘whisper in sb’s ear’ (TR) | <i>kayfü-tu-</i>  | ‘whisper’ (INTR)           |
- (Zúñiga 2013)

Puquina, an extinct language once spoken in the Southern Andes around Lake Titicaca and in the Peruvian coastal departments of Arequipa, Moquegua and Tacna, had a suffix *-que/-gue* [ke/ge], which seems to have had an applicative function or indicated accompaniment (Adelaar & Van de Kerke 2009), as in (81):

(81) *mutu-que-guina*

suffer-APPL-1.FUT(>2.O)

'I will suffer with you.' / 'I will help you to suffer.'

(Adelaar & Van de Kerke 2009: 135)

Adelaar and Van de Kerke (2009: 135) also mention that the text they analyzed contains instances of Quechua derivational suffixes, which preferably attach to roots of Quechua origin, like, for example, the factive suffix *-cha* in *pampa-cha* 'forgive' (lit. 'to level').

### 3. Discussion

Section 2 has shown that a certain morpheme varying in shape, but similar to *-tA* is apparently widespread in both the (North)western Amazon and the Southern Amazon. It tends to be a suffix that is attached immediately to the verb root. Its functions are related to valency, but it is not always a valency-increasing device, since at times it does not change the valency of the verb it attaches to or even decreases its valency. We found *-tA* and its allomorphs, among others, as an applicative, causative, verbalizer, factive, and middle voice marker.

The fact that *-tA* occurs as an applicative in some languages but as a causative in others, is not surprising in view of the conceptual link between applicatives and causatives. Both derivational devices usually allow for a new (nominal) argument to be introduced, thus deriving a new (di)transitive verb from a more basic intransitive or transitive verb (be it that causatives promote a new subject rather than a new object). Its function as a middle marker, however, is somehow less expected in view of the valency-decreasing implications. Nevertheless, we have seen that as an applicative *-tA* may decrease the valency of a verb, as in Shawi (see Example 5) and Mapudungun (80d). As yet we have not found any clear examples of languages displaying applicative/causative syncretism through *-tA*, but the spread of the morpheme in different valency-related functions does suggest different stages of grammaticalization. Peterson (2007: 64) discusses two basic types of applicative/causative isomorphism: (i) benefactive or malefactive applicative/causative and (ii) comitative or instrumental applicative/causative, the latter being a rarer phenomenon (Peterson 2007: 65). Further research is needed to see whether this holds for our data as well and whether these confirm the proposed grammaticalization path causative > sociative causative > comitative applicative, as proposed by Shibatani & Pardeshi (2002) and Peterson (2007: 137).



Table 2 gives an overview of the relevant markers discussed in the previous sections. Some of the languages have, in addition to *-tA*, other applicative or causative morphemes, such as the *-mo* and *-ma*-like forms, which are in principle unrelated to *-tA*, but which may have interfered in different ways with the valency system of the particular language. Such morphemes are therefore also included in the table.

**Table 2.** Overview of applicatives and causatives in Amazonia and beyond

Language	Family	Applicative(s)		Causative(s)		Other
		Prefix	Suffix/ postp.	Prefix	Suffix/ postp.	
<i>(North)western Amazon</i>						
Sikuani	GUAHIBOAN		-to			
Resígaro	ARAWAKAN				-(o)ta, -(o)tu	
Ashéninka	ARAWAKAN, CAMPA		-ako, -imo		-(t)ag	
Aguaruna	JIVAROAN		-hu, -tu			-tu, -hu '1SG.OBJ'
Shiwiar			-ru, -tu			
Shawi	KAWAPANAN		-ta, -ti			
Shiwilu			-ta, -tu, -pa'			
Munichi*	ISOLATE				-ta, -tʃa -ʃta, -ʃti	
Kokama- Kokamilla	TUPIAN, TUPI-GUARANI				-ta	
Yagua	PEBA-YAGUAN		-ta, -tya			
Witoto	WITOTOAN				-ta	
Andoa	ZAPAROAN		-ta			
Arabela			-ta, -tia			
<i>Southern Amazon</i>						
Baure	ARAWAKAN		-čho	i(mo)-		
Paresi				a-...	-ki	a-...-tya 'TRZ', 'VBLZ'
Harakmbut	HARAKMBUT- KATUKINAN	ta-		to-		
Aikanã	ISOLATE		-ita, -pa		-za	-sa 'MAL'
Kanoé	ISOLATE		-ta, -to			
Kwaza	ISOLATE		-ta, -tya			-ta '1SG.OBJ'
Latundê	NAMBIKWARAN,		-ka	ã(l)-	-tu	
Mamaindê	NORTHERN		-ka	ta(l)-		

Table 2. (continued)

Language	Family	Applicative(s)		Causative(s)		Other
		Prefix	Suffix/ postp.	Prefix	Suffix/ postp.	
Araona	PANO-TAKANAN					-tana 'MID'
Cavineña						-ta 'PASS' -tana 'PASS', 'ACAUS'
Karo	TUPIAN, RAMARAMA			ma-, ta-		
Tuparí	TUPIAN, TUPARI			ō(m)-, ete-, ite-		
Mundurukú	TUPIAN, MUNDURUKÚ			mu-/ muy-	-da/-ta, -at	
Canela	MACRO-JÊAN, JÊ		to		to	
Rikbaktsa	MACRO-JÊAN			ti-		
Bororo					-to/-do, dΛ	
<i>Bolivian foothills</i>						
Yurakaré	ISOLATE	n-, la-, y-			-che	-ta 'MID'
Mosetén	MOSETENAN	ti-, tĩ-	-tya, -tye, -tyi			
<i>Andes</i>						
Mapudungun	ARAUCANIAN		-tu		-(ü)m	
Aymara	AYMARAN		-t'a		-ya	-cha 'FACT'
Quechua	QUECHUAN		-pa(:)		-chi	-cha 'FACT' -ka(:) 'MID'
Chipaya	URU-CHIPAYAN					-ta 'VBLZ' -tan(a) 'COM/INS'
Puquina	UNCLASSIFIED		-que/-gue			-cha 'FACT'

\* Gibson (1996).

Table 2 contains several patterns of similarities across the languages listed. There is a genetic explanation for the *-mo* and *-ma*-like forms, which characterize many Arawakan and Tupian languages. Furthermore, a relationship between *-ta* and *-ka*-like forms is phonetically not surprising. However, parallel to the *Wanderwörter* mentioned in Section 1, there is no easy explanation for the widespread similarity of the *-tA*-like forms. This spread is undoubtedly due to language contact, which has been solidly attested for many of the immediately neighboring languages listed, such as Aikanã and Kwaza, but how and when it came to be so widely distributed over the entire western Amazon and beyond is a mystery. Also, the allomorphy *-ta/-tya* shared by several unrelated and geographically widely separate languages

is conspicuous. And although the shared homophony of the applicative morpheme and the first person object morpheme between the two unrelated non-neighboring languages Aguaruna (*-tu*, *-hu*) and Kwaza (*-ta*) is, as one reviewer correctly pointed out, not necessarily a coincidence, it does raise the question in which way these homophonies relate to the diffusion of *-tA*.

#### 4. By way of conclusion

Admittedly, the form *-tA*, which we have been looking at in this article, is very short, has a very common CV syllable structure and consists of a highly universal consonant and vowel. As emphasized by comparativists (e.g. Campbell 1998: 322) there is a greater possibility of chance resemblances between relatively short forms than between more complex forms. Nonetheless, the similarities in morphological form, position and function of the bound applicative morpheme *-tA* across the Amazon region and beyond are striking. In view of its multifunctionality in certain languages, however, it remains to be seen whether ‘applicative’, which we have used as a cover term in the present article, is the correct label for this morpheme.

There seem to be two ‘epicenters’ of the applicative *-tA* phenomenon and its reflexes in the causative realm. The northwestern area has been identified before, by Payne (1985), and was extended by Wise (2002) to include parts of Peru. In the present article we have attempted to further complete the picture. In view of the genetic diversity of the languages involved, and considering also other, independent shared traits in classifier systems, we regard the ubiquity of the *-tA* applicative across the (North)western and Southern Amazon region as an areal trait. Furthermore, even though the Mapudungun applicative *-tu* may or may not be related historically to Amazonian *-tA*, as already pointed out by Zúñiga (2013), it is intriguing how similar their functions are, as well as their positions relative to the verb root.

Perhaps it is precisely because the element *-tA* is so simple and useful that it could easily spread relatively unchanged, which occurred possibly under conditions of bilingualism or second language acquisition of neighboring languages. Matras (2009, 2014) has demonstrated the likely importance of bilingualism with fascinating instances of grammatical borrowing, in particular of derivational categories, between individual languages. Winford (2003) argues that the metaphor of borrowing does not hold for structural diffusion, but that it must be regarded as a substrate effect, possibly due to language shift. What these approaches imply for very widespread morphological traits like the one described here is an important topic for further debate.

## Acknowledgements

We thank the editors of the present volume for having taken the initiative to set up this long due and much deserved volume in honor of Pieter Muysken. We have always been greatly inspired by Pieter, first as students, then as postdocs, and lately as colleagues. We hope that even though he may not agree with our interpretation of the monosyllabic data we present in this article, he will find the facts intriguing. We are furthermore indebted to the highly appreciated comments and corrections by two anonymous reviewers. Any remaining flaws are our responsibility only.

## Abbreviations

A	subject of a transitive clause	DUR	durative
ABL	ablative	EP	epenthetic
ACC	accusative	F	feminine
ADV	adverbial	FACT	factitive verbalizer
AG	agentive	FNS	final noun suffix
AL	allative	FOC	focus
AM	associated motion marker 'do an action after arrival, away'	FRS	fossilized reality status
AO	animate object	FUT	future
APPL	applicative	GEN	genitive
APPLG	generalized applicative	IFUT	immediate future
ASBJ	agentive subject	IMP	imperative
AUX	auxiliary	IMPERS	impersonal
BEN	benefactive	IND	indicative
BFR	buffer	INF	infinitive
CAUS	causative	INS	instrumental
CLF	classifier	INT	interrogative
CNJ	conjunctural evidential	INTNS	intensive Aktionsart
COM	comitative	INTR	intransitive
CONC	concessive	IPFV	imperfective
COP	copula	IPST	intermediate past
DECL	declarative	IRR	irrealis
DEM	demonstrative	LOAF	low affectedness Aktionsart
DES	desiderative	LK	linker
DIM	diminutive	M	masculine
DIR	directional	MAL	malefactive
DIST	distal	mf	microfiche
DK	associated motion 'from'	MED	medial
DPST	distant past	MID	middle
DS	different subject	MOM	momentaneous
DU	dual	NEG	negative
		NFUT	non-future

NVIS	-non-visual	SOC	sociative
O	object	sp.	species
OBL	oblique	Sp.	Spanish
PASS	passive	ss	same subject
PFV	perfective	STAT	stative
POSS	possessive	TERM	terminative
PP	positive polarity	TOP	topic
PRF	perfect	TR	transitive
PRS	present	TRZ	transitivizer
PST	past	VBLZ	verbalizer
REAL	realis	VOL	volitional
REF	referential	VSM	verbal stem marker
REL	relativizer	1	first person
RPST	recent past	2	second person
S	subject; subject of an intransitive clause	3	third person
SG	singular	I	set I person index
SIM	simultaneous		

## References

- Adelaar, W. F. H. with Muysken, P. C. 2004. *The Languages of the Andes*. Cambridge: CUP.
- Adelaar, W. F. H. & van de Kerke, S. 2009. Puquina. In *Lenguas de Bolivia I: Ámbito andino*, M. Crevels & P. C. Muysken (eds), 125–146. La Paz: Plural editores.
- Aikhenvald, A. Y. 2001. Dicionário Tariana-Português e Português-Tariana. *Boletim do Museu Paraense Emílio Goeldi (Antropologia)* 17(1): 3–433.
- Aikhenvald, A. Y. 2002. *Language Contact in Amazonia*. Oxford: OUP.
- Alves, F. de Castro. 2014. Aumento de valência em Canela. In *Incremento de valência en las lenguas amazónicas*, F. Queixalós, S. Telles & A. C. Bruno (eds), 191–209. Bogota: Instituto Caro y Cuervo.
- 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. 2002. *Description phonologique, grammaticale et lexicale du Moré, Langue Amazonienne de Bolivie et du Brésil*, 2 vols. Porto Velho: Editora da Universidade Federal de Rondônia.
- Bacelar, L. N. 2004. Gramática da língua Kanoê. PhD dissertation, Katholieke Universiteit Nijmegen. <<https://repository.uibn.ru.nl/handle/2066/19429>> (26 April 2020).
- Baker, M. C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago IL: The University of Chicago Press.
- Birchall, J. T. R. 2014a. *Argument Marking Patterns in South American Languages* [LOT Dissertation Series 352]. Utrecht: LOT Publications. <<https://www.lotpublications.nl/argument-marking-patterns-in-south-american-languages-argument-marking-patterns-in-south-american-languages>> (26 April 2020).
- Birchall, J. T. R. 2014b. Verbal argument marking patterns in South American languages. In *The Native Languages of South America: Origins, Development, Typology*, L. M. O’Connor & P. C. Muysken (eds), 223–249. Cambridge: CUP.
- Boswood, J. 1974. Algumas funções de participante nas orações Rikbaktsa. *Série Lingüística* 3: 7–33. SIL Brazil.

- Brandão, A. P. 2014. Causatives in Paresi-Haliti (Arawak). In *Incremento de valencia en las lenguas amazónicas*, F. Queixalós, S. Telles & A. C. Bruno (eds), 19–31. Bogotá: Instituto Caro y Cuervo.
- Bybee, J. L. 1985. *Morphology: A Study of the Relation between Meaning and Form* [Typological Studies in Language 9]. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.9>
- Cacique Coronado, L. 2016. *Guía teórica asheninka*, G. Zerdin (ed.). Atalaya: Universidad Católica Sedes Sapientiae.
- Campbell, L. 1998. *Historical Linguistics: An Introduction*. Edinburgh: EUP.
- Campbell, L. 2006. Areal linguistics: A closer scrutiny. In *Linguistic Areas: Convergence in Historical and Typological Perspective*, Y. Matras, A. McMahon & N. Vincent (eds), 1–31. Houndmills: Palgrave MacMillan.
- Campbell, L. & Grondona, V. 2012. Languages of the Chaco and Southern Cone. In *The Indigenous Languages of South America: A Comprehensive Guide* [The World of Linguistics 2], L. Campbell & V. Grondona (eds), 625–667. Berlin: De Gruyter Mouton.
- Cerrón-Palomino, R. 2009. Chipaya. In *Lenguas de Bolivia, I: Ámbito andino*, M. Crevels & P. C. Muysken (eds), 29–77. La Paz: Plural editores.
- Cerrón-Palomino, R. & Carvajal Carvajal, J. 2009. Aimara. In *Lenguas de Bolivia I: Ámbito andino*, M. Crevels & P. C. Muysken (eds), 169–213. La Paz: Plural editores.
- Coler, M. 2014. *A Grammar of Muylaq' Aymara: Aymara as spoken in Southern Peru* [Brill's Studies in the Indigenous Languages of the Americas 8]. Leiden: Brill.
- Comrie, B. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell.
- Comrie, B., Golluscio, L. A., González, H. & Vidal, A. 2010. El Chaco como área lingüística. In *Estudios de lenguas amerindias, 2: Contribuciones al estudio de las lenguas originarias de América*, Z. Estrada Fernández & R. Arzápalo Marín (eds), 85–130. Hermosillo, Sonora: Universidad de Sonora.
- Crevels, M. 2006. Verbal number in Itonama. In *What's in a Verb? Studies in the Verbal Morphology of the Languages of the Americas* [LOT Occasional Series 5], G. J. Rowicka & E. B. Carlin (eds), 159–170. Utrecht: LOT Publications. <<https://dspace.library.uu.nl/handle/1874/296552>> (26 April 2020).
- Crevels, M. 2011. Who did what to whom in Magdalena (Itonama). *International Journal of American Linguistics* 77(4): 577–594.
- Crevels, M. 2012. Itonama. In *Lenguas de Bolivia, II: Amazonía*, M. Crevels & P. C. Muysken (eds), 233–294. La Paz: Plural editores.
- Crevels, M. & van der Voort, H. 2008. The Guaporé-Mamoré region as a linguistic area. In *From Linguistic Areas to Areal Linguistics* [Studies in Language Companion Series 90], P. C. Muysken (ed.), 151–179. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.90>
- Danielsen, S. 2007. *A Grammar of Baure* [Indigenous Languages of Latin America 6]. Leiden: CNWS Publications.
- Danielsen, S. 2012. Baure. In *Lenguas de Bolivia, II: Amazonía*, M. Crevels & P. C. Muysken (eds), 295–339. La Paz: Plural editores.
- Danielsen, S. 2014. Multiple valency derivation in Baure (Arawakan). In *Incremento de valencia en las lenguas amazónicas*, F. Queixalós, S. Telles & A. C. Bruno (eds), 285–312. Bogotá: Instituto Caro y Cuervo.
- de Araujo, G. Antunes 2004. *A Grammar of Sabanê: A Nambikwaran Language* [LOT Dissertation Series 94]. Utrecht: LOT Publications. <[https://www.lotpublications.nl/Documents/94\\_fulltext.pdf](https://www.lotpublications.nl/Documents/94_fulltext.pdf)> (26 April 2020).
- Dixon, R. M. W. 2012. *Basic Linguistic Theory*, Vol. 3. Oxford: OUP.

- Eberhard, D. M. 2009. *Mamaindê Grammar: A Northern Nambikwara Language and its Cultural Context* [LOT Dissertation Series 236]. Utrecht: LOT Publications. <[http://www.lotpublications.nl/Documents/236\\_fulltext.pdf](http://www.lotpublications.nl/Documents/236_fulltext.pdf)> (26 April 2020).
- Emkow, C. 2006. A Grammar of Araona, an Amazonian Language of Northwestern Bolivia. PhD dissertation, La Trobe University.
- Epps, P. 2006. Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica* 23(2): 259–288. <https://doi.org/10.1075/dia.23.2.03epp>
- Epps, P. 2007. The Vaupés melting pot: Tukanoan influence on Hup. In *Grammars in Contact: A Cross-linguistic Typology*, A. Y. Aikhenvald & R. M. W. Dixon (eds), 267–289. Oxford: OUP.
- Epps, P. 2009. Language classification, language contact, and Amazonian prehistory. *Language and Linguistics Compass* 3(2): 581–606.
- Epps, P. & Stenzel, K. (eds). 2013. *Upper Rio Negro: Cultural and Linguistic Interaction in Northwestern Amazonia*. Rio de Janeiro: Museu Nacional – Museu do Índio – FUNAI.
- Eriksen, L. 2011. Nature and Culture in Prehistoric Amazonia: Using G.I.S. to Reconstruct Ancient Ethnogenetic Processes from Archaeology, Linguistics, Geography, and Ethnohistory. PhD dissertation, Lund University. <<https://lup.lub.lu.se/search/ws/files/3626162/1890749.pdf>> (26 April 2020)
- Fernández Garay, A. V. 2001. *Ranquel-Español / Español-Ranquel. Diccionario de una variedad mapuche de La Pampa (Argentina)* [Indigenous Languages of Latin America 2]. Leiden: CNWS.
- Fleck, D. W. 2013. *Panoan Languages and Linguistics* [Anthropological Papers of the American Museum of Natural History 99]. New York NY: American Museum of Natural History. <https://doi.org/10.5531/sp.anth.0099>
- Gabas, N. J. 1999. A Grammar of Karo, Tupí (Brazil). PhD dissertation, University of California, Santa Barbara.
- Gibson, M. L. 1996. *El munichi: Un idioma que se extingue* [Serie Lingüística Peruana 42]. Yarinacocha: Instituto Lingüístico de Verano – Ministerio de Educación.
- Gomes, D. M. 2006. Estudo morfológico e sintático da língua Mundurukú. PhD dissertation, Universidade de Brasília. <<http://repositorio.unb.br/handle/10482/3754>> (26 April 2020).
- Grinevald, C. & Seifart, F. 2004. Noun classes in African and Amazonian languages. *Linguistic Typology* 8(2): 243–285.
- Guillaume, A. 2008. *A Grammar of Cavineña* [Mouton Grammar Library 44]. Berlin: Mouton de Gruyter.
- Guillaume, A. 2011. Third person agreement and passive marking in Tacanan languages: A historical perspective. *International Journal of American Linguistics* 77(4): 521–536.
- Guillaume, A. 2012. A tale of two passives in Cavineña. In *Ergativity, Valency and Voice* [Empirical Approaches to Language Typology 48], G. Authier & K. Haude (eds), 111–131. Berlin: De Gruyter Mouton.
- Hall, J., McLeod, R. A. & Mitchell, V. 2004 [1987]. *Pequeno dicionário Xavante Português / Português Xavante*, revised and updated by A. Harrison. Cuiabá: Sociedade Internacional de Linguística.
- Hart, H. 1988. *Diccionario chayahuíta-castellano* [Serie Lingüística Peruana 29]. Lima: Instituto Lingüístico de Verano – Ministerio de Educación.
- Hawkins, J. A. & Gilligan, G. 1988. Prefixing and suffixing universals in relation to basic word order. *Lingua* 74(2–3): 219–259.
- Haynie, H., Bower, C., Epps, P., Hill, J. & McConvell, P. 2014. Wanderwörter in languages of the Americas and Australia. *Ampersand* (1): 1–18.
- Kohlberger, M. 2013. Multitasking in Shiwiar verbal morphology: Applicatives, objects and beyond. Friday Afternoon Lecture, February 15, 2013, LUCL, Leiden.

- Krasnoukhova, O. V. 2012. *The Noun Phrase in the Languages of South America* [LOT Dissertation Series 301]. Utrecht: LOT Publications. <[http://www.lotpublications.nl/Documents/301\\_fulltext.pdf](http://www.lotpublications.nl/Documents/301_fulltext.pdf)> (26 April 2020).
- Krasnoukhova, O. V. 2014. The noun phrase: Focus on demonstratives, redrawing the semantic map. In *The Native Languages of South America: Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 250–273. Cambridge: CUP.
- Lathrap, D. W. 1970. *The Upper Amazon*. London: Thames & Hudson.
- Martins, A., Camargos, L., Costa, L. & Miranda, M. 2009. Causativização em línguas do tronco Macro-Jê: Primeiras aproximações. In *Anais do SILEL – Simpósio Nacional e Internacional de Letras e Linguística*, Vol. 1. Uberlândia: EDUFU. <[http://www.ileel.ufu.br/anaisdosilel/pt/arquivos/gt\\_lg13\\_artigo\\_1.pdf](http://www.ileel.ufu.br/anaisdosilel/pt/arquivos/gt_lg13_artigo_1.pdf)> (26 April).
- Matras, Y. 2009. *Language Contact*. Cambridge: CUP.
- Matras, Y. 2014. Why is the borrowing of inflectional morphology dispreferred? In *Borrowed Morphology*, F. Gardani, P. Arkadiev & N. Amiridze (eds), 47–80. Berlin: De Gruyter Mouton.
- Meader, R. E. 1967. *Iranxe: Notas gramaticais e lista vocabular* [Publicações Série Diversos Linguística II]. Rio de Janeiro: Museu Nacional – Universidade Federal do Rio de Janeiro.
- Meira, S. 1999. A Grammar of Tiriyo. PhD dissertation, Rice University.
- Messineo, C. 2011. Aproximación tipológica a las lenguas indígenas del Gran Chaco. Rasgos compartidos entre toba (familia guaycurú) y maká (familia mataco-mataguayo). *Indiana* 28: 183–225.
- Mihás, E. 2010. Essentials of Ashéninka Perené Grammar. PhD dissertation, University of Wisconsin-Milwaukee.
- Mihás, E. 2015. *A Grammar of Alto Perené (Arawak)* [Mouton Grammar Library 69]. Berlin: De Gruyter Mouton.
- Muysken, P. C., Hammarström, H., Birchall, J. T. R., Danielsen, S., Eriksen, L., Galucio, A. V., van Gijn, R., van de Kerke, S., Kolipakam, V., Krasnoukhova, O. V., Müller, N. J. & O'Connor, L. M. 2014. The languages of South America: Deep families, areal relationships, and language contact. In *The Native Languages of South America: Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 299–322. Cambridge: CUP.
- Muysken, P. C., Hammarström, H., Birchall, J. T. R., van Gijn, R., Krasnoukhova, O. V., & Müller, N. J. 2015. Linguistic areas, bottom up or top down? The case of the Guaporé-Mamoré region. In *Language Contact and Documentation*, B. Comrie & L. Golluscio (eds), 205–237. Berlin: De Gruyter Mouton.
- Nimuendajú, C. 1926. *Die Palikur-Indianer und ihre Nachbarn*. Göteborg: Elanders Boktryckeri Aktiebolag.
- Nonato, R. 2008. Ainore Boe egore: Um estudo descritivo da língua Bororo e consequências para a teoria de caso e concordância. MA thesis, Unicamp, Campinas. <<http://libdigi.unicamp.br/document/?code=vtls000432270>> (26 April 2020).
- Oldendorp, C. G. A. 1996 [1767–68]. *Criolisches Wörterbuch: Erster zu vermehrender und wo nöthig zu verbessernder Versuch [1767/68], herausgegeben, eingeleitet und mit Anmerkungen versehen von Peter Stein, sowie das anonyme, Johan Christopher Korbitz Thomsen Kingo (J. C. Kingo) zugeschriebene Vestindisk Glossarium, herausgegeben, eingeleitet und mit Anmerkungen versehen von Hein van der Voort*. Tübingen: Max Niemeyer.
- Overall, S. 2007. A Grammar of Aguaruna. PhD dissertation, La Trobe University.
- Overall, S. 2017. *A Grammar of Aguaruna (Iiniá Chicham)* [Mouton Grammar Library 68]. Berlin: De Gruyter Mouton.
- Padua, A. J. 2007. Contribuição para a fonologia da língua Apiaká (Tupi-Guaraní). MA thesis, Universidade de Brasília. Available at <[http://www.etnolingua.org/local--files/tese:padua-2007/padua\\_2007\\_apiaka.pdf](http://www.etnolingua.org/local--files/tese:padua-2007/padua_2007_apiaka.pdf)>.



- Payne, D. 2001. Causatives in Asheninka: The case for a sociative source. In *The Grammar of Causation and Interpersonal Manipulation* [Typological Studies in Language 48], M. Shibatani (ed.), 485–505. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.48>
- Payne, D. L. 1985. -ta in Zaparoan and Peba-Yaguan. *International Journal of American Linguistics* 51(4): 529–531.
- Payne, D. L. 1990. Some widespread grammatical forms in South American languages. In *Amazonian Linguistics: Studies in Lowland South American Languages*, D. L. Payne (ed.), 75–87. Austin TX: University of Texas Press.
- Payne, T. E. 1997. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: CUP.
- Peeke, C. 1953. Análisis tentativo de la morfología andoa [Información de Campo no. 32-c]. Peru: Instituto Lingüístico de Verano. (Microfiche).
- Peeke, C. 1959. Shimigae, idioma que se extingue. In *Estudios acerca de las lenguas huarani (auca), shimigae, y zapara*, 18–28. Quito: Publicaciones Científicos del Ministerio de Educación del Ecuador.
- Peperkamp, M. G. J. 2005. Shiwilu Wichi<sup>2</sup>- Sueño Jeberino: Un estudio de la lengua Shiwilu y las voces nativas representado en las tradiciones orales de Jeberos, Alto Amazonas, Perú. MA thesis, Leiden University.
- Peterson, D. A. 2007. *Applicative Constructions*. Oxford: OUP.
- Plaza Martínez, P. 2009. Quechua. In *Lenguas de Bolivia, Tomo 1: Ámbito andino*, M. Crevels & P. C. Muysken (eds), 215–284. La Paz: Plural editores.
- Polinsky, M. 2013. Applicative constructions. In *The World Atlas of Language Structures Online*, M. S. Dryer & M. Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info/chapter/109>> (26 April 2020).
- Polinsky, M. 2017. Antipassive. In *The Oxford handbook of ergativity*, C. Jessica, D. Massam & L. D. Travis (eds), 308–331. Oxford: OUP.
- Queixalós, F. 2014. Incremental valence changes in an ergative vs. an accusative language: A comparison between Katukina-Kanamari and Sikuani. In *Incremento de valencia en las lenguas amazónicas*, F. Queixalós, S. Telles & A. C. Bruno (eds), 313–327. Bogotá: Instituto Caro y Cuervo.
- Rich, R. G. 1975. Sufijos verbales y apuntes gramaticales sobre el idioma Arabela. In *Datos etno-lingüísticos* 18 (12 pp.). Lima: Instituto Lingüístico de Verano.
- Rich, R. G. 1999. *Diccionario Arabela-Castellano* [Serie Lingüística Peruana 49]. Lima: Instituto Lingüístico de Verano.
- Riester, J. n.d. Unpublished manuscript list of Napeka/Kitemoka words, APCOB, Santa Cruz de la Sierra, Bolivia.
- Rodrigues, A. D. & Caspar, F. 1958. Zur Sprache der Tuparí-Indianer (Westbrasilien). <<http://www.etnolingüística.org/caspar:grammar>> (26 April 2020).
- Rose, F. 2008. A typological overview of Emerillon, a Tupi-Guarani language from French Guiana. *Linguistic Typology* 12(3): 431–460.
- Rowan, O. 2001 [1978]. *Iraiti xawaiyehelakatyakala: Dicionário Paresí - Português*. Cuiabá: Sociedade Internacional de Lingüística.
- Sakel, J. J. 2004. *A Grammar of Mosestén* [Mouton Grammar Library 33]. Berlin: Mouton de Gruyter.
- Seifart, F. 2012. Causative marking in Resígaro (Arawakan): A descriptive and comparative perspective. *International Journal of American Linguistics* 78(3): 369–384.
- Seifart, F. & Payne, D. L. 2007. Nominal classification in the North West Amazon: Issues in areal diffusion and typological characterization. *International Journal of American Linguistics* 73(4): 381–387.

- Seki, L. 1999. The Upper Xingu as an incipient linguistic area. In *The Amazonian Languages*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 417–430. Cambridge: CUP.
- Shibatani, M. & Pardeshi, P. 2002. The causative continuum. In *The Grammar of Causation and Interpersonal Manipulation* [Typological Studies in Language 48], M. Shibatani (ed.), 85–126. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.48>
- Silva, L. de Jesus. 2011. Morphosyntaxe du Rikbaktsa (Amazonie brésilienne). PhD dissertation, Université Denis Diderot, Paris 7. <[http://www.etnolinguistica.org/tese:silva\\_2011b](http://www.etnolinguistica.org/tese:silva_2011b)> (26 April 2020).
- Singerman, A. R. 2018. The Morphosyntax of Tuparí. A Tupían Language of the Brazilian Amazon. PhD dissertation, University of Chicago. <<http://www.etnolinguistica.org/tese:singerman-2018>> (26 April 2020).
- Smeets, I. 2008. *A Grammar of Mapuche*. Berlin: Mouton de Gruyter.
- Song, J. J. 1990. On the rise of causative affixes: A universal-typological perspective. *Lingua* 82: 151–200.
- Sorensen, A. P. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69(6): 670–684. <<https://anthrosource.onlinelibrary.wiley.com/doi/pdf/10.1525/aa.1967.69.6.02a00030>> (26 April 2020).
- Stenzel, K. 2005. Multilingualism in the Northwest Amazon revisited. *Proceedings of the Conference on Indigenous Languages of Latin America-II, University of Texas at Austin*, 1–28. <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.422.6762&rep=rep1&type=pdf>> (26 April 2020).
- Stenzel, K. & Gómez-Imbert, E. 2009. Contato lingüístico e mudança lingüística no noroeste amazônico: O caso do Kotiria (Wanano). *Revista da ABRALIN* 8: 71–100.
- Swift, K. E. 1988. *Morfología del Caquinte (arawak preandino)* [Serie Lingüística Peruana 25]. Yarinacocha, Peru: Ministerio de Educación – Instituto Lingüístico de Verano.
- Telles, S. 2002. Fonología e gramática Latundê/Lakondê. PhD dissertation, Vrije Universiteit Amsterdam.
- Tripp, R. 1995. *Diccionario amarakaeri-castellano*. Yarinacocha: Ministerio de Educación e Instituto Lingüístico de Verano.
- Tuggy, D. 1988. Nahuatl causative/applicatives in cognitive grammar. In *Topics in Cognitive Linguistics* [Current Issues in Linguistic Theory 50], B. Rudzka-Ostyn (ed.), 587–618. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.50>
- Valenzuela, P. 2011. Argument encoding and pragmatic marking of the transitive subject in Shiwilu (Kawapanan). *International Journal of American Linguistics* 77(1): 91–120.
- Valenzuela, P. & Butler, L. 2009. Argument encoding and valence changing in Kawapanan. Paper presented at the SSILA Annual Meeting, San Francisco, 8–10 January, 2009.
- Vallejos, R. 2014. Cambio de valencia en Kokama-Kokamilla. In *Incremento de valencia en las lenguas amazónicas*, F. Queixalós, S. Telles & A. C. Bruno (eds), 261–283. Bogotá: Instituto Caro y Cuervo.
- van Baarle, P., Sabajo, M. A., Sabajo, A. L., Sabajo, L. L. & van der Stap, G. 1989. *Arhwaka Lokonong Djang: Arowakse taalkursus en woordenboek*. Haarlem/Amsterdam: Sociaal-culturele vereniging Ikyoshie and Instituut voor Algemene Taalwetenschap.
- van de Kerke, S. 2009. Leko. In *Lenguas de Bolivia, I: Ámbito andino*, M. Crevels & P. C. Muysken (eds), 287–331. La Paz: Plural editores.
- van der Voort, H. 2004. *A Grammar of Kwaza* [Mouton Grammar Library 29]. Berlin: Mouton de Gruyter.
- van der Voort, H. 2005. Kwaza in a comparative perspective. *International Journal of American Linguistics* 71(4): 365–412.

- van der Voort, H. 2013. Fala fictícia fossilizada: O tempo futuro em Aikanã. *Boletim do Museu Paraense Emílio Goeldi* (Ciências Humanas) 8(2): 359–377. <<http://www.scielo.br/pdf/bgoeldi/v8n2/09.pdf>> (26 April 2020).
- van der Voort, H. 2015. Sistemas de classificação nominal no Sudoeste amazônico. *Moara: Revista Eletrônica do Programa de Pós-Graduação em Letras* 43(2): 5–22.
- van der Voort, H. 2016. Recursive inflection and grammaticalized fictive interaction in the southwestern Amazon. In *The Conversation Frame: Forms and Functions of Fictive Interaction* [Human Cognitive Processing 55], E. Pascual & S. Sandler (eds), 277–299. Amsterdam: John Benjamins. <https://doi.org/10.1075/hcp.55>
- van der Voort, H. 2018. Development and diffusion of classifier systems in Southwestern Amazonia. *The Diachrony of Classification Systems* [Current Issues in Linguistic Theory 342]. W. B. McGregor & S. Wichmann (eds), 201–240. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.342>
- van Gijn, R. 2006. A Grammar of Yurakaré. PhD dissertation, Radboud University. <<http://repository.uibn.ru.nl/handle/2066/27381>> (26 April 2020).
- van Gijn, R. 2014. Subordination strategies in South America: Nominalization. In *The Native Languages of South America: Origins, Development, Typology*, L. M. O'Connor & P. C. Muysken (eds), 274–296. Cambridge: CUP.
- van Gijn, R. 2015. Verbal synthesis in the Guaporé-Mamoré linguistic area: A contact feature? *Linguistic Discovery* 13(2): 96–122. <<https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/464>> (26 April 2020).
- Van Linden, A. Forthcoming. Harakmbut. In *Amazonian Languages, Vol. 1: Smaller Language Families and Isolates* [HSK 44/1], P. Epps & L. Michael (eds). Berlin: De Gruyter Mouton.
- Wapishana Language Project. 2000. *Scholar's Dictionary and Grammar of the Wapishana Language - Tominpainao Ati'ò Wapichan Paradan Parada-karu Na'iki Paradauzo-karu Kaduzu*. San Jose via Lethem, Rupununi: Wapishana Language Project.
- Winford, D. 2003. *An Introduction to Contact Linguistics*. Malden MA: Blackwell.
- Wise, M. R. 1990. Valence-changing affixes in Maipuran Arawakan languages. In *Amazonian Linguistics: Studies in Lowland South American Languages*, D. L. Payne (ed.), 89–116. Austin TX: University of Texas Press.
- Wise, M. R. 1993. Algunas interrogantes en la clasificación de las lenguas indígenas sudamericanas. Paper presented at the 10th International Congress of the Asociación de Lingüística y Filología de la América Latina (ALFAL), Veracruz, Mexico, April 11–16, 1993.
- Wise, M. R. 1999. Small language families and isolates in Peru. In *The Amazonian Languages*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 307–340. Cambridge: CUP.
- Wise, M. R. 2002. Applicative affixes in Peruvian Amazonian languages. In *Current Studies on South American Languages* [Indigenous Languages of Latin America 3], M. Crevels, S. van de Kerke, S. Meira & H. van der Voort (eds), 329–344. Leiden: CNWS.
- Zúñiga, F. 2009. The applicatives of Mapudungun. Paper presented at the SSILA Annual Meeting, San Francisco, 8–10 January, 2009.
- Zúñiga, F. 2013. Some valency-neutral applicatives in the Americas. Paper presented at the Workshop *Valency Classes in the World's Languages*, Max Planck Institute for Evolutionary Anthropology, Leipzig, August 14–15, 2013.

# Transfer of Swahili ‘until’ in contact with East African languages

Maarten Mous  
Leiden University

Swahili has transformed the noun *mpaka* ‘boundary, border’ into a function word ‘until’, which has successfully spread to many other East African languages with locative and temporal readings. The grammaticalisation originated in a N-N construction without an associative ‘of’ interpreted as limiting the action adverbially. The main function is in the time interpretation of ‘until’. I provide an overview of this transfer in East Africa by looking at a large number of languages and argue that parallel independent grammaticalisation is not what is at stake but rather transfer of the function word and the preposition-like function.

**Keywords:** grammaticalisation, borrowing, East Africa, ‘until’

## 1. Introduction

The Swahili word *mpaka* is a class 3 noun meaning ‘border, boundary’ and is also used as a function word meaning ‘until’, for both locational and temporal goals. A function word for ‘until’ *mpaka* is used in most languages of East Africa. In many of these languages we can show that *mpaka* (or variants thereof) is a borrowing from Swahili. The function word *mpaka* resembles a preposition: it is preposed to the noun with no additional material, and it links the noun to the state of affairs expressed in the verb. For the many Bantu languages in East Africa an alternative view to this borrowing scenario is conceivable, namely that of independent parallel grammaticalisation of the noun ‘border, boundary’ to ‘until’. While this scenario cannot be ruled out for all the languages under discussion, we can show that this has not been the case for most of them.

## 2. The grammaticalisation of ‘boundary’ to ‘until’

In Swahili we can observe the grammaticalisation of the word for ‘boundary, border’, *mpaka*, to have developed the function of expressing ‘until’. It is a clear instance of grammaticalisation. Heine and Dunham (2010) in their article on grammaticalisation in Swahili and other Bantu languages provide a framework for grammaticalization of four parameters: (i) extension; (ii) desemanticization; (iii) decategorialization; (iv) erosion. I discuss these parameters one by one.

Parameter (i) extension: *Mpaka* ‘boundary’ has extended by pragmatic use in not only referring to the physical entity of a boundary but also using the concept of boundary as a limit for the state of affairs expressed in the verb. When a noun phrase “boundary X” is interpreted as functioning as an adverbial phrase the interpretation is that the boundary is the limit of the state-of-affairs expressed in the verb and hence translatable with ‘until’, see Krifka (1983: 125). In (1b) the constituent with boundary is interpreted as an adjunct. With that extension in pragmatic use, the interpretation of ‘boundary’ extends or shifts to ‘until’.

- (1) a. I see [the boundary of X]  
 b. I walk [the boundary X] = ‘I walk until X’  
 c. I walk [the boundary tomorrow] = ‘I walk until tomorrow’

Once the constituent with *mpaka* is interpreted as an adverbial phrase it is very easy to use a temporal noun in the position of X replacing a locative noun. But in that case *mpaka* no longer refers to the boundary of the temporal noun; boundary is no longer conceptualised as referring to X but as only limiting the state of affairs expressed in the verb; rather than the boundary of tomorrow, the concept of boundary of walking is evoked: “I walk until tomorrow”.

The crucial drive behind the extension is the interpretation of a noun phrase as an adverbial phrase. This in itself is very common in Swahili syntax (Krifka 1985: 125).

Parameter (ii) of desemanticization goes hand in hand with the extension sketched above. The semantic development is that the physical boundary becomes a limit to the action as a corollary of the constituent functioning as adverbial phrase. Limiting the action of the verb has as a consequence that a time interpretation comes to the foreground. As soon as what follows *mpaka* is an indication of time, this moment in time can be expressed just as easily in a clause as in a noun.

Parameter (iii) of decategorialization is again a consequence of the word *mpaka* now referring to the limit of an action and hence all properties of decategorialization of nouns that Heine and Durham mention apply to *mpaka* which loses “morphological distinctions of number, gender, case, etc., the ability to combine with adjective, determiners, etc., to be headed by adpositions...” (Heine & Durham 2010: 34).

Parameter (iv) of erosion is possibly also at stake in the reduction of the initial nasal of *mpaka* from a syllabic nasal to pre-nasalisation. However, I am not certain about the facts here. On the one hand, such an erosion is reported in the recent dictionary *Kamusi kuu ya Kiswahili* (Baraza la Kiswahili 2015) which indicates this distinction with a different placement of the stress mark: *m'paka* 'boundary' versus '*mpaka* 'until'. The syllabic nasal of the noun *mpaka* 'boundary' is the nominal class prefix of class 3. The initial nasal of the functional word 'until' is invariable and in the absence of any agreement to the function word 'until' there is little incentive to consider its initial nasal element as a separate morpheme and reduction to prenasalisation is favoured. However, I am not entirely sure that the reduction to prenasalisation is a linguistic fact. Swahili poets such as Farouk Topan, whom I consulted during the Swahili colloquium in Bayreuth in 2016 did not perceive any rhythmic difference between the two forms and conceived the form for 'until' as containing three syllables.

A final factor that needs to be discussed is the absence of an associative in the *mpaka X* construction. Swahili Noun + Noun constructions usually contain an associative, *mpaka wa X*, with *wa* referring to class 3 of *mpaka*. This issue is dealt with in Section 3.

Grammaticalisation from 'boundary' to 'until' seems perfectly logical but is in fact not very common. The lexicon of grammaticalization (Heine & Kuteva 2002: 61) mentions only Swahili and Mooré. I am not convinced, however, that the Moore example shows grammaticalization to 'until'. The most extensive dictionary of Mooré, Nikiema and Kinda (1997: 921), does indeed suggest grammaticalisation of the word for boundary as it gives for *teka* 'fin, limite' as well as 'à partir de, depuis' but the conceptual extension is rather different, developing into a point of departure rather than a telic goal.

But the same grammaticalisation happened in Arabic as I learned from my colleague Ahmad Al-Jallad. In Arabic *ḥattā* 'until' has undergone similar grammaticalisation from the noun *ḥadd* 'boundary' followed by a relative marker yielding *ḥaddī-tā* /border-which/ (this form exists in the Levantine dialects). The further developments in form are > *ḥaddtā* > *ḥattā*.<sup>1</sup> The same simplification of *ddt* happened in the parallel grammaticalisation of the conjunction *ḥadd-tay* > *ḥattay* 'until' with *tay* being a subordinating element (Al-Jallad 2017).

Next to the grammaticalisation of 'boundary' to 'until' Swahili has borrowed *hata* or *hati* from Arabic for 'until'. Thus, it may seem that the grammaticalisation of *mpaka* in Swahili is modelled on Arabic. However, that would require to assume quite intimate knowledge and analysis of Arabic on the part of the speakers that

1. In Arabic the conjunction can also express purpose, concession.

recognised the model at the time of grammaticalisation. Moreover, the grammaticalisation in Arabic including the devoicing of *dd* occurred in pre-Quranic times and hence long before the first contact of Arabic with Swahili. Therefore, I conclude that the Swahili grammaticalisation is not modelled on the Arabic one. The conceptual shift from ‘boundary’ to ‘until’ is the same in Arabic and in Swahili but also very straightforward. The developments in form are quite different. While the Arabic preposition and conjunction contain second elements expressing the relation, the Swahili preposition crucially does not include the relating associative. Swahili (or Bantu) internally such an inclusion of an associative in the preceding element is indeed not to be expected because an intonational brake in such a construction in a Bantu language would fall before the associative and certainly not after it and hence an associative would not merge with the preceding word for ‘boundary’ or ‘until’.

### 3. The construction with *mpaka* ‘until’ in Swahili

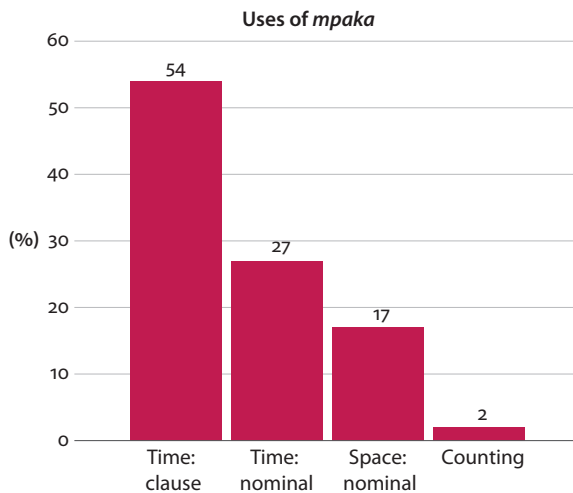
The Swahili dictionaries recognise the preposition-like usage of the word *mpaka* for ‘boundary’. Sacleux (1939) has ‘jusque’ with locative and time complements. Johnson (1939) derives *mpaka* ‘boundary’ from the verb *paka* ‘make a boundary’ and gives for the prepositional usage the following equivalents: ‘up to, to, as far as, till, until, to the time of’, and gives the words *hata*, *hadi* as comparable to *mpaka* and as preposition borrowed from Arabic.<sup>2</sup> TUKI (2001) has two entries for *mpaka*; one as the noun for ‘border’ and one as a preposition ‘until, to, up to’. TUKI (1981), a Swahili-Swahili dictionary, gives the Swahili equivalents *hadi* and *hata*.

In order to understand the transfer process, we should have a closer look at the meaning, functions and uses of *mpaka* in Swahili. To this end I consulted the Helsinki Swahili Corpus (HSC 2.0). The corpus contains 15,093 instances of *mpaka* (a comparable word such as *katika* ‘in’ has 281,094 instances). A first cursory investigation taking the first 100 instances of *mpaka* in the Helsinki Swahili corpus shows 54 instances with a sentential complement indicating a moment in time, 27 with a nominal complement indicating a point in time, and only 17 that have a locative complement and these are all nominal in nature, 2 instances that are used in “counting” (e.g. *tano mpaka nane* ‘five to eight’), and only 1 instance of ‘border’ was noted. The percentages are displayed in the following pie diagram. This shows an overwhelming majority of use with time complements and of use as clause introducer.<sup>3</sup>

---

2. Presently *hata* X is used as emphazier ‘even X’ rather than for ‘until’. Johnson (1939) also mentions *upeo* ‘limit’ used adverbially as intensifier of the verb rather than preposed to a noun, from the verb *pea* ‘become fully grown’

3. I have counted infinitival complements, which are grammatically nominal, as clausal since they too require the interpretation of an event as a moment in time.



Very common collocations with *mpaka* in the corpus are *mpaka sasa* ‘till now’ (1906 of 15093 or 12.6% of the instances of *mpaka*), others are *mpaka leo* ‘till today’ (850), *mpaka kesho* ‘till tomorrow’ (136), and *mpaka mwisho* ‘till the end’ (165). The use of *mpaka* as ‘border, boundary’ is far from obsolete. The Helsinki corpus contains at least some 3500 instances.<sup>4</sup>

Perusing the Helsinki corpus, I came across only two instances of *mpaka* used in the meaning of ‘until’ but followed by an associative marker (see discussion below why *mpaka* is not followed by an associative). Both instances came from the part of the corpus that is based on novels and not from the news and parliamentary proceedings. I consider these uses to be prescriptive attempts. Both examples come from materials for a university course in literature in Dar es Salaam’s Swahili department where correct Swahili is moulded.

- (2) *Mpango huu, u-na-o-patikana Mombasa mpaka wa leo*  
 3.plan 3:this 3-PRES-3.REL-available Mombasa until 3:of today  
*huko u-na-itwa ku-funga nyama*  
 3:this.one 3-PRES-call:PAS 15-close meat  
 ‘This plan, which is available in Mombasa, until today this is called “to close meat”’  
 (from HSC 2.0: *Makala za Semina ya Kimataifa ya Waandishi wa Kiswahili III Fasihi*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili cha Chuo Kikuu, 1983)

4. Obviously this is not easy to count such instances by looking for form only but I counted instances of *mpaka* plus associative ‘of’ (394), the plural form, *mipaka* (1971), and the locative forms *mpakani* ‘at the border’ (748), and its plural *mipakani* (338).



- (3) *Maelezo haya ya-me-vuka mpaka wa ku-ona ki-tu*  
 6.explanation 6:these 6-PRF-cross until 3:of 15-see 7-thing  
*kama ki-livyo*  
 like 7-it.is

‘These explanations extended until seeing something as it is.’

(from HSC 2.0: *Makala za Semina ya Kimataifa ya Waandishi wa Kiswahili III Fasihi Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili cha Chuo Kikuu*, 1983)

In short, the use of *mpaka* for ‘until’ in Swahili may indicate a location as endpoint of a path with focus on that end point, or it may indicate a goal. A goal is often expressed without the addition of *mpaka*; for example, *rudi nyumbani* ‘return home’. In examples of locational use of *mpaka* ‘until’, the occurrence of a path is common, but expressing a goal with *mpaka* without a path is not impossible.

By far the most common use of *mpaka* ‘until’ is for time. *Mpaka* indicates a time period with an endpoint and focus on that endpoint. *Mpaka* is also used in series of comparable entities (such as counting) with focus on their endpoint. Time is the major function of *mpaka* in actual language use. It offers the speaker an option to express a period of time as adjunct and hence also easily functions as a clause introducer.

The construction with *mpaka* ‘until’ is different from other prenominal nouns that have developed into function words in Swahili that contain the associative like *mbele ya* ‘in front of’, *kati ya* ‘in the middle of’, *juu ya* ‘on top of’, *ndani ya* ‘inside of’, *nyuma ya* ‘behind of’. The *mpaka* ‘until’ construction fits better in a set of other constructions with a preplaced noun and no associative such as *hadi N* ‘until N’, *tangu N* ‘since N’, *toka N* ‘from N’, *katika N* ‘in N’, *kama N* ‘like N’, *bila N* ‘without N’, *kuliko N* ‘than N’, *kila N* ‘every N’, *hata N* ‘even N’ (see Kapinga 1983). What these structures have in common is that they lack an associative and are often translated using prepositions but the set is not at all homogeneous; some of the initial elements are borrowed from Arabic (*hadi*, *hata*, *bila*, *kila*) other are not; some are verbal in origin (*toka* from *ku-toka* ‘to come from’) or originate in a construction (*ku-li-ko* 17-be-17:REL). The relevance of this list for understanding *mpaka X* is that there are models in the language for this construction.<sup>5</sup>

Next to the N associative N construction Swahili also has N N constructions where the second noun is not a modifier of the first noun but rather an equivalent of the first noun. This is the case in a construction like *ndugu X* ‘brother/friend X’ which is used in addressing people: X = brother, the meaning is not ‘brother of X’.<sup>6</sup> The conceptual frame is not one of modification and this is also the case in the

5. However, Makhuwa may use the associative in the function of ‘until’, in *mpákhá wa-ámútsy’ ááwe* ‘to his family’s place’ (van der Wal 2009: 48, example 129).

6. The structure resembles Swahili compounds such as *nya+X* ‘mother=X’ as in *Nyamwezi*.

*mpaka X* construction: *mpaka sasa* 'until now' does not develop from 'border of now' but from 'border is now'. The expression of time is semantically not a modification of the boundary and, therefore, the structure is not one with the associative marker. When *mpaka* is followed by a time expression, the time expression is a point on time and not a period of time. For example, *mpaka saa kumi* means 'until 4 o'clock' and cannot refer to the boundary of ten hours. Likewise, with locative complements the semantics evoked are boundary = X and not boundary of X; for example, *mpaka nyumbani* 'until home' does not refer of the limits of home but considers home as a limit.<sup>7</sup>

I propose that the origin of the until-construction is a N+N construction and not the N associative N construction on the basis of its semantics.<sup>8</sup> It has to be noted though that N associative N constructions can drop their associative. Examples are *maji moto* /water fire/ 'hot water' where *moto* does modify *maji* and the origin *maji ya moto* exists too; likewise, *gari moshi* /vehicle smoke/ 'train' as alternative to *gari la moshi*.<sup>9</sup>

When 'until' is followed by an expression indicating a moment in time the constituent with 'until' refers to the whole period from now (or another starting point) till the moment expressed after 'until' with emphasis on that endpoint. It cannot refer to only the end point. An example is 'You didn't bother until today, why do you tell me now?' (Ström 2013: 270, 1.583).

When the complement is a location, two interpretations are possible. One is comparable to the period of time and refers to the path from here (or another point) to the location expressed after 'until', as in Example (4):

- (4) *Makuyuni i-po ktk njia i-toka-yo Arusha ku-elekea*  
 M. 9-16.REL middle 9.road 9-come.from-9.REL A 15-direct  
*Babati mpaka Dodoma*  
 B until D.  
 'Makuyuni is on the road halfway from Arusha direction of Babati until Dodoma.'<sup>10</sup>

7. There is a third option with the general preposition *na* 'and, with' instead of the associative, *mpaka na X*. This construction is almost exclusively attested in the meaning of 'boundary' in the HSCorpus, after the English model 'boundary with'.

8. In Kagulu, locational nouns such as 'near' and 'inside' equally need no associative (Petzell 2008: 59–60).

9. I thank one of the reviewers for pointing this out to me.

10. From <<http://tembeatz.blogspot.nl/2010/08/njia-panda-ya-makuyuni-mpaka-mto-wa-mbu.html>> (8 December 2016).

With locational complement of *mpaka* reference can either be made to the whole path as in (4) or only to the endpoint as in (5).<sup>11</sup> The same is true for languages borrowing ‘until’ from Swahili.

- (5) *ni-ka-m-geuza farasi wangu ni-ka-anza ku-rudi kwa haraka*  
 1SG-CSEC-O1-change horse 1:mine 1SG-CSEC-start 15-return with hurry  
*mpaka nyumbani*  
 until home  
 ‘and I changed horses and started to return home in a hurry’.

(from HSC 2.0 Salim A. Kibao 1975. Matatu ya Thamani Nairobi:  
 Heinemann Educational Books)

This expression of path can also be found in languages borrowing *mpaka* from Swahili. In Konso, we find ‘Sweep this house up to the gate! (Orkaydo 2013: 181). In Mbugu we have an example expressing locational path *kulí kinhko mpaká mag-amba ni muda wa madamo mahai to* ‘(the distance) from Kinko until Magamba is a period of four hours only’ (Mous 2003: 197). In this last example there is no movement involved; expression of path is more common with movement as in the Rangi example ‘S/he is fat, s/he will not be able to go to Kondoia by foot’ (Gibson 2013: 83). For time we often need to refer to periods but for location, reference for path may be relatively less common. There are plenty examples where only goal is indicated using ‘until’ and no path is intended as in Ndenggeleko *Ywaa nkólongwa kaabóya mpaka ukááya* ‘That/The man returned home’ (Ström 2013: 178 example 302).

#### 4. Borrowing *mpaka* in East Africa

The function word *mpaka* ‘until’ is remarkably often borrowed into the languages of East Africa. Most languages in East Africa use a word like *mpaka* for ‘until’. In the Table in the Appendix I present the results of searching for *mpaka* ‘until’ in East Africa for more than 30 languages. Some non-Bantu languages borrow *mpaka* from Swahili. For the non-Bantu languages without *mpaka* I included information in the

---

11. The combination *mpaka nyumbani* is rare though in the Swahili corpus and occurs only 29 times. A few times *mpaka nyumbani* occurs with the verb *sindikiza* ‘to accompany a guest for a while upon departure’ in which it probably expresses the unexpectedness of the endpoint. This may also be the reason to choose *mpaka nyumbani* with the verb *rudi* ‘to return, go back’ over simply *nyumbani*. Simply *nyumbani* is far more common as complement of the verb *rudi*; over 750 instances can be found of *nyumbani* with *rudi* against only two instances of *mpaka nyumbani* in HSC 2.0. In the example above *mpaka* may have been used to express the unexpectedness too, ‘all the way home’.

table about how they express 'until'. I found some 18 Bantu languages where there is no evidence for borrowing of 'until' from Swahili.<sup>12</sup>

The grammaticalisation of 'boundary' to 'until' in Swahili has a parallel in Arabic with differences in the detail, as mentioned above; this parallel was considered to be coincidental and not due to historical influence. Swahili may not have copied this model from Arabic, but did the other languages in East Africa copy the model from Swahili and followed the same path of grammaticalisation and developed 'until' from the noun for 'boundary'? Most people in East Africa are speakers of other languages but know Swahili well enough to be able to connect the noun and the function word in Swahili. It is thus perfectly imaginable that the East African Bantu languages derived the function word from their own reflex of the word for 'boundary' following the Swahili model. Yet I will argue that for many East African Bantu languages it can be shown that they did not. Rather they borrowed *mpaka* from Swahili in the function of 'until'.

A number of non-Bantu East African languages have borrowed the Swahili preposition *mpaka* for 'until'. These are at least Ik (Kuliak, Uganda), Borana (Cushitic, Kenya, Ethiopis), Konso (Cushitic, Ethiopia) and Sandawe (Khoisan, Tanzania). The word is often adjusted to the phonotactics of the receiving language that does not like nasal-oral stop compounds. It is obvious that in these cases only the function word *mpaka* is borrowed because these languages have completely different words for the concept of 'boundary'. For some of these languages there is a competing word expressing 'until'. Konso is an interesting case since Swahili is not strong in South Ethiopia and the most likely scenario is that Konso borrowed it from Borana that borrowed it in turn from Swahili; the pronunciation of Konso *haka* is actually identical to Borana *haga* despite the difference in spelling because intervocalic stops are voiced in Konso (Orkaydo 2013: 39–40).

There is no indication of such borrowing in some other East African non-Bantu languages such as Iraqw (Cushitic, Tanzania), Rendille (Cushitic, Kenya), Maasai (Nilotic, Kenya, Tanzania), Turkana (Nilotic, Kenya). We cannot rule out the possibility that these languages actually do use *mpaka*. In some grammars and dictionaries there is no mention of *mpaka* but one does frequently encounter instances of *mpaka* 'until' in texts. These cases are apparently considered instances of code-switching, or such a grammar writer has a purist stance favouring inherited function words above borrowed ones.

---

12. Makonde (Leach 2010), Kiitharaka (Muriungi 2009), Luluyia (Marlo 2007), Yao (Ngunga 2000), Rutara (Muzale 1989), Bukusu (Mutonyi 2000), Jita (Kagaya 2005), Digo (Nicolle (*hata* 'until' and *mpaka* 'boundary'), ChiNgoni (Ngonyani 2003), Lingala (Edema 1994 *kiná conj*), Gogo (Rugemalira 2009 Instead Gogo uses *sunga* +S)<sup>13</sup>, Ha (Harjula 2004), Pangwa (Stirnimann 1983), Rimi (Olson 1964), Nyiha (Busse 1960), Mwera (Harries 1950), Daiso (Nurse 2000), Kuria (Cammenga 2004).

It is equally clear that a number of Bantu languages borrowed *mpaka* as ‘until’ because their word for ‘border’ is completely different: Nyamwezi, Kivunjo Chaga, Luganda, Matengo, Ekoti, Haya, see the table in the Appendix for details. For example, Luganda has borrowed *paka* for ‘until’ from Swahili while it has several different words for ‘boundary’: *èsalè*, *èmbobò*, *enkingi*.

And in some other Bantu languages the word for ‘until’ resembles *mpaka* while the word for ‘boundary’ is a clearly different even if it is a reflex of the same root: Mbugwe, Rangi, Pare, Mbugu, Ndamba, Zigula, Shambala, Gweno. For example, Ndamba has *mupaka* ‘boundary, border’, but *mbaka* ‘until’. In this instance the Swahili loan is adapted to the phonological system of the receiving language while it remains different from the regular reflex of the Bantu noun root for ‘boundary’. In some other languages the loan is not adapted at all. For example, Mbugu *mpaka* ‘until’ shows the prenasalised voiceless stop which is not otherwise attested in Mbugu because such prenasalisations need to be voiceless in the language as in *mhpaka* ‘boundary’. These languages borrowed the Swahili preposition ‘until’.

There is another category of languages for which we have a strong indication that the word for ‘until’ is borrowed from Swahili in the similarity in form but we cannot contrast it to the word for ‘boundary’ due to lack of information: Ndengeleko *mpaka* ‘until’, Makonde *mpáaka* ‘until’.

For some languages we cannot rule out a parallel grammaticalisation scenario; these languages are presented in the last part of the table in the Appendix. This is particularly the case if the word for ‘until’ is not identical to the Swahili form. For example, Ilwana has a form for ‘until’ that is slightly different from Swahili namely *mupaka*. Nurse (2000: 210) indicates that the preposition is borrowed from Swahili; *mupaka* being an automatic adaptation of Swahili *mp* to Ilwana phonology which does indeed render all syllabic nasals of Swahili loans as *mu* (while Swahili prenasalised stops are maintained as such, see Nurse 2000: 115). Thus, a scenario that Ilwana borrowed the function word from Swahili is indeed likely. However, their word for ‘boundary’ is identical and presumably also borrowed from Swahili. Hence, we cannot be sure whether both are borrowed in parallel or whether Ilwana has developed a parallel grammaticalisation of noun ‘border’ to functional element ‘until’. In Matengo both ‘until’ and ‘boundary’ are *mpaka* though the word for ‘boundary’ is indicated as having a syllabic nasal and the word for ‘until’ is not, but I cannot be sure that there is truly such a difference. Such a difference in syllabicity is also reported for Pare or Chasu: it has *mháka* for ‘until’ and *mháka* with a syllabic nasal for ‘boundary’. It is possible that *mhaka* is the automatic adjustment in Pare for prenasalised voiceless stops that the language does not allow. But the closely related Mbugu breaks that restriction when it borrows *mpaka* from Swahili, next to the inherited noun *mhpaka* with a voiceless nasal for ‘boundary’. Thus, the borrowing scenario requires automatic phonological adjustment of the loan while the parallel grammaticalisation scenario requires a reduction in form from a voiceless nasal

stop compound to voiced nasal plus *h*. I consider this latter reduction less likely because the existing variation in closely related Mbugu is between a voiceless nasal plus voiceless stop compound [mp] and just a voiceless nasal without the stop part [m]; and there is no indication of variation between (aspirated) *p* and *h*, and the nasal in this context in inherited words is always voiceless. Hence, here too I would opt for the borrowing scenario. In sum, there is no strong evidence that parallel grammaticalisation of 'boundary' along the Swahili model is what is at stake in any of the East African languages.

The word *mpaka* for 'until' has spread from Swahili to many languages of East Africa in this function. Its use in Swahili and the recipient East African languages goes in tandem in a situation of extensive bilingualism. The main use is expressing a temporal period by indicating an end point. The construction *mpaka* + Noun or Clause is grammaticalised in Swahili from a conceptual frame 'boundary' = X. The general construction is not common in the East African languages but in this instance of *mpaka* it is part of Swahili grammar and copied as such in the grammars of the recipient languages.

A conceivable reason for the easy transfer of *mpaka* is that bilingual speakers appreciate the efficiency of a constituent with an initial element expressing an endpoint and feel the need for that in their other language(s). The attraction of a preposition-like *mpaka* is the fact that the relationship is expressed in a separate word dedicated to that relationship. This has advantages over an alternative in which this needs to be inferred from other relators with wider functions or from a state of affairs indicating duration and expressed in the verb in combination with a noun phrase interpreted but not marked as endpoint. This explains for example why some languages do not borrow *mpaka*. Some of those East African languages express the notion of 'until' with a comparable preposition-like element with a wider application. Iraqw, a Southern Cushitic language in Northern Tanzania uses the general preposition *ay* 'to' in *ay hamti* 'till now', *ay dír deelodá* '/to place:of day:that/ 'until that day' but also *ay dír* + relative clause expressing point in time for 'until'. When asked in elicitation speakers give *afiqoomár* /mouth:period:of/ as equivalent of 'until'.<sup>13</sup>

In a similar vein, it could be raised as an argument why Maasai does not use *mpaka*. Maasai has a proclitic to a time adverbial or a subordinate clause that is used "to indicate that the main clause event continues until the time or situation indicated by the time adverbial of subjunctive clause event." (Payne & Ole-Kotikash 2008: lemma ɔ).<sup>14</sup>

13. *Mpaka* is also used occasionally in elicitation but it is not used in texts

14. Mol (1996: 254) mentions for Maasai the use of verbs in subjunctive 'to reach' and 'to be big' to express 'until'.

The argument of experienced efficiency as incentive for the transfer can however not explain why *mpaka* would be borrowed by those languages that have a comparable competitor for *mpaka*. In some cases, the Swahili loan is clearly replacing a similar item with seemingly identical function. For example, in Ik *páka* and *gone* seem interchangeable. In Rangi the Swahili replacement is considered to be equivalent to the Rangi original but the Swahili form is simply more often used (Stegen 2011: 191). Shambala has two competing elements *shuti* and *mpàkà*.<sup>15</sup> *Mpaka* is used with clause complements referring to a moment in time in the examples that LangHeinrich supplies and he provides ‘till that [bis dass]’ as translation equivalent but he indicates that it can also be followed by locative pronouns while *shuti* has the same kind of complements. *Shuti* has a homonym in his dictionary that means ‘it must be that’ and this is followed by a dependent verb form while it is invariant itself (like Swahili *lazima*). Roehl (1911: 103) reports only this latter function for *shuti* and already in 1911 gives the Swahili loan *mpàkà* for ‘till that...’. If *shuti* in the meaning of ‘until’ has developed from the function ‘it must be that’ then both competitors for ‘until’ are lexical innovations, one internal and one borrowed. If languages that have a similar element to *mpaka* to express ‘until’ preceding a noun or a clause, and indicating a moment in time or place, then the motivation for borrowing from Swahili cannot be the efficiency that a bilingual speaker discovered in Swahili in preposition/conjunction-like element for ‘until’. It would simply reflect the dominance of Swahili.

Other East African languages have different grammatical means that have a wider application and are used for ‘until’. Sandawe uses the additive suffix *kí* in *swê-nà-kí* /now-to-additive/ ‘until now’ (Eaton 2010: 34) or it uses durative aspect in the verbal complex together with a time indicating constituent and ‘until’ is part of interpretation, e.g. ‘I keep waiting [until] the millet sprouts’ (Eaton 2010: 72 example 78).

The general pattern is that languages of East Africa borrow the preposition *mpaka* ‘until’ and do not copy the Swahili grammaticalization. In most of these speech communities Swahili is a strong second language and most speakers are bilingual in Swahili. For some languages the function word is an additive rather than a replaceive borrowing. For four languages in our files there is a competing preposed functional element meaning ‘until’ (Ik, Makua, Rangi, Shambala). The languages that use *mpaka* like a preposition often include time in the contexts of use. The data is not always rich enough to establish this but it looks as if time is part of the meaning of *mpaka* from the start. It is probable that it is precisely in time expressions that speakers copy Swahili patterns into the other language that they

---

15. Which is a transfer from Swahili since the regular reflex for ‘boundary’ is *mhaka* (LangHeinrich 1921: 269, 320)

speak. The use of a preposition 'until' has the advantage that a time period and its endpoint can be expressed in one constituent, for example *pàkâ ʔà làʔté* 'until they died' (Sandawe, Eaton 2010: 103). Speakers will use *mpaka* in the same functions in which they can use in Swahili and hence we can find *mpaka* with clausal complements. *Mpaka* is also used to indicate the end of an enumeration (X,Y up to Z) in Mbugu and possibly other languages too, as it is in Swahili.

## Acknowledgements

I would like to thank the audience of Bayreuther Swahili colloquium of 2016 and two anonymous reviewers for their extremely useful comments and suggestions.

## References

- Al-Jallad, A. 2017. The etymology of ḥattā. In *To the Madbar and Back Again, Studies in the Languages, Archaeology, and Cultures of Arabia Dedicated to Michael C.A. Macdonald* [Studies in Semitic Languages and Linguistics 92], L. Nehmé & A. Al-Jallad (eds), 338–345. Leiden: Brill.
- Alnet, A. 2009. The Clause Structure of the Shimaore Dialect of Comorian (Bantu). PhD dissertation, University of Illinois at Urbana-Champaign.
- Baraza la Kiswahili la Taifa (Kenya). 2015. *Kamusi kuu ya Kiswahili*. Nairobi: Longhorn.
- Besha, R. M. 1993. *A Classified Vocabulary of the Shambala Language with Outline Grammar*. Tokyo: ILCAA.
- Busse, J. 1960. *Die Sprache der Nyiha in Ostafrika* [Deutsche Akademie der Wissenschaften zu Berlin. Institut für Orientforschung, Veröffentlichung 41]. Berlin: Akademie Verlag.
- Cammenga, J. 2004. *Igikuria Phonology and Morphology: A Bantu Language of South-West Kenya and North-West Tanzania*. Cologne: Rüdiger Köppe.
- Devos, M. 2008. *A Grammar of Makwe (Palma; Mozambique)*. Munich: Lincom.
- Dimmendaal, G. 1983. *The Turkana Language*. Dordrecht: Foris.
- Dunham, M. 2005. *Éléments de description du langi: Langue bantu F 33 de Tanzanie: Phonologie – grammaire – lexicque*. Leuven: Peeters.
- Eaton, H. 2010. *A Sandawe Grammar*. SIL. <[http://www-01.sil.org/silepubs/Pubs/52718/52718\\_EatonH\\_Sandawe\\_Grammar.pdf](http://www-01.sil.org/silepubs/Pubs/52718/52718_EatonH_Sandawe_Grammar.pdf)> (12 October 2016).
- Edelsten, P. & Lijongwa, C. 2010. *A Grammatical Sketch of Chindamba, A Bantu Language (G52) of Tanzania* [East African Languages and Dialects 21]. Cologne: Rüdiger Köppe.
- Edema Atibakwa Baboya. 1994. *Dictionnaire bangála-français-lingála = Bagó na mɔnɔkɔ na bangála-frasé-lingála: Suivi d'un lexique lingála-bangála-français et d'un index français-bangála-lingála*. Paris: ACCT.
- Gibson, H. 2013. Auxiliary Placement in Rangi: A Dynamic Syntax Perspective. PhD dissertation, University of London, SOAS.
- Harjula, L. 2004. *The Ha Language of Tanzania: Grammar, Texts and Vocabulary*. Cologne: Rüdiger Köppe.
- Harris, L. P. 1950. *A Grammar of Mwera*. Johannesburg: Witwatersrand University Press.



- Heine, B., Güldeman, T., Kilian-Hatz, C., Lessau, D. A., Roberg, H., Schladt, M. & Stolz, T. 1993. Conceptual shift: A lexicon of grammaticalization processes in African languages. *Afrikanistische Arbeitspapiere* 34–35. Cologne: Institut für Afrikanistik.
- Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP.
- Heine, B. & Dunham, M. 2010. Grammaticalization in Bantu with special reference to Swahili. In *Bantu Languages, Analyses, Description and Theory*, K. Legère & C. Thornell (eds), 31–45. Cologne: Rüdiger Köppe.
- Hendle, J. 1907. *Die Sprache der Wapogoro (Deutsch-Ostafrika) nebst einem deutsch-chipogoro und chipogoro-deutschen Wörterbuche* [Archiv für das Studium deutscher Kolonialsprachen 6]. Berlin.
- HSC 2.0 = Helsinki Swahili Corpus 2.0, curated by Arvi Hurskainen. 2016. <<http://metashare.csc.fi/repository/browse/helsinki-corpus-of-swahili-20-hcs-20/d544fe38184411e291cf-005056be118e78589842c66b4efda6ad46f4034679c9/>> (27 April 2020).
- Johnson, F. 1939. *A Standard English-Swahili Dictionary*. Oxford: OUP.
- Kagaya, R. 1989. *A Classified Vocabulary of the Pare Language*. Tokyo: ILCAA.
- Kagaya, R. 2005. *Jita Vocabulary*. Tokyo: ILCAA.
- Kagaya, R. 2006. *A Kiwošo Vocabulary*. Tokyo: ILCAA.
- Kapinga, M. C. 1983. *Sarufi maumbo ya Kiswahili sanifu*. Dar-es-Salaam: TUKI.
- Kaji, S. 2000. *A Haya Vocabulary*. Tokyo: ILCAA.
- Kisbey, W. H. 1906. *Zigula-English Dictionary*. London: Society for the Promotion of Christian Knowledge.
- Kossmann, M. 2005. *Berber Loanwords in Hausa*. Cologne: Rüdiger Köppe.
- Krifka, M. 1983. *Zur semantischen und pragmatischen Motivation syntaktischer Regularitäten: Eine Studie zur Wortstellung und Wortstellungsveränderung im Swahili*. [Studien zur Theoretischen Linguistik 5]. Paderborn: Wilhelm Fink.
- LangHeinrich, F. 1921. Schambala-Wörterbuch [Abhandlungen des Hamburgischen Kolonialinstituts, Reihe B., Völkerkunde, Kulturgeschichte und Sprachen 23]. Hamburg: Friederichsen.
- Leach, M. B. 2010. *Things Hold Together Foundations for a Systemic Treatment of Verbal and Nominal Tone in Plateau Shimagonde*. Utrecht: LOT.
- Maganga, C. & Schadeberg, T. C. 1992. *Kinyamwezi: Grammar, Texts, Vocabulary*. Cologne: Rüdiger Köppe.
- Marlo, M. R. 2007. The Verbal Tonology of Lunyala and Lumarachi: Two Dialects of Luluyia (Bantu, J.30, Kenya). PhD dissertation, University of Michigan, Ann Arbor.
- Mol, F. 1996. *Maasai Language & Culture Dictionary*. Lemek, Kenya: Maasai Centre.
- Morrison, M. 2011. A Reference Grammar of Bena. PhD dissertation, Rice University. <<https://scholarship.rice.edu/handle/1911/70359>> (27 April 2020).
- Mous, M. 1993. *A Grammar of Iraqw*. Hamburg: Helmut Buske.
- Mulira, E. M. K. & Ndawula, E. G. M. 1952. *A Luganda-English and English-Luganda Dictionary*. London: Society for Promoting Christian Knowledge.
- Muriungi, P. 2009. Phrasal Movement inside Bantu Verbs: Deriving Affix Scope and Order in Kĩitharaka. PhD dissertation, CASTL Universtiy of Tromsø. <<https://munin.uit.no/bitstream/handle/10037/1720/thesis.pdf?sequence=1>> (27 April 2020).
- Murphy, J. D. & Kiggundu, C. 1972. *Luganda-English Dictionary*. Washington DC: Consortium Press for Catholic University of America Press.
- Mutonyi, N. 2000. Aspects of Bukusu Morphology and Phonology. PhD dissertation, Ohio State University.
- Muzale, H. 1989. *A Reconstruction of the Proto-Rutara Tense/Aspect System*. PhD Memorial University of Newfoundland, St. John's.

- Ngonyani, D. 2003. *A Grammar of Chingoni*. Munich: Lincom.
- Ngunga, A. 2000. *Phonology and Morphology of the Ciyao Verb*. Stanford CA: CSLI.
- Nicolle, S. 2013. *A Grammar of Digo: A Bantu Language of Kenya and Tanzania*. Dallas TX: SIL.
- Nikiema, N. & Kinda, J. 1997. *Dictionnaire orthographique du moore*. Ouagadougou: Sous-commission Nationale du Moore.
- Nurse, D. 2000. *Inheritance, Contact, and Change in Two East African Languages*. Cologne: Rüdiger Köppe.
- Olson, H. S. 1964. The Phonology and Morphology of Rjimi. PhD dissertation, University of Hartford, Connecticut.
- Orkaydo, O. O. 2013. *A Grammar of Konso*. Utrecht: LOT.
- Payne, D. L. & Ole-Kotikash, L. 2008. *Maa Dictionary, Maasai (IlKeekonyokie, IlPurko, IlWua-sinkishu) and Samburu*. <<http://darkwing.uoregon.edu/~maasai/>> (12 October 2016).
- Petzell, M. 2008. *The Kagulu Language of Tanzania: Grammar, Texts and Vocabulary*. Cologne: Rüdiger Köppe.
- Petzell, M. 2015. Three Kagulu stories: Annotations, analysis, and word lists. *Studia Orientalia Electronica* 3. <<http://ojs.tsv.fi/index.php/StOrE/article/view/48196>> (27 April 2020).
- Philippson, G. & Nurse, D. 2000. Gweno, a little known Bantu language of Northern Tanzania. In *Lugha za Tanzania/Languages of Tanzania: Studies Dedicated to the Memory of Prof. Clement Maganga*, K. Kahigi, Y. Kihore & M. Mous (eds), 233–284. Leiden: CNWS.
- Pillinger, S. & Galboran, L. 1999. *A Rendille Dictionary Including a Grammatical Outline and an English-Rendille Index*. Cologne: Rüdiger Köppe.
- Roehl, K. 1911. *Versuch einer systematischen Grammatik der Schambala sprache*. [Abhandlungen des Hamburgischen Kolonialinstituts 2]. Hamburg: Friederichsen.
- Rugemalira, J. 1993. Runyambo Verb Extension and Constraints on Predicate Structure. PhD dissertation, University of California at Berkeley.
- Rugemalira, J. 2009. *Cigogo: Kamusi ya Kigogo-Kiswahili-Kiingereza, Kiingereza-Kigogo, Kiswahili-Kigogo=Gogo-Swahili-English, English-Gogo, Swahili-Gogo Dictionary*. Dar es Salaam: Mradi wa Lugha za Tanzania, Chuo Kikuu cha Dar es Salaam.
- Sacleux, C. 1939. *Dictionnaire Swahili – Français*. Paris: Institut d'Ethnologie.
- Schadeberg, T. C. 2000. *Ekoti: The Maka or Swahili Language of Angoche*. Cologne: Rüdiger Köppe.
- Schrock, T. 2014. *A Grammar of Ik (Icé-tód) Northeast Uganda's Last Thriving Kuliak Language*. Utrecht: LOT.
- Stegen, O. 2011. In Quest of a Vernacular Writing Style for the Rangi of Tanzania: Assumptions, Processes, Challenges. PhD dissertation, University of Edinburgh.
- Stirnimann, H. 1983. *Praktische Grammatik der Pangwa-Sprache (SW-Tansania)*. Freiburg, Schweiz: Universitätsverlag.
- Ström, E.-M. 2013. The Ndengeleko Language of Tanzania. PhD dissertation, Gothenburg Univ. <<https://gupea.ub.gu.se/handle/2077/32111>> (27 April 2020).
- Stroomer, H. 1995. *A Grammar of Boraana Oromo (Kenya): Phonology, Morphology, Vocabulary*. Cologne: Rüdiger Köppe.
- TUKI. 1981. *Kamusi ya Kiswahili sanifu*. Dar es Salaam: OUP and Taasisi ya Uchunguzi wa Kiswahili (TUKI).
- TUKI. 2001. *Kamusi ya Kiswahili – Kiingereza*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili (TUKI).
- van der Wal, J. 2009. *Word Order and Information Structure in Makhuwa-Enahara*. Utrecht: LOT.
- Yoneda, N. 2006. *Vocabulary of the Matengo Language*. Tokyo: ILC.AA.
- Yukawa, Y. 1989. *A Classified Vocabulary of the Nilamba Language*. Tokyo: ILC.AA.

## Appendix. Table of transfer of ‘until’ in East Africa

The table represents what I found in the languages of East Africa for which I had documentation. The second column gives the loan ‘until’ or a hyphen when there is no such loan. The next two columns indicate whether the complement can be place or time. This is mostly based on example sentence provided or sentences found in texts. Hence in many instances, information is lacking. I have not found a language that uses borrowed ‘until’ for only one of the two meanings. The next column indicates whether the preposition is attested with a clausal complement. This is only explicitly excluded for Makua. For many languages the evidence is inconclusive due to lack of information. The next column present alternatives to the preposition ‘until’ in the languages. The last but one column present the word for ‘boundary’ and whether is different (D), (3/4) indicates the noun class; sometimes the plural of the noun is given after |. The last column indicates the source for the data.

Language	‘until’	Place	Time	_S	Other	‘boundary’	Source
<i>Non-Bantu languages</i>							
Ik	páka	Y		Y	gone P	D: akáni’	Schrock (2014: 250f, 612)
Borana	haga	Y		Y		D	Stroomer (1995)
Konso	haka + GEN*	Y				D	Orkaydo (2013: 180f)
Sandawe	páká		Y	Y	Y		Eaton (2010)
Rendille	–				geeddi C, deeddi C		Pillinger and Galboran (1999)
Maasai	–				ɔ-	ol-pólósíé (and other words)	Payne (2008)
Turkana	–				táni`		Dimmendaal (1983: 481)
Iraqw	–				ay, aydíř +S, afiqoomár < mouth period:of P		Mous (1993: 55, 101)
<i>Borrowed as preposition because different word for ‘boundary’</i>							
Nyamwezi	mpáká					D: loříimbí; ma-lu-	Maganga & Schadeberg (1992)
Luganda	paka					D: èsalè, èmbobò, èkingi,	Murphy (1972), Mulira & Ndawula (1952)
Ekoti	mpakha	Y	Y	?	–	D: mwiinano   nyiinano (3/4)	Schadeberg (2000: 194)

Language	'until'	Place	Time	_S	Other	'boundary'	Source
Kivunjo	mpaka	?	Y	?	?	D: ndasá   mi-ndasá	Kagaya (2006)
Chaga							
Nilamba	mpáka					mím̄bí	Yukawa (1989)
Haya	mpáka					olubibi   embibi (11/10)	Kaji (2000)
Makonde	mpáaka					?	Leach (2010: 155)
Makua	mhpaka	Y	Y	N	hata	?	Van der Wal (2009)
Ndengeleko	mpaka					?	Ström (2013)
Kagulu	mbaka		Y	Y		?	Petzell (2008, 2015: 60)
Bena	mpagha	Y	?	?	?	D: mbeto, lugengetsi**	Morrison (2011)
Shimaore	paka					?	Alnet (2009)
Mbugwe	mpaka					mupáka (3/4)	Mous (2004: 22)
Rangi	mpaka	Y	Y	Y	f̄oʊʊ Adv	m̄o-haka (3/4)	Gibson (2013: 50,84) Dunham (2005: 83,141,192)
Mbugu	mpaka	Y	Y			mhpaka	Mous (2003)
Ndamba	mbaka	Y		Y		mupaka	Edelsten and Lijongwa (2010: 131)
Zigula	mpaka					mhaka   mi- (3/4)	Kisbey (1906)
Shambala	mpaka				shuti	mhaka   mi- (3/4)	Besha (1993)
Gweno	mpaká	Y	Y			mw.aká 3/4	Philippson & Nurse (2000)
Pare	mháká				?	m̄haka	Kagaya (1989)
Gikũyũ	mpaka	?	?	?	?	mũhaka	Samuel Mdogo p.c.
Kisii	mbaka	?	?	?	?	oroḅago	Geofred Osoro p.c.
Matengo	mpaka	?	?	?	?	m̄paka	Yoneda (2006)
Makwe	`mpáaka	?	?	?	?	`mpáaka	Devos (2008)
Ilwana	mupaka	?	?	?	?	mupaka	Nurse (2000)
Pogoro	mpaka	?	?	?	?	mpaka	Hendle (1907)

\* The genitive consists of the vowel *a* which merges with the final vowel of *haka*; hence it is an issue of analysis whether this genitive is present or not.

\*\* Angelus Mnenuka p.c. Bayreuth colloquium. The first noun is most common.



PART 3

## Mixed languages and language mixing



# Turkish-German code-switching patterns revisited

## What naturalistic data can(not) tell us

Jeanine Treffers-Daller

University of Reading

This paper offers a review of what we know about Turkish-German code-switching patterns as found in naturalistic conversations, based on the typology of code-switching offered by Muysken (2000, 2013). Providing information about code-switching patterns that are commonly found in naturalistic data is important as it can be used to inform research into language switching, that is, externally induced switching between languages in experimental settings (Gullberg, Indefrey & Muysken 2009). Such experimental evidence on language switching is needed to supplement the evidence that is available from naturalistic data, but researchers using experimental approaches are not always aware of the range of intrasentential code-switching patterns which occur naturalistically, nor of the ways in which this variability has been captured in Muysken's typology, although this typology is highly relevant for theories of processing and cognitive control in bilinguals (Treffers-Daller 2009; Green & Wei 2014; Hofweber, Marinis & Treffers-Daller 2016, 2019). This paper aims to bridge the gap between researchers working on naturalistic code-switching and those working on language switching by summarizing what naturalistic data can tell us and which questions can only be answered with experimental approaches.

**Keywords:** code-switching, language switching, Turkish, German, bilingualism

### 1. Introduction: Explaining variability in code-switching patterns

Code-switching has always been a core interest of Pieter Muysken, probably because the use of two languages within one utterance, as in (1), bears testimony to the creativity of bilinguals who can effortlessly produce utterances which draw upon two grammatical systems without interrupting the flow of speech.



- (1) J'étais au balcon *op mijn gemakske* en train de regarder les étoiles.  
 'I was sitting on the balcony at my ease, watching the stars.'

(Pierre, Dutch/French, Treffers-Daller 1994: 131)

Apart from syntactic reasons, there are also sociolinguistic reasons why Muysken and many of his PhD students find code-switching such an interesting research topic: The mixture of words and structures from two languages reveals the rebellious nature of bilinguals who defiantly transgress the monolingual norms imposed upon them by those who can speak only one language.

The Dutch/French bilingual who produced (1) was one of the most prolific code-switchers in my Brussels data set in which I studied a group of mainly elderly bilinguals who spoke the local varieties of French and Dutch (Treffers-Daller 1994). Pierre was not his real name of course. When choosing nick names for my informants, I decided to use the first names of members of staff in the Department for General Linguistics at the University of Amsterdam. It is not hard to guess whose name I borrowed for this informant. Pierre was possibly such a talented code-switcher because he was illiterate and therefore did not feel the normative pressures to use either Dutch or French in one sentence to the same extent as other Brusselers who could read and write. His wife, who pretended not to speak any Dutch, certainly tried to discourage him from switching between the local varieties but this only encouraged him further. A true Brussels *ketje* 'kid' if ever there was one.

Negative views of code-switching were prevalent even among prominent linguists until at least the 1950s, but this began to change with Pfaff (1979) and Poplack (1980) who flagged up the creative potential of this behaviour as well as the fact that there were patterns and rules behind the seemingly random mixture of words and grammar rules from different languages.

Through these first publications on code-switching it became clear that bilinguals' switching behaviour was not random and that certain switches were more common or plausible than others. This triggered the search for universal constraints on code-switching. While it is not possible to discuss the history of the field in any detail here, a few key approaches will be mentioned here. The most widely cited general constraint is the Equivalence Constraint (Poplack 1980), which predicts that switches occur only at points where the surface structures of the languages are parallel. Myers-Scotton & Jake's (2000, 2017) 4-M model is based on an analysis of the roles of four different morpheme types in situations of language contact. The most parsimonious theory is offered by MacSwan (1999: 2017) who advocates a constraint-free approach and suggests that nothing constrains code-switching apart from the requirements of the mixed grammars.

A major contribution to this debate was made by DiSciullo, Muysken and Singh (1986) who made a link between the patterns found in code-switching and

principles of generative grammar, in particular the government constraint, which predicts that elements which are in a government relation must remain in the same language. This meant that switching between a Moroccan verb and its Dutch complement *intercultureel werk* ‘intercultural work’ as in (2) would be “out.”

- (2) Anaka-ndir *intercultureel werk*.

I I-am-doing *intercultureel werk*

(Moroccan-Arabic Dutch (Nortier 1990: 131))

In subsequent work Muysken (2000: 25) explains that the government constraint as formulated in 1986 is “clearly inadequate” because there are lots of counter examples to this constraint, for example in the work of Nortier (1990). Spurred on rather than deterred by examples such as (2), Muysken (2000) suggests that absolute constraints which can be invalidated by even a few counter examples are not the best approach to the analysis of bilingual speech production data.

Production data can be problematic for syntacticians because speakers often suddenly decide to abandon a structure they had originally started in favour of another structure. In (3), for example, the speaker starts off with a German main clause but changes the structure within the verb phrase by using a German Turkish mixed verbal compound *geben yaptılar* ‘give did,’ instead of the German past participle *gegeben* ‘given,’ which would have been expected at this point if the speaker had continued with the German structure. The Turkish glosses are based on Kornfilt (1997).

- (3) Damals haben die mir *biraz* Kartoffeln *geben yap-tı-lar*.

At that time have those me a few potatoes give do-PAST-PL.

‘At that time they gave me some potatoes.’

(Ayşe, in Treffers-Daller & Yalçın 1995, analysed in Muysken 2000)

Thus, in (3) there are two inflected verbs in the same utterance, *haben* ‘have’ in the first half of the utterance and *yaptılar* ‘they did’ in the second half, both of which are marked for tense, which is enough to give syntacticians of any persuasion a major headache. For switches such as those in (3), where the switched parts do not form a constituent together and/or there is double marking (doubling)<sup>1</sup> of some elements in the sentence, Muysken (1990) created the term *ragged* or *non-constituent switching*.

Karabağ (1995) found such ragged switches frequently among Turkish returnees who were no longer in everyday contact with German. In (4) the modal verb *muß* ‘must’ is in German. Instead of finishing the utterance with a German

1. ‘Doubling’ refers to examples where “the semantic value of the switch is the same as that of another morpheme in the original language also found in the utterance, as e.g. when plurality is marked twice” (Muysken, Deuchar & Wang 2007: 1317).

infinitive, the speaker finishes with the Turkish verb *gerek-* ‘must’ marked with the progressive aspectual marker *-iyor*. The subject agreement marker on the finite form is zero in the third person (Kornfilt 1997), which is why agreement is not visible here. In (4) doubling does not only occur in the modal verbs: the German preposition *mit* ‘with’ is also repeated in the Turkish suffix *-le* ‘with’.

- (4) Deutschland muß mit dies-en Hippie-ler-le baş-a  
 Germany must with these-DAT. hippie-PL.-INST. head-DAT.  
*çık-ma-sı gerek-iyor.*  
 leave-ANOM-3.SG must-Pr.Prog.-Ø  
 ‘Germany must cope with these hippies.’ (Karabağ 1995)

A similar example is found in the Turkish-Dutch data from Backus (1989), further analysed in Backus (1996) and Boeschoten (1991), here reproduced as (5).

- (5) Je moet naar een Türk ev-i-ne bak-acağ-m.  
 You must to a Turkish house-3.SG.-DAT look-FUT.-2.SG.  
 ‘You have to look at a Turkish house.’ (Backus 1989; Boeschoten 1991)

In this example the speaker uses a Dutch modal verb *moet* ‘must’, which is normally followed by a main verb in the infinitive (Dutch *kijken* ‘to watch’ in this case), but she abandons this structure and instead switches to Turkish. There is doubling in this utterance in that it contains an inflected Dutch modal verb and an inflected Turkish verb. The Dutch preposition *naar* ‘to’ and the dative case on *ev* ‘house’ also both indicate direction, which makes this example very similar to (4) where a German preposition *mit* ‘with’ was repeated in a Turkish instrumental case *-le* ‘with’.

The fact that universal grammatical constraints can be invalidated by a single counter example, as in (2), and the existence of extensive variability in patterns as well as ragged switches meant that the search for universal constraints which *only* take into account typological differences between languages was no longer a very fruitful avenue for research (cf. Deuchar, Muysken & Wang 2007). This does not mean that the search for unifying models of code-switching needs to be abandoned but rather that such models need to be refocused. Unifying models need to be more comprehensive and take into account a far wider range of factors than just morphosyntactic ones. Muysken, therefore, turned his attention to the development of a probabilistic approach aimed at explaining the variability in code-switching patterns by looking at the interplay between (morpho)syntactic patterns, theories of speech processing and sociolinguistic factors.

In a seminal paper Muysken (2013) proposes such a comprehensive framework for modelling and interpreting a wide range of language contact phenomena, which was developed on the basis of his typology of code-switching, but also includes

contact-induced language change and creole formation. It is this framework that is being used in this paper to analyse variability in code-switching patterns as found among Turkish-German bilinguals in Germany and Turkey. In a few cases, examples from Turkish-Dutch bilinguals are included when it is important to demonstrate that this is a more widely found pattern or when a suitable Turkish-German code-switch which illustrates a particular pattern is not available.

The key aim of this paper is to summarize what we know about code-switching between these language pairs so that this information can be used to inform research into *language switching*, that is, the externally induced switching between languages in experimental settings (Gullberg et al. 2009). This is important if research into code-switching is to have an impact on research in other areas of linguistics and psychology, for example research into the relationship between executive functions and bilingualism. As Costa, Hernández, Costa-Faidella and Sebastián-Gallés (2009) point out, it is possible that the monitoring needed to keep track of which language to speak to whom, and when to use code-switching, is at the heart of the advantages that bilinguals may have in executive functions. Investigating the link between code-switching and executive functions can thus help resolve the controversy over the conditions under which bilingualism enhances executive functions. However, progress in this area can only be made if we take into account the different types of code-switching bilinguals engage in (cf. Hofweber, Marinis & Treffers-Daller 2016, 2019). Code-switching is most definitely not monolithic behaviour and simplistic classifications of informants into a group who “does it” and a group who “doesn’t do it” are unlikely to advance our understanding of the ways in which code-switching relates to cognitive control.

In a recent special issue (henceforth SI) of *Linguistic Approaches to Bilingualism* on methodologies for the study of intrasentential code-switching, Badiola Delgado, Sande, and Stefanich (2018) note the importance of ensuring that experiments are based on naturalistic observations. However, in the papers in the SI little mention is made of the variability in patterns that are found in naturalistic data, and the authors in the SI do not appear to be aware of Muysken’s (2000, 2013) typology of code-switching, which has been developed on the basis of a study of a wide range of typologically and sociolinguistically different data sets. The relevance of the typology for experimental research appears to be poorly understood, although it is not difficult to see that acceptability judgements of stimuli will be influenced by what is (un)conventional in a specific sociolinguistic setting, and in a specific language pair. To give just two examples, it is helpful for researchers developing stimuli involving code-switched pronouns to know that insertions of single pronouns are extremely rare in naturalistic data (Muysken 2000) and researchers working with insertions of nouns or verbs would benefit from knowing that such patterns tend to be unidirectional (see Deuchar et al. 2007, and the discussion in 2.1). Any discussions about

whether or not it is more difficult to insert an L1 noun into an L2 sentence than vice versa will need to be informed by the literature on the sociolinguistic reasons for the directionality of code-switching in naturalistic data.

As pointed out by Valdés Kroff, Guzzardo Tamargo, and Dussias (2018) some researchers who are interested in language switching create experimental stimulus material containing switches which are very far removed from those found in naturalistic settings. Methodologies based on eliciting reactions from bilinguals to such artificial stimuli are unfortunately unlikely to provide meaningful insights into the practices of the bilinguals under study. Others obtain information about code-switching behaviour with a questionnaire which asks bilinguals to explain the types of code-switching they engage in. Soveri, Rodriguez-Fornells, and Laine (2011) invite respondents to indicate (dis-) agreement with twelve statements which tap the direction of code-switching (from language A to language B or vice versa), or the contexts in which they switch (“There are situations in which I always switch between languages”), and whether or not switching happens intentionally (“It is difficult for me to control the language switches I make during a conversation”). Such questionnaires do not capture the variety of patterns found in transcripts of naturalistic observations either. In an attempt to avoid these problems in their study of the relationship between code-switching behaviour and cognitive control, Hofweber et al. (2016, in press) took a different approach and elicited reactions from German-English bilinguals to a set of code-switches which were sampled from the available literature on German-English code-switching (see also Gullberg et al. 2009, for a range of techniques that can be used to elicit such reactions). This approach, which ensures that experimental stimuli used in language switching are not artificial but draw upon real examples of naturalistic code-switching, has more ecological validity and is therefore more likely to generate genuine insights into language switching and its relationship to other variables, such as measures of executive functions. Apart from the work of Goschler, Schroeder & Woerfel (accepted) so far few psycholinguistic studies have focused on Turkish-German bilinguals, although the Turkish community in Germany is the largest Turkish community outside Turkey (cf. Daller & Treffers-Daller 2014) and code-switching is widely practiced among this group (cf. Treffers-Daller 1997).

This paper sets out to give an overview of the kinds of code-switching that can be found among Turkish-German bilinguals, with the aim to provide information to other researchers who are interested in developing stimuli for use in studies of language switching. In the next section the four different types of code-switching distinguished in Muysken (2013) will be presented and in the final section of the paper the implications of the findings for further research in the field will be discussed.

## 2. The four types of code-switching

The four types of code-switching distinguished in Muysken (2013), which build on the three-way distinction developed in Muysken (2000), will be illustrated here with examples from Turkish-German and Turkish-Dutch code-switching as found in Backus (1989, 1996, 2001), Boeschoten (1991), Daller, Treffers-Daller & Furman (2011), Dođruöz & Backus (2009), Karabađ (1995), Treffers-Daller (1995, 1997, 2006), Treffers-Daller & Yalçın (1995), and Treffers-Daller, Daller, Furman & Rothman (2016). Further details on the diagnostic criteria for identifying the different types of code-switches can be found in Deuchar et al. (2007). In all examples, the Turkish part of the utterances is given in italics and the German or Dutch part in regular type face.

### 2.1 Insertions

Insertional code-switching involves the embedding of words or constituents from a guest language into a matrix language, which is often the speaker's first language. This is exemplified in (6), where the German *Flughafen* 'airport' is embedded into a Turkish clause and marked with a Turkish accusative case. The direct object *Flughafen* is selected by the main verb *buldum* 'I found,' and nested (Muysken 2000) in the Turkish construction.

- (6) *bütün Flughafen'ı bul-dum.*  
 entire airport-ACC. found-PAST-1.SG.  
 'I found the entire airport.' (Sedef, in Treffers-Daller 2006)

Example (6) was produced by Sedef, a 17-year-old second generation Turkish-German bilingual who was born in Germany and had lived there all her life. In the recording of her conversation, which was made by a Turkish-German research assistant (named Emre, a returnee) there are 22 insertions of German words in Turkish and only seven insertions of Turkish words in German. In total there are only 292 Turkish tokens but 1128 German tokens, which clearly illustrates the dominance of German in this conversation. Turkish is more prominent in the conversations Emre recorded in Turkey with Gizem, a 23-year-old Turkish-German returnee (267 Turkish tokens, 205 German tokens), as explained in Treffers-Daller (1997).

Insertional code-switching in the data from Sedef thus mainly consists of German insertions in Turkish and insertions in the opposite direction are much less frequent. The same strong preference for insertions mainly to occur in one direction was noted by Deuchar et al. (2007) for the Welsh-English data, where 96% of the insertions were English insertions in Welsh and by Blokzijl, Deuchar &

Parafita Couto (2017) for Spanish-English in Miami. In their data set there were more switches from Spanish determiners to English nouns than the reverse. The authors note that the switches tend to be in the direction of the official language, English. Similar asymmetrical patterns were found by Hofweber et al. (2016), who used a frequency judgement task<sup>2</sup> to study German-English code-switching among first generation immigrants in the UK and fifth generation heritage speakers in South Africa. In their study German insertions in English were judged to be much less frequent than English insertions into German.

It is interesting to take a closer look at a few utterances in which speakers do insert Turkish words into German. The ways in which this is done can depend on the speaker's proficiency in German. The speaker who produced (7) was Emre, who had grown up in Germany, while (8) was produced by his mother Ayşe who learned German as an adult but had been back for more than ten years at the time of recording, and did not speak German that often anymore at the time of recording.

- (7) Bist du mit dem *maaş* zufrieden?  
 Are you with the-DAT. salary content  
 'Are you happy with the salary?' (Emre)
- (8) Wenn man so schneidet oder mit Ø *çatal* drüber geht.  
 If one so cut or with Ø fork across goes  
 'If one cuts it this way or uses a fork to spread it out.' (Ayşe)

Emre integrates the Turkish noun *maaş* 'salary' into the German structure by allocating the neuter gender of the German translation equivalent *das Gehalt* 'the salary' to this noun and using it with an article in the dative case, namely *dem* 'the+DAT' that is normally used with the preposition *mit* 'with' in German. By contrast, Ayşe does not attempt to integrate *çatal* 'fork' in this way: *çatal* is a bare noun which is not accompanied by any determiners.

Turkish verbs can also sometimes be integrated into German, but as insertions from Turkish into German are rare overall, lone Turkish verbs are not often seen in German utterances. In (9) mixing takes place within the past participle: Sedef creates the stem *ayrıl-* by attaching a passive suffix *-il* to the Turkish root *ayr-* 'separate', after which German pre- and suffixes are attached to the stem *ayrıl-*:

---

2. This task was developed by Onar Valk (2014) who used it to measure the conventionality of Turkish grammatical patterns among Dutch-Turkish bilinguals. As code-switching is a stigmatised form of language behaviour, asking for grammaticality judgements is difficult as these will be influenced by attitudes towards code-switching (Badiola et al. 2018). In Hofweber et al. participants were instead asked to rate the frequency with which they would encounter utterances similar to the stimuli on a scale from 1=never to 7=all the time.

(9) Die sind *ge-ayr-il-t*.

They are PAST-separate-PASS.-PAST

'They are separate.'

(Sedef)

In his analysis of English-Afrikaans code-switching, Muysken (2011) refers to the 'masking' of morphological material from the guest language (in this case Turkish *ayrıl-*) as a *neutralization strategy*.

While the relative dominance of the languages under question is beyond doubt a key reason for the prevalence of insertions from German into Turkish, semantic factors can explain which kind of expressions are most likely to be used in insertions. Backus' (2001: 128) *specificity hypothesis*, given in (10), predicts that basic words such as *tree* are much less likely to occur as insertions than more specific items such as *oak*.

(10) *Specificity hypothesis*: Embedded language elements in codeswitching have a high degree of semantic specificity

This is true for example, for *intercultureel werk* in (2), which is a highly specific description of the kind of work done by the speaker. The rather basic word *werk* on its own would be much less likely to be used as an insertion than the expression *intercultureel werk*.

The specificity hypothesis can also explain why speakers sometimes produce insertions in the opposite direction, and use a Turkish expression in German, because the Turkish term has a high degree of specificity and finding a German translation equivalent is hardly possible. This is the case, for example, in (11), where Emre explains that friends of his decided to get married in a Turkish ceremony called *nikah*. Note that the Turkish word for the official marriage is preceded and followed by the German adverb *halt* 'just, you know', which functions as a flag (Poplack 1988), and potentially signals the problems the speaker has in trying to find a German expression which matches the concept of *nikah*. Failing that he decides to switch to Turkish.

(11) Jaja, und da haben sie sich halt. entschlossen zuerst

yes yes, and there have they themselves Interj. decided first

halt *nikah* zu machen.

Interj. marriage ceremony to make

'Yes, yes, and then they decided to do the marriage ceremony first, you know'

(Emre, in Treffers-Daller 2006)

As Examples (9) and (11) illustrate, it is possible to insert Turkish words into German. It is important to point out here that morphosyntactic considerations cannot account for the lack of popularity of insertions in this direction. Rather,



it is a matter of dominance relations between the two languages whereby the pre-dominant language (Silva-Corvalán 2014) is more likely to be the source of the insertions. In addition, it is semantic considerations which explain the prevalence of patterns favouring insertions of German words into Turkish discourse.

A very common strategy for integrating German or Dutch verbs into Turkish is the use of the helping verbs *yap-* 'do, make' or *et-* 'do.' Compound verbs with *yap-* are found more frequently among Turkish-Dutch/ Turkish-German bilinguals than compound verbs with *et-* (Doğruöz & Backus 2009; Muysken 2000; Pfaff 2000). As (12) and (13) from Sedef illustrate, both types of compound verbs occur in data from Turkish-German heritage speakers in Germany, although *yap-* is clearly the favourite among this group:

- (12) Çok *faszinieren* et-ti ben-i, bil-iyor mu-sun.  
 Very fascinate do-PAST I-ACC., know-Pr.Prog Q-2.SG.  
 'I was totally fascinated, you know.' (Sedef, in Treffers-Daller 2006)
- (13) Und *ben* feiern *yap-a-ma-dı-m*.  
 and I party do-ABIL.-NEG.-PAST-1.SG  
 'And I could not party.' (Sedef, in Treffers-Daller 2006)

Compound verbs with *yap-* also occur among returnees in Turkey, as shown in (14), so among returnees it is not necessarily the case that they only use compound verbs with *et-*, although constructions with *et-* are used more frequently by returnees (Treffers-Daller et al. 2016).

- (14) *Hemen* *H.bey* auch zustimmen *yap-tı* *filan*.  
 immediately H Sir also agree do-PAST Interj.  
 'Immediately Mr H also agreed.' (Gizem, in Treffers-Daller 2006)

## 2.2 Alternations

In the second type of code-switching, called alternation, fragments of different languages can be combined independently of the grammars involved, as exemplified in (15), where the apposition *hep böyle genç genç*<sup>3</sup> 'all so very young' accompanies the noun *Passagiere* 'passengers' and *yalnız* 'lonely' is an apposition to the subject *ich* 'I'. Neither of these appositions is embedded into the syntactic structure of the utterance, which was produced by Sedef, a second generation Turkish-German bilingual living in Germany:

---

3. As one reviewer has pointed out, reduplication is very common in Turkish, which could explain the repetition of *genç* in (15).

- (15) Es waren wirklich alles nur Passagiere, *hep böyle genç genç* ich kam  
 It were really all only passengers all this young young I came  
 mir auch ein bisschen doof vor, *yalnız*, weißt du.  
 me also a little stupid for lonely, know you  
 ‘There were really only passengers there, all so very young, I found myself a bit  
 funny, alone, you know.’ (Sedef, in Treffers-Daller 2006)

Sedef and Emre produce far more alternations (95) than insertions (29) in their conversation, and in the conversation between Gizem and Emre there are 20 insertions against 45 alternations. This confirms findings of Backus (1996) and Muysken (2000, 2013), who note that alternational code-switching is more common among second generation speakers. These have good competence in the societally dominant language, whilst first generation speakers, such as Ayşe, whose competence is more limited, produce more insertions, as exemplified in (8).

A different example of alternation can be found in the doubling of an impersonal construction, as in (16), where German *es gibt* ‘there is’ and Turkish *var* ‘there is’ are both used in the same sentence:

- (16) Batterie-*ler-de* *de şey var böyle, es gibt gümgümgüm*  
 battery-PL.-LOC. also thing there-is so there is dong-dong-dong  
 Anhänger *var*.  
 pendants there is  
 ‘There are such dong-dong-dong pendants with batteries.’  
 (Sedef, in Treffers-Daller 2006)

Another case of doubling can be found in (17) where the speaker uses a Turkish subordinating conjunction *ki* ‘whether’ as well as the German translation equivalent *ob* ‘whether.’ These two are used in the same utterance, which means the German subordinate clause is not truly embedded into the main clause, and the construction is likely to be an alternation rather than an insertion:

- (17) *On-dan sonra* *T ar-ıyor ki ob* wir nicht nach  
 that-Abl. after T call-Pr.Prog. whether whether we not to  
 K. könnten.  
 K could  
 ‘After that, T calls whether we couldn’t go to K.’  
 (Gizem, in Treffers-Daller 2006)

A third example of structures involving alternation in the data set is the occurrence of Turkish adverbial clauses such as *master yaptıktan sonra* ‘after doing a masters’ in (18) at the start of a German utterance. The adverbial clause is to a certain extent syntactically embedded into the utterance, as the adverbial clause occupies the first

position in the main clause and triggers verb second. We can see that because it is the inflected verb *kannst* 'you can' which follows the adverbial clause and not the subject *du* 'you'. However, the adverbial clause is at the periphery of the sentence, and not really nested into German in the way described in Muysken (2000). This is therefore still likely to be a case of alternation. The sentence was produced by a returnee who was in regular contact with German through his university studies:

- (18) *Master yap-tık-tan sonra, kannst du an*  
 Master do-FNOM-ABL. after can you at  
 der Universität *öğretim görevli-si* werden.  
 the university student clerk-3.SG. become  
 'After doing your masters, you can become a research assistant at the university.'  
 (Turkish-German returnee, Karabağ 1995)

Turkish adverbial clauses of different types can occupy the first position in a German utterance and trigger verb second, as in (19) where the adverbial clause *baloya gittiğimizizde* 'when we went to the ball' fulfils this role:

- (19) *On-dan sonra balo'ya git-tiğ-imiz-de sind* wir telefonieren  
 that-Abl. after balo-DAT. go-FNOM-1.PL-LOC. are we telephone  
 gegangen.  
 gone  
 'After that, when we went to the Ball, we went out to give a call.'  
 (Emre, in Treffers-Daller, 1997)

As the code-switch at the start of the sentence consists of two Turkish adverbial constituents in a row, the adverbial phrase *ondan sonra* 'after that' and the adverbial clause *baloya gittiğimizizde*, it is likely that this is an example of alternation rather than insertion, despite the fact that there is some integration into German syntax. In (19) the adverbial clause triggers Verb Second in the German main clause as we can see from the fact that *sind* 'they are' immediately follows the adverbial clause. As shown in Huls (1989, cited in Muysken 1990), simple Turkish adverbial phrases can also trigger Verb Second, so it seems that at least some bilingual speakers recognise different types of Turkish constituents as elements that can occupy the first position in a Dutch sentence.

A similar construction can be found in Demirçay and Backus (2014), whose data contain Example (20), which consists of a Turkish conditional clause which precedes a Dutch main clause, but this time the conditional clause is not embedded into the syntactic structure of the remainder of the sentence. The Dutch adverb *dan* 'then' takes the first position and triggers verb second in the main clause. The conditional clause is attached to the left periphery of the sentence, which indicates this construction is probably best analysed as alternation.

- (20) *Allah koru-sun bi al-mas-sa-k dan* moeten we die  
 God protect-IMP. we pass-NEG.-COND.-1.PL. then must we that  
 sowieso herkansen toch?  
 in any case resit right  
 ‘God forbid if we cannot pass it then we should take the resit anyway, right?’  
 (Demirçay & Backus 2014: 39)

It would be interesting to find out whether switching between a Turkish adverbial clause and a German main clause is typical of certain subgroups of Turkish-German or Turkish-Dutch bilinguals only, or a more wide-spread phenomenon. In addition, the occurrence of verb second in this type of constructions, and the link with language proficiency or language dominance, is worth investigating further. This will be taken up again in the discussion.

### 2.3 Congruent lexicalisation

The third strategy, congruent lexicalisation, sometimes called *dense code-switching* (Green & Wei 2014), is one where both languages share the same syntactic structure and elements can be drawn from each language. This is an unlikely scenario in Turkish-German language contact because of the typological distance between the two languages. However, a few examples in the data sets from Sedef and Gizem could be analysed as congruent lexicalisation, because it is not really possible to distinguish between a matrix language and a host language, partly because the utterances contain homophonous diamorphs (Clyne 2003), that is words which are phonetically similar in both languages, such as *August* and *ağustos* ‘August’ and proper names such as *Ali* which are the same in both languages and cannot therefore be attributed to one specific language. The switches are short, which makes alternation an unattractive option. Because the switches consist of prepositions or copulas, it is difficult to argue that these are switched because of their semantic specificity (Backus 2001). Therefore, it seems unlikely that these are insertions. Finally, there is syntactic convergence of Turkish to German in some of these utterances, which means there is evidence of some shared grammatical patterns which is typical of congruent lexicalisation.

In (21) and (22) the German prepositions *ohne* ‘without’ and *mit* ‘with’ occur inside a Turkish PP. Both of these sentences were produced by Ayşe, the Turkish-German returnee who had learned German as an adult after immigrating to Germany. In (21), the conjunction *ama* ‘but’ is a homophonous diamorph which closely resembles its German equivalent *aber* ‘but’, which is often pronounced as *ama* in colloquial German. It is possible that this homophonous diamorph functioned as a trigger for the switch to German for the preposition *ohne*:

- (21) *Ben biraz getir-di-m ama ohne et.*  
 I a bit bring-PAST-1.SG. but without meat  
 'I have brought a little bit but without meat.' (Ayşe, in Treffers-Daller 1995)

In (22) *mit* 'with' precedes its complement *pirinç* 'unboiled rice.' In a monolingual Turkish version of this sentence *with* would be expressed with the suffix *-le* 'with', so that the expression would be *pirinç-le* 'with rice'. While the internal structure of the PP is German, the overall word order of the utterance is Turkish. It is therefore an utterance in which both Turkish and German grammatical patterns co-occur, which is typical for congruent lexicalisation:

- (22) *Pilav mit pirinç yap-ıl-ır.*  
 (boiled) rice with (unboiled) rice make-PASS.-AOR.  
 'Boiled rice is made from unboiled rice.' (Ayşe, in Treffers-Daller 1995)

A similar compromise between Turkish and German word order can be found in (23), which was produced by Gizem:

- (23) *Üniversite-ye başvur-acağ-ım für asistanlık.*  
 university-DAT. apply-PL-1.SG. for assistant-ship  
 'I will apply to the university for an assistantship.'  
 (Gizem, in Treffers-Daller 2006)

In (23), *für* 'for' is followed by a homophonous diamorph *asistanlık* 'assistantship', the first half of which closely resembles the German *Assistenz* 'assistant.' The switch of *für* could be triggered by the following homophonous diamorph. Analysing *für* as an insertion is not an attractive option, because on its own it does not carry any specific semantic information (Backus 2001) that would explain why the speaker would want to insert it into a German structure. The fact that the entire PP *für asistanlık* is placed at the end of the utterance is interesting as many Turkish-German bilinguals put goal PPs after the verb instead of before it, also in utterances which do not contain code-switches (Daller et al. 2011). The canonical Turkish word order, shown in (24), is OV, although other options are possible depending on pragmatic conditions (Kornfilt 1997). The occurrence of both homophonous diamorphs and structural convergence make it plausible that (23) is an example of congruent lexicalisation.

- (24) *Üniversite-ye asistanlık için başvur-acağ-ım.*  
 University-DAT. assistant-ship for apply-FUT.-1.SG.  
 'I will apply for an assistantship at the university.'

Examples (25) and (26) involve switches of the German copula *ist* 'is' in an otherwise Turkish sentence:

- (25) *Ağustos ist iğrenç.*  
 August is disgusting  
 ‘August is disgusting.’ (Gizem, in Treffers-Daller 1997)
- (26) *Ali ist ayrı.*  
 Ali is separate  
 ‘Ali has separated.’ (Sedef, in Treffers-Daller 2006)

At first sight, the German copula *ist* ‘is’ in (25) and (26) looks like an insertion, as it is surrounded by Turkish words, but it is hard to claim that *ist* is nested in a Turkish structure as the word order is clearly German. In a monolingual Turkish version the Turkish epistemic copula *-Dir*<sup>4</sup> (Kornfilt 1997) would be affixed to the adjective (*iğrenç-tir; ayrı-dir*), while in German the copula precedes the adjective complement. These examples might look like insertions, but there are good reasons for not considering them as such. First of all, insertions tend to consist of content words (or multiword items) which are inserted into a matrix language because of the specificity of their lexical meaning. Alternatively, they could belong to a semantic domain that is typically encoded in the embedded language (Backus 2001). As copulas are function words, they are unlikely to carry any such specific meanings or belong to semantic domains that cannot be provided by Turkish. Second, we already know from Muysken’s (2011) analysis of Sinhala-English code-switching that insertion of single code-switched function words is highly restricted because of the lack of categorical equivalence between function words from different languages. Upon closer scrutiny it becomes clear that the subject is a homophonous diamorph in both utterances. This may again have triggered the switch to German and have paved the way for a compromise structure which is typical of congruent lexicalisation. An alternative analysis could therefore be that the first half of the utterance is actually German because the subjects *Ağustos* (*August* in German) in (25) and the proper name *Ali* in (27) are homophonous diamorphs which could be attributed to either language. If we assume that the first half of the utterance is indeed German, the adjectival complements *iğrenç* ‘disgusting’ and *ayrı* ‘separate’ can then be analysed as alternations to.

Switching of single function words is indeed very unusual in Turkish-German code-switching, but it is not completely ‘out’ as can be seen in (27), where the reported past (evidential) marker *-miş* is added to a German utterance to indicate

---

4. In the suffixes, capital letters indicate which sounds vary depending on consonant harmony or vowel harmony. Consonants in the suffix share voicing features with the last consonant in the root (*t* and *d* for  $\pm$  voice) and front and back vowels (*i* and *ı*) are also shared between root and suffix. In addition, the vowels *i* and *ı* in the suffix tend to become *ü* and *u* respectively when the root contains a rounded vowel. Thus, there are eight different allomorphs of the copula: *-tir, -tir, -tur, and -tür* and *-dir, -dir, -dur, and -dür* (Kornfilt 1997).

that the speaker is not sure whether the statement is true or has inferred it from some clues or evidence (Kornfilt 1997):

- (27) Er hat das Buch gelesen- *mış*.  
 He has the book read-*rep.PAST/INFER.PAST*.  
 ‘(I believe) he has read the book.’ (Turkish-German returnee in Istanbul)

A final example of congruent lexicalisation can be found in (28), where switching takes place within the relative clause which depends on *Tag* ‘day’. Because of the structural differences between Turkish and German with respect to relative clauses, one would not predict there to be any code-switching at this point, but the speaker who produced (28) was able to construe equivalence between Turkish and German relative clauses in such a way that code-switching became possible at this point. In this relative clause the word order looks German at the beginning, and the relative pronoun *wo* ‘where’ is German too, but the inflected verb *gid-ecek-ti-m* ‘I was going to’ is Turkish. This inflected verb would only appear in a main clause in Turkish as spoken in Turkey, as in a subordinate clause a non-finite, participial construction would be used (e.g. *gid-eceğ-im* go-fut-3sg). The goal PP *Klassenfahrt’a* ‘to the school trip’ occurs on the left of the main verb, which is common in both languages in subordinate clauses. Despite the important typological differences in relative clause formation between Turkish and German, the speaker constructs an utterance in which both grammars appear to operate in parallel and lexical items from both languages can be drawn in:

- (28) Und *ben* feiern *yap-a-ma-dı-m*, *çünkü* an dem Tag wo  
 and I party do-Abil.-NEG.-PAST-1.SG because on the-DAT. day where  
*Klassenfahrt’a gid-ecek-ti-m*, *akşam-a* konnt’ keine Fete  
 school trip-DAT. go-FUT.-PAST-1.SG evening-DAT. could no party  
 Machen. (Gizem, in Treffers-Daller 2006)  
 make  
 ‘And I could not go to the party because on the day that I was going to the  
 school trip I could not have a party until the evening.’

## 2.4 Backflagging

In the final strategy, called backflagging, the matrix language is the societally dominant language, and not the speaker’s heritage language. Bilinguals use elements from the heritage language to flag up their allegiance to their heritage language.

Turkish-German bilinguals in Germany frequently make use of backflagging, for example in using the calque *biliyormusun* ‘you know’ from German *weisst du* ‘you know’, as in (12), which is not commonly used in Turkish as spoken in Turkey.

Backflagging can also take the form of conjunctions and discourse markers from Turkish in otherwise completely German sentences, as in (29):

- (29) *Ama* nach Australien darfst du nichts Eßbares mitnehmen, *çünkü* es  
Aber to Australia dare you nothing eatable with take because it  
könnte ja Bakterien und so, *o yüzden*.  
could of course bacteria and so on that's why  
'But you can't take anything edible to Australia, *because* there could be bacteria  
in there and so on, *that's why*.' (Sedef, in Treffers-Daller 2006)

In (29), *ama* 'aber', *çünkü* 'because' and *o yüzden* 'that's why' do not contribute any specific content to the utterance that could not be expressed in German, but they do have the social function of flagging up the speaker's ethnic identity, even though the speaker may no longer use the heritage language as their main language. Similar examples can be found in Turkish-Dutch codeswitching, as in (30), where the only Turkish word is *da* 'also.' Again, the speaker could easily have used the Dutch translation equivalent *ook* 'also' and switching to Turkish does not add anything to the content of the utterance:

- (30) *Da's* niet te vertalen man! Woordenboek *da* weerdeloos.  
that's not to translate man! Dictionary also useless  
'That can't be translated, man! A dictionary is also useless.' (Boeschoten 1991)

### 3. Discussion and conclusion

In this paper an attempt has been made to provide an overview of different types of Turkish-German code-switching as found in spontaneous conversations. It is probably true that this overview is not complete, and that code-switching corpora based on chat fora such as that of Çetinoğlu (2016) contain types not yet covered in this overview, also because the medium (online written messages) is different from the recordings of spontaneous conversations from which examples in the current study were drawn. However, trying to put together an exhaustive list of all possible switch types in a language pair is not a realistic undertaking. As shown in Muysken (2013), what is possible in language contact does not just depend on structural factors, but on a great number of other factors, including language proficiency of speakers (and hearers), the frequency with which they use the languages, the practice they have in switching, the bilingual mode of the conversation, the attitudes towards mixing in the speech community, etc. Bilinguals using the same language pair can therefore be found to engage in different code-switching practices. Hofweber et al. (2016) show, for example, that fifth generation heritage speakers



of German in South Africa engage more often in congruent lexicalization than first generation German immigrants in the UK, as predicted by Muysken's model. The situation is very different in Brussels where code-switching is not commonly practiced because the codes symbolise social groups that are in conflict with each other (Treffers-Daller 1992). For researchers developing stimuli for experimental studies of code-switching or language switching it is essential to be aware of the variability in the code-switching patterns among the bilinguals in their study and of the factors that affect this variability.

The examples discussed in this chapter – in particular switching within a relative clause or switches involving compound verbs – also illustrate that code-switching is creative and innovative behavior. Highly skilled bilinguals can adjust structures from both languages to facilitate code-switching and thus produce structures that are not found in monolingual varieties (e.g. Turkish as spoken in Turkey). These innovations are therefore not easily accounted for in approaches which assume that only the requirements of the mixed grammars are needed to account for code-switching (MacSwan 1999).

The current chapter shows that all four types of code-switching distinguished by Muysken (2013) can be found in data from spontaneous conversations among Turkish-German bilinguals, although they are not all equally frequent. The frequency of the different patterns remains very difficult to establish as code-switching corpora tend to be very small. As pointed out by Muysken (2000, p. 29) not much information is currently available about the frequency distribution of specific grammatical patterns in monolingual data, which makes it very difficult to make strong claims about the frequency of such patterns in bilingual data. From the information that is available so far it appears that alternational code-switching is the most common form of code-switching among second generation bilinguals, whilst insertional code-switching seems more frequent among first generation bilinguals, as predicted by Backus (1996) and Muysken (2000). Examples of congruent lexicalisation can also be found but this is not the most common strategy among Turkish-German bilinguals because of the typological distance between the languages. By contrast, backflagging is clearly present among Turkish-German bilinguals, particularly so among those living in Germany, but perhaps less among returnees. The former appears to use this strategy as a way to flag up their allegiance to their heritage language, even when they do not use this language most often in everyday conversation.

While the current overview has therefore shown that there is great variety in code-switching patterns, it remains difficult to explain this variability. It is clear that typological differences between Turkish and German play a role, but individual differences with respect to migration status (first, second and third generation immigrants in Germany, or returnees in Turkey), and societal language

dominance are likely to be partly responsible too. However, how these variables affect code-switching behaviour cannot be determined in detail on the basis of the available evidence, which is based on spontaneous data samples. For these kinds of projects, informants are not screened for language proficiency, language learning history or other variables that might be relevant such as working memory or attitudes to the different languages and cultures. Experimental approaches where informants are systematically sampled and compared is needed in order to provide answers to acceptability of specific patterns by different groups of informants. However, some of the patterns could just be relatively arbitrary conventionalised routines used by a subgroup of speakers, and these are not necessarily the same in another subgroup (Muysken 1990; Valdés Kroff 2016). It is clear though that the variability found in the data does not bode well for universal constraints on code-switching, such as Myers-Scotton's 4-M model, as pointed out at the beginning of this chapter. Some speakers are able to switch between Turkish and German in highly unexpected places and produce, for example, switches within a relative clause, or in clauses containing copulas or evidential markers, where there does not appear to be any categorical equivalence between elements from the two languages at all. As the exceptional examples discussed here have shown clearly, absolute constraints which predict that certain forms of codeswitching are completely 'out' can easily be invalidated by single counter examples. A probabilistic approach which takes into account the interaction between structural, sociolinguistic, and processing factors, and keeps track of the effects of conventionalisation is more likely to be able to explain the patterns observed in the data.

An example of patterns which are due to conventionalisation can be found in the use of *yap-* versus *et-* in mixed compounds among Turkish-German heritage speakers in Germany who prefer *yap-* and returnees in Turkey who use more constructions with *et-* (Treffers-Daller et al. 2016). A more difficult issue is that of switches from Turkish to German after a Turkish adverbial clause. Some bilinguals apply the German Verb Second rule after the Turkish adverbial clause and thus treat the Turkish adverbial in the same way as any German constituent that occupies the first position in the utterance, while others add a German adverb to the Turkish adverbial clause before switching to German. In this paper, I have argued that in the latter case it is probably the German adverb which triggers Verb Second, and not the Turkish adverbial clause, which is attached to the left periphery of the utterance. Similar phenomena were found among French-Dutch bilinguals who tended to avoid putting some borrowed French adverbs in the first position of a sentence (Treffers-Daller 1994). It remains unclear why *some* borrowed or code-switched adverbial constituents are allowed to appear in the first position of a German or a Dutch sentence whilst others are not. The presence or absence of such adverbial constructions in this position could either be related to respondents' competence

in the different varieties, or to their frequency of use of these varieties, or to the bilingual mode the informants were in when recorded (Grosjean 1998), or to conventionalisation or to a combination of (some of) these variables, or to other variables not yet understood. Importantly, it is not really possible to explain the variability on the basis of evidence from observations only, that is transcriptions of naturalistic recordings of French-Dutch or Turkish-German bilinguals, because such corpora are generally small and thus unlikely to reveal patterns very clearly. In addition, such corpora do not provide evidence about code-switching types that respondents disprefer, nor about the reasons for such (dis)preferences. Of course, this could be investigated by exploiting the non-occurring switches (Muysken 1990) although the limited size of codeswitching corpora makes this difficult. Further sharing of corpus data between researchers, for example through MacWhinney's Talk Bank <<https://talkbank.org/>> would help to make it possible to compare between data sets (see also Toribio 2017).

To throw further light on these issues experimental approaches are needed which focus on *language switching*, that is a form of switching that is externally induced in a laboratory setting (Gullberg et al. 2009; Myers-Scotton 2006; MacSwan & McAlister 2015). In such approaches it is possible to manipulate utterances containing switches between an adverbial clause and a main clause, and to elicit reactions to these different manipulations, which could help explain why groups of informants engage in this type of code-switching whilst others do not or do so much less often. It could be the case, for example, that Turkish-German bilinguals in Germany who have a higher competence in German are more likely to accept Turkish adverbial clauses in the first position of a German sentence, whilst returnees in Turkey whose German has attrited are less tolerant of such constructions or even avoid them altogether. In the experimental manipulation a German adverb could be inserted into the first position of the sentence after the Turkish adverbial phrase in some utterances to see if this improves the acceptability of the utterance for some participants. It could also be the case that competence in either language is not the major factor here: the choice of particular structures could depend on where the speaker is on a bilingual mode continuum (Soares & Grosjean 1984). If speakers find themselves towards the bilingual endpoint of the continuum they might code-switch more freely than if they are in a more monolingual mode. The impact of the bilingual mode on the type of code-switching in which bilinguals engage has not been investigated in any depth but could be analysed in experimental approaches.

As pointed out by Badiola et al. (2018), we need converging evidence from a range of research techniques to deal with a phenomenon as complex as code-switching. Data from corpora and experimental data are complementary sources of evidence which have advantages and disadvantages but are both equally needed if progress

is to be made in this field of research. Of course, experimental studies should be based on the available evidence that has been obtained from naturalistic studies so that we can avoid eliciting responses to examples of code-switching that are very far removed from those occurring in spontaneous data and that thus have low ecological validity. I therefore hope that the overview presented in this paper will be useful to researchers wishing to embark on studies of language switching, and that we will continue to benefit from Muysken's inspirational work in the field of language contact for many years to come.

## Abbreviations

1.SG.	first person singular	FNOM.	factive nominal
1.PL.	first person plural	IMP.	imperative
2.SG.	second person singular	INST.	instrumental
2.PL.	second person plural	LOC.	locative
3.SG.	third person singular	NEG.	negation
3.PL.	third person plural	PASS.	passive
ABIL.	abilitative	PAST	past
ACC.	accusative	REP.PAST/	reported past/inferential past
ANOM.	Action nominal	INFER.PAST	
AOR.	aorist	PR.PROG.	present progressive
COND.	conditional	Q	interrogative particle
DAT.	dative	Ø	zero marking.
FUT.	future		

## References

- Backus, A. 1989. *Code-switching onder Turkse jongeren in Tilburg*. Tilburg: Katholieke Universiteit Brabant.
- Backus, A. 1996. *Two in One. Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: Tilburg University Press.
- Backus, A. 2001. The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In *Codeswitching Worldwide*, R. Jakobson (ed.), 125–154. Berlin: Mouton de Gruyter.
- Badiola, L., Delgado, R., Sande, A. & Stefanich, S. 2018. Code-switching attitudes and their effects on acceptability judgment tasks. *Linguistic Approaches to Bilingualism* 8: 5–24.
- Blokzijl, J., Deuchar, M. & Parafita Couto, M. C. 2017. Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages* 2(4). <https://doi.org/10.3390/languages2040020>
- Boeschoten, H. 1991. Asymmetrical code-switching in immigrant communities. Paper presented to the Workshop on Constraints, Conditions and Models, 85–100. London: European Science Foundation.

- Çetinoglu, Ö. 2016. A Turkish-German Code-Switching Corpus. Paper presented to the Proceedings of the Tenth International Conference on Language Resources and Evaluation, Paris, 2016.
- Clyne, M. G. 2003. *Dynamics of Language Contact: English and Immigrant Languages*. Cambridge: CUP.
- Costa, A., Hernández, M., Costa-Faidella, J. & Sebastián-Gallés, N. 2009. On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition* 113:135–149.
- Daller, M. H. & Treffers-Daller, J. 2014. Moving between languages: Turkish returnees from Germany. In *Rückkehr in die Fremde? Ethnische Remigration russlanddeutscher Spätaussiedler* [OstWestExpress. Kultur und Übersetzung 21], B. Menzel. & C. Engel (eds), 185–212. Berlin: Frank & Timme.
- Daller, M. H., Treffers-Daller, J. & Furman, R. 2011. Transfer of conceptualization patterns in bilinguals: The construal of motion events in Turkish and German. *Bilingualism, Language and Cognition* 14: 95–119.
- Demirçay, D. & Backus, A. 2014. Bilingual constructions: Reassessing the typology of code-switching. *Dutch Journal of Applied Linguistics* 3: 30–44.
- Deuchar, M., Muysken, P. C. & Wang, S.-L. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10: 298–340.
- DiSciullo, A.-M., Muysken, P. C. & Singh, R. 1986. Code-mixing and government. *Journal of Linguistics* 22:1–24.
- Doğruöz, A. S. & Backus, A. 2009. Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change. *Bilingualism, Language and Cognition* 12: 41–63.
- Goschler, J., Schroeder, C. & Woerfel, T. To appear. Convergence in the encoding of motion events in heritage Turkish in Germany: An acceptability study. In *Studies in Turkish as a Heritage Language*, F. Bayram (ed.). Amsterdam: John Benjamins.
- Green, D. W. & Li Wei. 2014. A control process model of code-switching. *Language, Cognition and Neuroscience* 29: 499–511.
- Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism, Language and Cognition* 1: 131–149.
- Gullberg, M., Indefrey, P. & Muysken, P. C. 2009. Research techniques for the study of code-switching. In *The Cambridge Handbook of Linguistic Code-switching*, B. Bullock & A. J. Toribio (eds), 1–17. Cambridge: CUP.
- Hofweber, J., Marinis, T. & Treffers-Daller, J. 2016. Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism* 6: 648–668.
- Hofweber, J., Marinis, T. & Treffers-Daller, J. 2019. Predicting executive functions in bilinguals using ecologically valid measures of code-switching behavior. In *Bilingualism, Executive Function and Beyond: Questions and Insights* [Studies in Bilingualism 57], I. A. Sekerina, L. Spradlin & V. Valian (eds). Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.57>
- Huls, E. 1989. *Interactienetwerken in Turkse gezinnen (Interaction networks in Turkish families)*. Ms, Tilburg University.
- Karabağ, İ. 1995. Türkisch-deutsche Sprachmischung bei Rückkehrern: Ein Arbeitsbericht. In *Zwischen den Sprachen: Sprachgebrauch, Sprachmischung und Sprachfähigkeiten türkischer Rückkehrer aus Deutschland*, J. Treffers-Daller & M. H. Daller (eds), 133–135. Istanbul: The Language Center, Boğaziçi University.
- Kornfilt, J. 1997. *Turkish*. London: Routledge.
- MacSwan, J. 1999. *A Minimalist Approach to Intrasentential Code-switching*. New York NY: Garland.

- MacSwan, J. & McAlister, K. 2015. Naturalistic and elicited data in grammatical studies of code-switching. *Studies in Hispanic and Lusophone Linguistics* 3: 521–532.
- Muysken, P. C. 1990. Concepts, methodology and data in language contact research. In *Papers for the Workshop on Concepts, Methodology and Data*, 15–30. Basel: European Science Foundation.
- Muysken, P. C. 2000. *Bilingual Speech: A Typology of Code-mixing*. Cambridge: CUP.
- Muysken, P. C. 2011. Bridges over troubled waters: Theoretical linguistics and multilingualism research. *Stellenbosch Papers in Linguistics* 40: 20–38.
- Muysken, P. C. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism, Language and Cognition* 16: 709–730.
- Muysken, P., Deuchar, M. & Wang, S.-L. 2008. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10: 298–340.
- Myers-Scotton, C. 2006. Natural codeswitching knocks on the laboratory door. *Bilingualism, Language and Cognition* 9: 203.
- Myers-Scotton, C. & Jake, J. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second language acquisition. *Linguistics* 38: 1053–1100.
- Myers-Scotton, C. & Jake, J. 2017. Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism* 21: 340–366.
- Nortier, J. W. R. 1990. *Dutch-Moroccan Arabic Code Switching among Moroccans in the Netherlands*. Dordrecht: Foris.
- Onar Valk, P. 2014. Convergent developments on Dutch Turkish word order: A comparative study between ‘production’ and ‘comprehension’ data: Converging evidence? *Applied Linguistics Review* 5: 351–372.
- Pfaff, C. W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55: 291–318.
- Pfaff, C. W. 2000. Development and use of et- and yap- by Turkish/German bilingual children. In *Studies on Turkish and Turkish Languages*, A. Göksel & C. Kerslake (eds), 365–373. Wiesbaden: Harrassowitz.
- Poplack, S. 1980. Sometimes I’ll start a sentence in Spanish y termino en español. Toward a typology of code-switching. *Linguistics* 18: 581–618.
- Poplack, S. 1988. Contrasting patterns of code-switching in two communities. In *Aspects of Multilingualism*, E. Wande, J. Anward, B. Nordberg, L. Steensland & M. Thelander, 51–77. Uppsala: Borgström.
- Silva-Corvalán, C. 2014. *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: CUP.
- Soares, C. & Grosjean, F. 1984. Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition* 12: 380–386.
- Soveri, A., Rodriguez-Fornells, A. & Laine, M. 2011. Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology* 2: 1–7.
- Toribio, A. J. 2017. Structural approaches to code-switching: Research then and now. In *Romance languages and linguistic theory 12: Selected papers from the 45th Linguistic Symposium on Romance Languages (LSRL) Campinas, Brazil* [Romance Languages and Linguistic Theory 12], R. E. V. Lopes, J. Ornelas de Avelar, & S. M. L. Cyrino (eds), 213–234. Amsterdam: John Benjamins. <https://doi.org/10.1075/rllt.12>

- Treffers-Daller, J. 1992. French-Dutch codeswitching in Brussels: Social factors explaining its disappearance, *Journal of Multilingual and Multicultural Development* 13: 143–156.
- Treffers-Daller, J. 1994. *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: Walter de Gruyter.
- Treffers-Daller, J. 1995. Deutsch-türkische Sprachmischung. Eine Einführung. In *Zwischen den Sprachen. Sprachgebrauch, Sprachmischung und Sprachfähigkeiten türkischer Rückkehrer aus Deutschland*, J. Treffers-Daller & M. H. Daller (eds), 69–83. Istanbul: Language Center, Boğaziçi University.
- Treffers-Daller, J. 1997. Variability in codeswitching styles: Turkish-German code-switching patterns. In *Code-switching Worldwide*, R. Jacobson (ed.), 277–297. Berlin: Mouton de Gruyter.
- Treffers-Daller, J. 2006. Accommodation in Turkish-German code-switching. Paper presented at the Workshop on Linguistic and Psycholinguistic Approaches to Code-switching and Language Switching, Nijmegen.
- Treffers-Daller, J. 2009. Code-switching and transfer: An exploration of similarities and differences. In *The Cambridge Handbook of Linguistic Codeswitching*, B. E. Bullock, & A. J. Toribio (eds), 58–74. Cambridge: CUP.
- Treffers-Daller, J., Daller, M. H., Furman, R. & Rothman, J. 2016. Ultimate attainment in the use of collocations among heritage speakers of Turkish in Germany and Turkish-German returnees. *Bilingualism, Language and Cognition* 19: 504–519.
- Treffers-Daller, J. & Yalçın, K. 1995. Variabilität türkisch deutscher Sprachmischung. Paper presented at the Germanistensymposium Eskişehir.
- Valdés Kroff, J. 2016. Mixed NPs in Spanish-English bilingual speech. Using a corpus-based approach to inform models of sentence processing. In *Spanish-English Codeswitching in the Caribbean and the US* [Issues in Hispanic and Lusophone Linguistics 11], R. E. Guzzardo Tamargo, C. M. Mazak & M. C. Parafita Couto (eds), 281–300. Amsterdam: John Benjamins. <https://doi.org/10.1075/ihll.11>
- Valdés Kroff, J., Guzzardo Tamargo, R. E. & Dussias, P. E. 2018. Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading codeswitches. *Linguistic Approaches to Bilingualism* 8: 98–133.

# Mixing and semantic transparency in the genesis of Yilan Japanese

Luis Miguel Rojas-Berscia  
University of Queensland

In this short piece, the so-called ‘Yilan Creole’ spoken in Yilan County, Taiwan (Republic of China) is reanalysed in light of contemporary discussions on mixing and the transparency principle as factors behind the genesis of languages. I argue that there is evidence to conceive the emergence of this vernacular as a case of Atayal+Japanese language mixing (rather than creolisation), whereby a large amount of lexicon, e.g. nouns, verbs, and particles, have been retained maintaining not only their original phonological form, but operating in the original Atayal way, within an all-encompassing Japanese grammar skeleton. In addition, a discussion on the transparency principle as a key factor in the genesis of this mixed language is brought into focus.

**Keywords:** mixed language, semantic transparency, Yilan Creole, Japanese, Atayal, Formosan languages

## 1. Introduction

The work of Pieter Muysken has been characterised as an overarching quest for universals, without ignoring the rampaging typological diversity of the world languages. His work on established mixed languages such as Media Lengua (Muysken 1981), pidginised varieties of Quechua, such as the Quechua vernacular of Lowland Ecuador (Muysken 2000), and his recent claims on the possibility of widespread language mixing in South America (O’Connor & Muysken 2014) are a modern point of departure in typological and formally oriented approaches to language (cf. Muysken 2008 for an all-encompassing volume of his ideas).

This brief article deals with two topics that have been of interest to Pieter Muysken, i.e. language mixing and the principle of semantic transparency (PST), as originally conceived in Seuren & Wekker (1986). The workings of these two processes are illustrated with examples from Yilan Japanese, originally dubbed



Yilan Creole by Chien and Sanada (Chien & Sanada 2010; Chien 2015, 2016). As such, the article has two orientations. On the one hand, based on the evidence presented by Chien & Sanada and a short two-week fieldtrip to Hanhsi and Chinyang villages in Yilan, Taiwan, I claim that it is preferable to classify this vernacular as a mixed-language, rather than as a creole, on both sociohistorical and formal grounds.<sup>1</sup> Sociohistorical as well as linguistic evidence is discussed in § 3. On the other hand, in § 4, I evaluate the genesis of Yilan Japanese in light of the three dimensions of semantic transparency: *uniformity*, *universality*, and *simplicity* (Seuren & Wekker 1986: 64), as factors not only present in creole genesis, but also shared across language genesis scenarios, as also observed in Muysken (2000) for Lowland Ecuadorian Quechua.

## 2. The Yilan “creole” of Taiwan

Yilan Creole, as dubbed by Chien and Sanada (2010), is the native language of some 3,500 people in Yilan County, Taiwan (Republic of China) (see Map 1). To date, it is spoken in four villages: Tungyueh, Chinyang, Aohua, and Hanhsi. The name ‘Yilan creole’ is not used in any of the villages in question, nor do villagers know about the existence of a creole in their communities. In Hanhsi, for example, they refer to the language as *Kangke no ke* (ㄎㄤㄎㄝ ㄋㄛ ㄎㄝ).<sup>2</sup> *Kangke* literally means ‘Hanhsi’, *no* is the Japanese genitive marker, and *ke* is the Atayal for ‘word’. In Chinyang, they refer to the language as *Nihongo* (ㄋㄧ ㄞㄥ ㄇㄛ), which is the Japanese word for ‘Japanese language’. People identified as speakers of Yilan Creole consider themselves members of the Atayal ethnic group. Many of them speak Atayal. However, the younger the speaker, the less likely he/she will have some command of the

---

1. The data for this article was collected in Hanhsi and Chinyang in August 2018. The corpus mainly consists of elicited sentences to test some of the hypotheses presented in Chien & Sanada (2010). The language used for elicitation was Mandarin Chinese. My main consultant in Chinyang was Mr. Toli (m, 70s). In Hanhsi, I worked with Wu LingMei (f, 50s), Zhang YingFen (f, 40s), and Kaluo (Kanox Yikai) (m, 30s). However, it must be noted that most of the examples in this article come from my interactions with Kaluo and YingFen. It was easier to translate sentences with them, given their familiarity with foreigners and their degree of schooling. As such, the corpus is more Hanhsi-biased. This should be taken into account for future studies that take more varieties and cross-generational variation into account.

2. Throughout the article, Zhuyin Fuhao (注音符號) or Bopomofo (ㄅ ㄆ ㄇ ㄏ), the major transliteration system of Mandarin Chinese in Taiwan (ROC), will be used to transcribe Yilan Japanese. Since our consultants are not familiar with the Latin transliterations of their sentences, I thought this might be the easiest way to provide an accessible representation for them for future reference.

ancestral language. Elders in Chinyang can carry spontaneous conversations in Yilan Creole and vernacular Japanese. This is not the case for younger generations. Conversely, in Hanhsi, even people in their late twenties have a good command of the “creole”. When speakers were asked about their language, they would say they speak a sort of “mixed” Japanese which is not the “original” one, and which displays elements of Atayal and Chinese. They would also highlight their Atayal identity, and, in many cases, would emphasise that they have lost or are on the way of losing the ancestral language, Atayal.

The pioneering work of Chien and Sanada (2010) put Yilan Creole on the spot as the only Japanese-lexifier creole in the world. Below I present some of the main grammatical characteristics of the language.



**Map 1.** Villages in Yilan County where Yilan Japanese is still in use<sup>3</sup>

The phonological system of Yilan Creole is quite similar to that of Japanese. Below I present the list of consonant phonemes:

3. The Yilan within Taiwan small map was taken from Luuva – Own work, CC BY-SA 3.0, <<https://commons.wikimedia.org/w/index.php?curid=4461770>>

**Table 1.** Consonants (taken from Qiu 2015)

	Bilabial	Alveolar	Alveo-Palatal	Palatal	Velar	Glottal
Stops	p b	t d			k g	ʔ
Fricatives		s z	ɕ		x	h
Affricates		ʃs	tɕ dʒ			
Nasals	m	n			ŋ	
Liquids		r l				
Glides				j	w	

As correctly pointed in Qiu (2015: 32), the phonemes /l/, /x/, and /ʔ/ (see bolds) are of Atayal origin. In addition, the Japanese bilabial fricative /ɸ/ and the uvular nasal /ŋ/ are absent in Yilan Creole.

As for the vowel system, Yilan creole has five vowels:

**Table 2.** Vowel inventory

	Front	Central	Back
High	i		u
Mid	e	(ə)	o
Low		a	

In terms of vowels, Yilan Creole equals Japanese in number. However, it displays a rounded /u/, unlike the unrounded Japanese /u/. In addition, it sometimes displays a schwa, following the rules of Atayal for consonant clusters (Rau 1992: 22), and must therefore be considered a transition support, rather than a vocalic phoneme (pace Qiu 2015: 38).

The system of pronouns of the language, compared to that of Japanese, has been highly reduced. Standard Japanese has some 41 pronouns (cf. Kaiser, Ichikawa, Kobayashi, & Yamamoto 2013: 168), while Yilan Creole only displays 11 of these, some of which originate in Atayal. According to Chien (2015: 528), “it seems that the simplification of the Japanese system in Yilan creole is due to substrate influence, since Atayal does not have polite forms and gender distinctions”.

**Table 3.** Pronoun inventory

	Singular	Plural
1	washi/wa/wax/waha	watatachi
2	anta/su	antatachi
3	jibun/jin/hya	jibuntachi

The pronouns of Atayal origin are in bold. As can be seen, pronouns of Japanese origin can be pluralised by means of *-tachi* suffixation. Special attention should be paid to the first person singular pronouns. The four forms in the table are used interchangeably by speakers of Yilan Creole. The first two forms originate in vernacular Japanese. However, both *wax* and *waha* show a rephonologisation of the Japanese pronouns from an Atayal perspective.

As regards the lexicon, Chien and Sanada (2010: 354) claim that “Yilan creole draws roughly 70 percent of its vocabulary from Japanese, roughly 30 percent from Atayal”. Chien (2015) adds that, “the proportion of Atayal-derived words in the Yilan Creole basic vocabulary is 18.3%, and that of Japanese-derived words is 35.6%”. According to Chien, Japanese is the basis for the lexicon and, as such, should be called the lexifier. Atayal, being the ancestral language would be the substrate language, while Taiwanese Mandarin and Southern Min would be adstrate languages.

In terms of syntax, Yilan Creole is mainly SOV. The original nominative-accusative marking of Japanese is absent. The distinction between the syntactic subject and object is made by means of word order (1)–(2). Modern generations sometimes resort to SVO structures, (3) (Chien 2015: 514). Words of Atayal origin are in bold.

- (1) X Y Yム去Y イセ口々 口ーセ日X°  
 Wa asta chyomen miy-e-ru.  
 S O V  
 I tomorrow book see-POT-N.PAST  
 ‘I will read a book tomorrow.’
- (2) X Y厂 厂尤厂せ 了セ 了せ 口ウX°  
 Wax kangke no ke yu-bu.  
 S O V  
 I Hanxi GEN word speak-N.PAST  
 ‘I speak the language of Hanhsi (Yilan Creole).’
- (3) X Y 了セ 口Y口Y 一口Y 去Y 夕X 口Y口ー  
 Wa no mama ima **tapuy** **mami**.  
 S V O  
 I GEN mother now cook porridge  
 ‘My mother is now cooking porridge.’

I could confirm this characterisation of the language by Chien, Sanada and Qiu with my own data collected in Hanhsi and Chinyang in August 2018. However, what are the features that make this language a creole beyond a label? Chien and Sanada (2010) resort to sociohistorical and formal arguments that are explored in more detail in the following section.

### 3. Is Yilan Creole a creole?

What is or what is not a creole lends itself to a squishy answer: “we cannot identify creoles with any certainty” (Smith 1994: 332). However, one of the most solid definitions of the notion of creole comes from sociohistorical studies.

Chien and Sanada (2010) provide some sociohistorical background for their main claim, i.e. Yilan Creole being a creole. They position the emergence of the language in the late nineteenth century and the first half of the twentieth century (1895–1945), when Taiwan underwent a process of invasion and colonisation by Japan. The beginning of the Japanese occupation of Formosa was characterised by massive slaughter and relocation of Formosan aboriginals from the mountains to the plains. Although there was an “intention” regarding the education of aboriginals, this did not result in any realistic outcome until as late as 1916–18 (Tsurumi 1977: 231). The beginning of the Japanese educational programmes with relocated aboriginal groups was referred to as “taming”, and only later on did they become fully-fledged programmes. Tsurumi describes the system as follows:

Courses were not to be fixed: standards were to be low; the length of the school year was to be flexible. Five years later the administration outlines a plan of studies in the Japanese language, ethics and arithmetic for aborigines who were judged ready for instruction. The plan called for four years of elementary education modelled on the common school curricula but of a much lower standard.

(Tsurumi 1977: 232)

These newly improvised schools were meant to provide aboriginal children with formal Japanese schooling. One of the main shortcomings of the programmes conducted in these institutions was the lack of proper teaching personnel. Police officers, with no pedagogical background, were in many cases the ones who had the duty to teach Japanese in the aboriginal schools. Additionally, this was only one of the many obligations they had. Police officers had to supervise the teaching of children, livestock raising, medical aid, transportation, etc. According to Tsurumi (1977), the language taught by these police officers was conversational Japanese. There was simply neither enough time nor the proper background to teach the standard variety of the language.

It is in this scenario that Atayal and Seediq groups came together. Although Atayal and Seediq are sister languages within the Atayalic subfamily of Austronesian, their languages are mutually unintelligible. According to Chien and Sanada (2010), the two groups needed a *lingua franca* that would act as a tool for inter-group communication, as well as a way of reaching the Japanese dominant group. These authors claim that it simplified Japanese was the language the Atayal and Seediq

used and which later on became a creole. Now, the question that pops up is what the population number of each one of the groups was for this to happen.

Chien and Sanada (2010: 354ff.) mention that the Seediq were fewer than the Atayal, and “[a]ccording to [the] Institute of Ethnology, Taihoku Imperial University (1935), in 1931, the population of the Atayal Nan-ao group was 1692. Atayal made up 85.6% of the population, while Se[e]diq made up 14.3%.” If the Atayal were a vast majority, why would they want to switch to a simplified variety of Japanese to communicate with a minority group? It seems that the *lingua franca* explanation is not sustained by numbers.

Beyond considerations of what a creole is from universalist perspectives (cf. Muysken & Smith 1994 for a survey of these), one that does not fail to categorise the group of creole languages as a cluster as mentioned earlier is the one coming from a sociohistorical perspective (Arends 1994; Mufwene 2001). For scholars such as Mufwene, creole languages are an exclusive product of European colonisation of the rest of the world, “starting in the seventeenth century, typically on island or coastal colonies between the tropics, *in the contact settings of plantations* (Mufwene 2001: 10). Plantations were the setting for the coming about of so-called creole languages. These were demographically characterised by relocated labourers from many different regions which were constantly replaced, due to the high levels of exploitation that led to a high rate of early deaths, and where there was little to no proper instruction of the major European language. This socio-historical scenario triggered a recursive accommodation of forms and styles in the linguistic competence of the diverse inhabitants of the plantations. This process eventually resulted in the coming about of what we know as creole languages.

If we avoid the “European” feature of this type of language genesis and transport it to the site in question, i.e. Taiwan under Japanese control, the numbers do not add up. There was no such situation. The majority population was Atayal. Some Seediq were relocated to the newly created Atayal villages. Both groups maintained their languages and received some type of explicit education in vernacular Japanese. They were not slaves. They could have perfectly well maintained their vernaculars (which they did in most villages). It seems that the concept of creole does not fit for such a situation. In Section 4 a new proposal is sketched.

### 3.1 Creole-like grammatical structures in Yilan Creole

Chien and Sanada (2010: 355) claim that Yilan Creole displays a remarkable simplification of morphology and morphosyntax, noticeable in the systematic lack of inflectional morphemes both in nominal and adjectival sentences. According to

the authors, “tense in nominal sentences and adjectival sentences in Yilan Creole is marked by adverbs instead of inflectional morphemes as in Japanese. This can be seen as one kind of simplification in morphology and morphosyntax” (2010: 355), see (4) and (5), slightly adapted from Chien and Sanada:

- (4) Yilan Creole: Aci ga gako.  
 There FOC school  
 ‘There is a school.’  
 Japanese: Asoko wa gakkoo da  
 there FOC school COP  
 ‘There is a school.’
- (5) Yilan Creole: Aci lela ga gako.  
 There yesterday FOC school<sup>4</sup>  
 ‘There was a school.’  
 Japanese: Asoko wa gakkoo da-tta.  
 there FOC school COP-PAST  
 ‘There was a school.’
- (6) Yilan Creole: Washi ima kukuy.  
 I now itchy  
 ‘I feel itchy.’  
 Japanese: Watashi ima kayui  
 I now itchy  
 ‘I feel itchy now.’
- (7) Yilan Creole: Washi kino kukuy.  
 I yesterday itchy  
 ‘I felt itchy yesterday.’  
 Japanese: Watashi kinoo kayu-katta.  
 I yesterday itchy-PAST  
 ‘I felt itchy yesterday.’

I can confirm this from my own data, see (8). Elements from Atayal appear in bold.

- (8) ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ ㄉㄞ  
**Ladan** kochi **utux** gako a-ru.  
 Before here one school be-N.PAST  
 ‘Before there was a school here.’

However, one of my consultants suggested that (9) is also grammatical. This could be due to dialectal or generational differences.

---

4. Remarkably, Chien and Sanada (2010) never explain why this sentence does not mean ‘there was a school yesterday’. Would *lela* be used to convey a past tense meaning even when the event/state in the proposition occurred ‘a year ago’ or ‘ten years ago’? This question remains open.

- (9) カ ヲ カ マ コ ト コ イ ヌ ク ヌ フ ≪ ヲ コ ト ヲ カ ヲ  
**Ladan** kochi **utux** gako a-da.  
 Before here one school be-PAST  
 ‘Before there was a school here.’

Conversely, the authors also suggest that this phenomenon “appears to be due to influence from Atayal” (ibid: 355). In Atayal, adjectives indeed can sometimes act as the main predicate of a sentence. (10) illustrates this phenomenon, whereby the Atayal adjective ‘es ‘happy’ (Rau 1992: 24), is prefixed with the agent focus marker *m-* (Gorbunova 2017), which links it to the syntactic subject of the sentence. Interestingly too, this sentence is full of Atayal grammatical elements, such as the phasal polarity particle *la* (Gorbunova 2014) with the meaning ‘already’.

- (10) マ ク ヲ □ - セ ク ヲ カ ヲ □ ' - セ ム  
 Anta **m-ieta** =**la**<sup>5</sup> **m-es**.  
 2 AF-look PPP AF-happy  
 ‘You already look as if you were happy.’

It seems that Yilan Creole can also deploy an Atayal grammatical skeleton in adjectival and nominal sentences. As such, this would not be a case of simplification.

In addition, it must also be noted that the “remarkable” simplifications in morphology and syntax are not present in sentences that have a verb as their main predicate. Below I present examples that show that Yilan Creole employs inflectional morphemes, sometimes identical to Japanese, but also identical to Atayal.

- (11) ヲ □ セ フ コ - ム コ - ク ト コ マ カ ヲ ヌ ヲ コ -  
 Are **hoying**<sup>6</sup> kino kan-da wa-ni.  
 That dog yesterday bite-PAST 1-DAT  
 ‘That dog bit me.’

5. It must be noted that an interpretation of the phasal polarity particle =*la* here as the Japanese conditional suffix *-ra* would not render the same meaning. A reviewer suggested the phrase could be translated as ‘when [I] looked, you were happy’: This seems not to be what the speaker means.

In addition, as the same reviewer pointed out, it seems that, in this case, we could be dealing with the Japanese predicate *mi-ye-* ‘see-POT-’. I agree that this is possible. There are two possible scenarios: Either this is the Atayal verb *k-ita* or the Japanese verb *mi-*, both meaning ‘to see’. In the former case, a phonological restructuring based on Japanese would have taken place, adding [e] to the base. In the latter case, the Japanese base would be allowed to carry the Atayal phasal polarity particle =*la* (see reasons above). Although I am more inclined to the former, a detailed grammatical survey is necessary to solve this question.

6. Atayal elements in the examples appear in bold. The translations of these words were originally provided by our consultants and later on checked against the Atayal grammars of Gorbunova (2017) and Rau (1992).



- (12) ×Y 4-1-3-1 去 所 厂 工 巧 × 去-1-么 力 Y ム せ 去 Y  
 Wa jin-ni Taihoku **tiyao-dase-ta.**  
 I 3-DAT Taipei work-CAUS<sup>7</sup>-PAST  
 ‘I made him work in Taipei.’

In (11), what is considered a tense-marking adverb by Chien and Sanada, i.e. *kino*, is indeed referring to a precise moment in the past, ‘yesterday’.<sup>8</sup> However, the main tense reference is given in the verb, by means of a past tense inflection *-da*. Similarly, in (12) the past tense reading is provided by *-ta*, which follows a causative operator.

- (13) ×Y 去 工 ×Y 3 工 カ-カ- Y ム 工 工-去 せ 工 ×  
 Wa to wa no *titi*<sup>9</sup> asobi-te-ru.  
 I COM I GEN younger.brother play-PROG-N.PAST  
 ‘I am playing with my younger brother.’

Sentence (13) also illustrates inflectional morphology. In this case, the progressive aspect marker *-te* is used along the non-past marker *-ru*.

#### 4. Yilan Japanese as a mixed language

One of the main claims of this short article is that the vernacular spoken in Hanhsi, Chinyang, Tungyueh, and Aohua is not a creole, but a mixed language. I subscribe to the definition of mixed language provided in Muysken (2008: 211), whereby by mixed language we mean a language that shows an asymmetry in its origin, i.e. which shows a degree of *genetic ambiguity* (Thomason & Kaufman 1988). As such, a mixed language is “a more or less stable language, [...] with substantial parts of their grammar and/or their basic lexicon from specific historically different sources” (Muysken 2008: 211). The sources in this case are vernacular Japanese and Atayal.

I tentatively sketch the history of the coming about of Yilan Japanese into four main periods compatible with known historical facts and *in situ* observations:

---

7. Here, the causative *-dase* formally resembles the Japanese *-sase*. However, as one reviewer suggested, it could be the case that *-dase* originates in the Japanese root *das-* ‘put out’. This would explain the oddity of the [t] segment. Although this is plausible, I surmise this is just the Atayal Japanese *-sase*. A study that collects data from more speakers would solve this problem.

8. This is also the case in (7).

9. Elements from Mandarin Chinese are in italics in all examples.

1. *Introduction of vernacular Japanese and pidginisation.* The progressive learning and use of vernacular Japanese in colonial times could be considered a second-order index (Silverstein 2003) of higher status and adaptation to Japanese customs. This would lead to the learning of vernacular Japanese and the triggering of a scenario of *semi-shift* (Croft 2000), whereby the ancestry language was “left behind” (Atayal, Seediq) and there was a preference for the newly introduced language (Japanese). However, given the short period of occupation, and the lack of a properly supervised education in the standard, a complete shift to Standard Japanese never occurred. The use of vernacular Japanese, as such, became widespread, and newer strategies for its expression were implemented by speakers. I surmise this scenario to be a case of *pidginisation*, understood as a general process used by L2 learners, which characterises random and sporadic contacts, and which could be related to our innate linguistic competence (Todd 1980).
2. *Japanese-Atayal code-switching.* After the liberation of the island, and therewith a cut of cultural ties with Japan, the use of vernacular Japanese did not index a positive value anymore. Therefore, there was a need to recover traditional cultural customs. This was performed through code-switching. The *re-invention* of an aboriginal identity was indexed through the use of Atayal words and sentences in daily speech (cf. Thomason 2003 for a discussion of mixed languages as a means for indexing an ancestral group membership).

To date, a general dispreference for Japanese seems to be still prevalent in Atayal Japanese speaking villages. When I arrived in Hanhsi people were concerned as regards my intentions towards their language. They told me they did not speak Japanese and did not want to learn that language.

3. *Conventionalisation of code-switching.* Code-switching conventionalised and a new mixed-language emerged (cf. Myers-Scotton 1993, 2003). The generations that were born after liberation were not exposed to Japanese or Japanese education anymore, but did have input from a variety of Japanese displaying insertional and alternational code-switching with Atayal that eventually became their first language. As such, *language intertwining*, i.e. a process of creation of a new language that has lexical morphemes from one language and grammatical morphemes from another language (Bakker & Muysken 1994: 42), occurred.

Kaluo and Yingfen from Hanhsi, two of my main consultants, told me that this was a common phenomenon in their households as children. They said their parents spoke the two languages all the time and that it was in this particular context that their vernacular came to be. Lingmei from Hanhsi also confirmed this. She said her mother used to speak the two languages constantly. I indeed confirmed this when trying to elicit some Yilan Japanese sentences



- (15) X Y 3 ㄛ ㄇ ㄩ ㄇ ㄩ 去 ㄩ ㄉ ㄨ 一 去 ㄛ ㄇ X  
 Wa no mama **tapuy-te-ru**.<sup>13</sup>  
 1 GEN mother COOK-PROG-N.PAST  
 ‘My mother is cooking.’
- (16) ㄇ 去 ㄩ ㄇ 一 ㄛ 去 ㄩ ㄉ ㄩ ㄩ 去 ㄩ ㄇ ㄩ ㄨ ㄛ ㄉ ㄩ ㄩ 一 ㄛ  
 Anta **m-ieta** =**la** atama **belayo**.<sup>14</sup>  
 2 AF-see PPP intelligent very  
 ‘You seem to be very intelligent.’

Interestingly, in (16) the object NP *atama belayo* retains a traditional Modified-Modifier structure, as in Atayal. It seems that the nuclei of NPs retain the grammar of their source language. In (17) a Japanese noun *ningen* ‘person’ is used. Here, the modifier *kenmai* ‘many’ precedes the modified, as expected from a Japanese grammar point of view.<sup>15</sup>

- (17) 去 ㄇ ㄨ ㄉ ㄉ ㄨ ㄨ ㄇ ㄩ ㄉ ㄨ ㄨ ㄨ ㄨ ㄨ ㄨ  
 Tanux **kenmai** ningen.  
 outside many person  
 ‘There are many people outside.’

However, two consultants (Kaluo and Yingfen) said that the same structure was possible with the Atayal word *s’uli*, as in (12).

- (18) 去 ㄇ ㄨ ㄉ ㄉ ㄨ ㄨ ㄇ ㄩ ㄉ ㄨ ㄨ ㄨ ㄨ ㄨ ㄨ  
 Tanux **kenmai** **s’uli**.  
 outside many person  
 ‘There are many people outside.’

As Meakins suggests, “often mixing is more intricate, involving different layers of the lexeme or morpheme, such as combining the forms from one source language that have a function from another language” (Meakins 2013: 187). In the examples

13. This is an interesting case of relexification (Muysken 1981 for a detailed account on the phenomenon in *Media Lengua*). However, as shown in (3), it is still possible to use *tapuy* without Japanese morphology.

14. As one reviewer commented, it seems that *belayo* originates in the Atayal *balay*. However, as he/she also commented, it could be the case that this form originated in the Atayal *blaq* ‘good’. Although I prefer the first analysis, the second one is still possible and a study on the lexicon with bilingual Atayal/Yilan Japanese speakers should prove useful to answer that question.

15. However, as a reviewer pointed out, it could be the case that in (17) and (18) the Atayal *kenmai* is the main predicate, thus retaining the predicate-initial (VS) structure of Atayal. This issue needs further exploration.

above, this was quite clear. Some verbs undergo a Japanese structuring, while other an Atayal one.

Essentially, Yilan Japanese seems not to be a *symbiotic mixed language* (Smith 1994: 333), i.e. a type of mixed language, whereby the grammatical structure of one language, and a varying number of lexical items often from the original native language are combined. The grammars of Atayal and Japanese seem to be in a constant competition.<sup>16</sup> Certainly, more in-depth studies of the grammar, dialectology and geographical as well as generational variation of the language are necessary to provide a better picture of what is going on. It could be the case that the variation I report has to do with generational lects. This is just a new attempt to bring this vernacular into focus.

The next section deals with the creation of Yilan Japanese, in light of the Principle of Semantic Transparency.

## 5. Semantic transparency in Yilan Japanese

The creation of Yilan Japanese, as sketched in § 3, seems to have been characterised by an initial introduction of vernacular Japanese and a subsequent restructuring through conventionalised code-switching, triggering the emergence of a new vernacular, unintelligible both to Japanese and Atayal speakers. This scenario, although not identical in all details, seems to be comparatively similar to that of Lowland Ecuadorian Quechua (Muysken 2000), which appears to be the product of an initial expansion of the Quechua language in pre-Incaic times, followed by a gradual restructuring due to substratal influence (Muysken 2009). As such, given the similarity of both emergence scenarios, I consider it of theoretical relevance to examine to what extent more/less transparent the language became. I hypothesise that pidginisation strategies in period 1, would trigger features that made Yilan Creole more *semantically transparent*, while the conventionalisation of code-switching would have triggered the opposite. With “more” semantically transparent I mean that surface structures (SS) must be almost identical to shallow structures (ShS) as a result of an absence of language specific post-cyclic rules. This view is embedded within the point of view of Semantic Syntax (q.v. Seuren, 1972, 2018).

---

16. It must be added that this is not a case of code-switching, since most speakers of Yilan Japanese have little to no knowledge of Japanese. Even so, many Yilan Japanese speakers have no command of Atayal and think of their language as a variety of Atayal, until they get to meet Atayal speakers in major cities. To their surprise, the “real” Atayal is unintelligible.

The *transparency principle* must not be considered just a defining factor in creole genesis, but a characteristic of creole languages, possibly shared with other ‘emergence’ scenarios as well. As such, I suggest that the principle of semantic transparency (PST) is an important factor in the emergence of new languages in general, such as creoles or mixed languages and is, therefore, rightly considered ‘universalist’.

Below I present some characteristics that made Yilan Japanese a more/less semantically transparent language, based on the three following strategies: *uniformity, universality, and simplicity*.

### 5.1 Uniformity

The strategy known as *uniformity* states that in the emergence of new contact languages “one will expect few arbitrary distinctions, as with grammatical gender or conjugational idiosyncrasies, or with derivational processes in morphology” (Seuren & Wekker 1986: 65). This is precisely the case of Yilan Japanese with regard to pronouns. The complex system of pronouns of Japanese, which includes male/female distinction, together with a formality distinction in respect of the interlocutor (Kaiser et al. 2013, see Table 5.3), was completely reduced to a few tokens (Table 3). In addition, the case-marking system of the language was divested of the nominative-accusative marking, and most space-related oblique case-markers have fallen together as *ni*:

**Table 4.** Case marking of Yilan Japanese compared to Japanese, adapted from Chien (2016)

	Japanese	Yilan Japanese
Nominative	<i>ga</i>	$\emptyset$
Accusative	<i>o</i>	$\emptyset$
Dative	<i>ni</i>	<i>ni</i>
Instrumental	<i>de</i>	<i>de</i>
Locative	<i>de</i>	<i>ni</i>
Allative	<i>e</i>	<i>ni</i>
Ablative	<i>kara</i>	<i>kara</i>
Comitative	<i>to</i>	<i>to</i>
Genitive	<i>no</i>	<i>no</i>

Also, the optional cyclic rule of copula insertion (Seuren 2018: 111) of Japanese is nonexistent in Yilan Japanese. As such, predicates can stand on their own (cf. 19–20).

- (19) 彼女 は 誰 (ですか)? [Japanese]  
 Kanojo wa dare (desu ka)  
 She FOC who (be INT)  
 ‘Who is she?’ (Aoyama p.c.)  
 カ ヲ 誰 セ ヲ ー ヲ ヌ ヲ ㄱ ㄱ  
 Dare jibun ga?  
 who 3 INT  
 ‘Who is he/she?’
- (20) 彼 の 髪 は 長い (です) [Japanese]  
 Kare no kami wa nagai (desu)  
 He/she GEN hair FOC long be  
 ‘His/her hair is long.’ (Aoyama p.c.)  
 ヲ ヲ ヌ ヲ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ  
 Jibun no yonux doyux.  
 3 GEN hair long  
 ‘His/her hair is long.’

Essentially, in terms of uniformity, Yilan Japanese is semantically transparent as regards pronouns, case-marking and lack of copula insertion. All these processes seem to be a consequence of the initial pidginisation of the language through L2 acquisition. However, in the case of adjectives in predicative sentences (see (7)), adjectives act as verbal predicates as in Atayal. The latter complexity makes the uniformity strategy less strong in the case of adjectival predicative sentences, and is possibly due to the fossilisation of Atayal sentence-forming strategies after the conventionalisation of code-switching.<sup>17</sup>

Another characteristic of the language that goes against the uniformity strategy is its variation of word order. Examples (4), (7), and (10) display an SVO word order, rather than the most common SOV order. This goes beyond the simple clause level. In the case of embedded clauses there is also a variation in word order.

- (21) ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ ㄱ  
 [Ano hito suki [nihongo yu(bu)]].  
 S V [O V]  
 That man like Japanese speak  
 ‘That man likes to speak Japanese’.

17. As a reviewer pointed out, it could be the case that *m'es* is a stative intransitive verb. This needs to be explored in detail in the future. Either way, both possibilities clearly show that an Atayal grammatical strategy is being used.





- (24) ー 口 ヲ ム ヲ 口 × ー ム ー ㄣ ㄣ ム ヲ 口 × ー ム ㄣ  
 Ima samuy-si-nay / \*samuy-sa-ng  
 now cold-NEG-REAL cold-NEG-IRR  
 ‘Now is cold.’
- (25) ヲ ム ㄣ ヲ ム ヲ 口 × ー ム ー ㄣ ㄣ ム ヲ 口 × ー ム ㄣ  
 Asta \*samuy-si-nay / samuy-sa-ng  
 Tomorrow cold-NEG-REAL cold-NEG-IRR  
 ‘Tomorrow will be cold.’

Examples (23)–(25) show the use of the realis negation marker *-nay* and the irrealis negation marker *-ng*. The former originates in standard Japanese and the latter in western Japanese dialects. The examples illustrate the productivity of the marker and how their use is sometimes not possible with certain temporal adverbs, e.g. the irrealis negation marker is incompatible with sentences that convey a past or present meaning. I suggest this is a result of grammar mixing too.

### 5.3 Simplicity

The strategy known as *simplicity* implies that “the amount of processing needed to get from a SA (Semantic Analysis) to a ShS (Shallow Structure), and vice-versa, is kept to a minimum” (Seuren & Wekker 1986: 66–67). Here, TMA markers are expected to be rendered as individual particles, which would virtually mirror the AUX area of a semantic analysis (SA) (Seuren 2018: 94), resorting to little or no LOWERING rules. In addition, SUBJECT and PREDICATE RAISING cyclic rules are dispensable.

As seen in the previous two subsections, the simplification of the case-marking system has diminished the transformational load, avoiding redundant agreement rules. As such, there is no need to mark what the subject or object is. Apparently, word order is enough. TMA marking remains to be studied in more detail. As observed in Chien and Sanada (2010) adjectival predicates can stand on their own. This would be simplification. However, verbs display tense and aspect markers, as well as phasal polarity and agent focus particles when they are of an Atayal origin. This latter case is against simplicity.

Conversely, the Japanese causative has remained almost intact, see (5) and Chien (2015: 520–521). The causative predicate retains the PREDICATE RAISING rule of Japanese. No serialisation or periphrasis is used in the language to create causative sentences. This, for example, is also against simplicity.

## 6. Concluding remarks

This short article has presented evidence that supports the idea that Yilan Japanese is not a creole but a less understood mixed language. The genesis of this vernacular seems to have been characterised by an initial introduction and learning of vernacular Japanese, followed by Atayal-Japanese code-switching, and a subsequent conventionalisation of this code-switching. The grammar of Yilan Japanese resembles that of mixed languages, such as Light Warlpiri or Gurindji Kriol. However, more work is necessary to provide a final answer.

Additionally, following the research programme of Seuren and Wekker (1986) and Muysken (2000), I have attempted to bring into focus this complex language in light of the strategies of *uniformity*, *universality*, and *simplicity* of the PST. The more semantically transparent grammatical features of Yilan Japanese seem to be related to its initial history of learning of vernacular Japanese, while the less semantically transparent features seem to correlate with the scenarios where code-switching and mixing operated.

Below I present a summary of structural and lexical sources of Yilan Japanese. Rather than a final word, this is a preliminary first step towards the understanding of this understudied mixed language. Only after a careful description of the language is conducted, taking into account cross-generational variation, will we be able to classify this vernacular and claim a rightful spot for it in modern typologies of mixed languages.

Table 5. Preliminary account on structural and lexical sources of Yilan Japanese

Structural feature	Language of origin	Lexical category	Language of origin (based on Chien, 2015)
<i>word order</i>	Japanese and Atayal (depends on the origin of the main predicate)	<i>N-body parts</i>	Atayal (some Japanese words)
<i>TMA markers</i>	Japanese and Atayal	<i>N-persons</i>	Atayal & Japanese
<i>bound verbal morphology</i>	Japanese and Atayal	<i>N-nature</i>	Atayal and Japanese
<i>case morphology</i>	Japanese	<i>N-plants</i>	Atayal and Japanese
<i>other nominal morphology</i>	Japanese	<i>N-animals</i>	Atayal and Japanese
<i>negation</i>	Atayal (using Japanese forms)	<i>V-basic</i>	Japanese
<i>pronouns</i>	Japanese and some Atayal (2, 3)	<i>Adjectives</i>	Mostly Atayal, also Japanese
<i>interrogative pronouns</i>	Japanese	<i>Numerals</i>	Atayal and Japanese
<i>deictic elements</i>	Japanese		

## Acknowledgements

This article is dedicated to a mentor and a friend, Pieter C. Muysken. Thank you, Pieter, for introducing me to the world of language contact studies as an MA student, and for shaping my thoughts on language variation and change during my four-year *séjour* in the Netherlands as your doctorate student.

I would also like to thank Felicity Meakins and Norval Smith for their initial comments and suggestions. Two anonymous reviewers must be thanked and acknowledged. Without their insightful comments, this article would not have attained its current shape. I would also like to thank my main consultant and friend Kaluo from Yilan, for his online patience and support. All inaccuracies in the article remain my own responsibility.

## List of conventions (following the Leipzig glossing rules)

1	first person	NEG	negative
2	second person	N.PAST	non-past
3	third person	O	object
AF	actor focus	PAST	past
CAUS	causative	POT	potential
COP	copula	PPP	phasal polarity particle
DAT	dative	PROG	progressive
FOC	focus	REAL	realis
GEN	genitive	S	subject
IRR	irrealis	V	verb

## References

- Arends, J. T. G. 1994. The socio-historical background of creoles. In *Pidgins and Creoles: An Introduction*, Arends, Muysken & Smith (eds), 15–24.
- Arends, J., Muysken, P. C. & Smith, N. (eds). 1994. *Pidgins and Creoles: An Introduction* [Creole Language Library 15]. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.15>
- Bailey, C.-J. N. 1973. *Variation and Linguistic Theory*. Arlington VA: Center for Applied Linguistics.
- Bakker, P., & Muysken, P. C. 1994. Mixed languages and language intertwining. In Arends, Muysken & Smith (eds), 41–52.
- Chien, Y. 2015. The lexical system of Yilan Creole. In *New Advances in Formosan Linguistics*, E. Zeitoun, S. F. Teng & J. J. Wu (eds), 513–532. Canberra: Asian Pacific Linguistics. <<http://hdl.handle.net/1885/14354>> (27 April 2020).
- Chien, Y. 2016. Yilan Creole case marking. *NINJAL Research Papers* 10: 1–17.
- Chien, Y. & Sanada, S. 2010. Yilan Creole in Taiwan. *Journal of Pidgin and Creole Languages* 25(2): 350–357.

- Chien, Y. & Sanada, S. 2011. Taiwan no giran kureouru ni okeru hiteiji: -nay to -ng no henyō o megutte (Negation in Taiwan's Yilan Creole: Focusing on -nay and -ng). *Gengo Kenkyū [Journal of the Linguistic Society of Japan]* 140: 73–87.
- Coşeriu, E. 1981. *Lecciones de lingüística general*. Madrid: Gredos.
- Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. London: Longman.
- Gorbunova, I. M. 2014. The category of phasal polarity in Atayal. *ВОПРОСЫ ЯЗЫКОЗНАНИЯ* 3.
- Gorbunova, I. M. 2017. Грамматическая семантика атайльского глагола в типологической перспективе (The Grammatical Semantics of Atayal Verbs from a Typological Perspective). PhD dissertation, Russian State University for the Humanities, Moscow.
- Kaiser, S., Ichikawa, Y., Kobayashi, N., & Yamamoto, H. 2013. *Japanese: A Comprehensive Grammar*. London: Routledge.
- Meakins, F. 2013. Mixed languages. In *Contact Languages: A Comprehensive Guide*, P. Bakker & Y. Matras (eds), 159–228. Berlin: Mouton de Gruyter.
- Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
- 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. 2000. Semantic transparency in Lowland Ecuadorian Quechua morphosyntax. *Linguistics* 38(5): 973–988.
- Muysken, P. C. 2008. *Functional Categories*. Cambridge: CUP.
- Muysken, P. C. 2009. Gradual restructuring in Ecuadorian Quechua. In *Gradual Creolization: Studies Celebrating Jacques Arends* [Creole Language Library 34], R. Selbach, H. Cardoso & M van den Berg (eds), 77–100. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.34>
- Muysken, P. C. & Smith, N. S. H. 1994. Introduction: The study of pidgin and creole languages. In *Pidgins and Creoles: An Introduction*, Arends, Muysken & Smith (eds), 3–14.
- Myers-Scotton, C. 1993. *Duelling languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, C. 2003. What lies beneath: Split (mixed) languages as contact phenomena. In *The Mixed Language Debate: Theoretical and Empirical Advances*, Y. Matras & P. Bakker (eds), 73–106. Berlin: Mouton de Gruyter.
- O'Connor, L. M. & Muysken, P. C. 2014. *The Native Languages of South America. Origins, Development, Typology*. Cambridge: CUP.
- Qiu, P. 2015. A Preliminary Investigation of Yilan Creole in Taiwan: Discussing Predicate Position in Yilan Creole. MA thesis, University of Alberta.
- Rau, D.-H. V. 1992. A Grammar of Atayal. PhD dissertation, Cornell University.
- Sanada, S. 2010. Yilan Creole in Taiwan: A Japanese-lexicon creole. Lecture presented at the International Centre for Japanese Studies. Retrieved from: <[http://www.tufs.ac.jp/icjs/en/activityreports/pdf/20101016\\_01.pdf](http://www.tufs.ac.jp/icjs/en/activityreports/pdf/20101016_01.pdf)> (27 April 2020).
- Seuren, P. A. M. 1972. Autonomous versus semantic syntax. *Foundations of Language* 8(2): 237–265.
- Seuren, P. A. M. 2018. *Semantic Syntax*, 2nd rev. edn. Leiden: Brill.
- Seuren, P. A. M. & Wekker, H. 1986. Semantic transparency as a factor in Creole genesis. In *Substrata Versus Universals in Creole Genesis: Papers from the Amsterdam Creole Workshop, April 1985* [Creole Language Library 1], P. C. Muysken & N. Smith (eds), 57–70. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.1>
- Silverstein, M. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3–4): 193–229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2)

- Smith, N. S. H. 1994. An annotated list of creoles, pidgins, and mixed languages. In Arends, Muysken & Smith (eds), 331–374.
- Thomason, S. G. 2003. Social factors and linguistic processes in the emergence of stable mixed languages. In *The Mixed Language Debate: Theoretical and Empirical Advances*, Y. Matras & P. Bakker (eds), 21–40. Berlin: Mouton de Gruyter.
- Thomason, S. G. & Kaufman, T. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley CA: University of California Press.
- Todd, L. 1980. Pidginization and creolization. *Annual Review of Applied Linguistics* 1: 19–24.  
<https://doi.org/10.1017/S0267190500000477>
- Tsurumi, E. P. 1977. *Japanese Colonial Education in Taiwan, 1895–1945*. Cambridge MA: Harvard University Press.

# Pottefers Cant, Groenstraat Bargoens, and the development of “have” and “be” in the wider context of contact

Norval Smith and Frans Hinskens

University of Amsterdam / Meertens Instituut &  
Radboud Universiteit Nijmegen

Our article falls into two parts. In the first part we compare two “secret” or replacive Dutch languages, Potteferstaal (Pot-repairers language) and Groenstraat Bargoens (Groenstraat cant), which developed in and around eastern Belgium and the southeastern fringe of the Netherlands, respectively. Nothing has been published on the first as yet (but now see Van Hauwermeiren (2018), while the second has been well-studied. Speakers of both cryptolects used similar strategies, the replacement of frequent lexical items by alternate ones with the purpose of inhibiting comprehension by outsiders. We also provide a general account of the social conditions surrounding the use of this type of replacive language. The social history of Groenstraat Bargoens is covered in some detail, while that of Potteferstaal is described as fully as possible, given our limited information. In the second part of our article we treat a striking structural difference shared by Potteferstaal and Groenstraat Bargoens and other similar replacive languages. This concerns the fact that various uses of “have” and “be” verbs, including “main verb”, auxiliary and copular usages are expressed by a single form. We add three other so-called “Bargoens” varieties from northern Belgium and Dutch Limburg to the study, brief information on the Kempen traders’ cant, as well as “Henese Fleck” spoken in Breyell just over the German border, and discuss their parallels and differences in connection with this feature. We relate this phenomenon to the retention of a small but persistent number of Romani lexical items by most of these languages. One group which exhibits the same identity of expression of “have” and “be” as well as formerly similar historical ways of life, is the Sinti Romani. We ascribe this shared structural pattern to early contact between Dutch and Belgian “travellers” and Romani, despite the fact that the languages of the first group are inflection-poor, whereas Romani languages are inflection-rich.

**Keywords:** secret language, cryptolect, language contact, borrowing, lexical replacement, combiverb, Judeo-German, Romani, Sinti, travellers, Bargoens, woonwagenbewoners

## 1. Introduction

On the whole it can fairly be said that, maybe with the exception of Groenstraat Bargoens, not much has been done in the way of serious linguistic description on the so-called “secret languages” of the Low Countries. There are two reasons. These languages had by and large lost most of their (original) functions by the inter-war period in the 20th century. Partly this was due to the dying-out of various, frequently peripatetic, trades such as those of travelling salesmen of cloth and clothing, repairers of pots and pans, knife-grinders, and other tradesmen, horse-traders, hawkers, and so on. The trades which provided them with a livelihood had become more and more redundant as the twentieth century progressed.

As the general public became more and more mobile, in step with improvements in communications, they had less and less need, or inclination, to wait for tradesmen to come to them, whether it was to repair things that had got so vastly cheaper that it was easiest just to replace them, or to sell them things from the limited range necessitated by having to physically transport them.

This meant that researchers (Endepols 1924; Moormann 1932, 1934, 2002; Hermant 1933, 1933/1934; Dewulf 1980; Hinskens 1985, Van Hauwermeiren 2012)<sup>1</sup> were usually limited to what they could extract from a few, usually elderly, speakers.

The list of trades we have just mentioned displays a certain hierarchy, in terms of respectability. However the users of secret languages were by no means restricted to “respectable” people. We can continue the list of users down the social scale, to thieves, pickpockets and others of less repute. In particular in the larger towns and cities, similar “secret languages” were employed by petty criminals (or “ruffians”) of various types. There was however a significant distinction in terms of register (Moormann 2002: 68). Moormann draws a four-way distinction between:<sup>2</sup>

- (1) a. Ambulant traders’ languages (Kramertaal)
- b. Bargoens, the language of tinkers (in Dutch: *woonwagenbewoners*, lit. ‘caravan-dwellers’ or *reizigers* ‘travellers’), practising various trades.
- c. Thieves’ language (Boeventaal)
- d. The language of cattle-traders (“Jargon” as defined by Moormann)

In practice, however, as Van Hauwermeiren (2004: 169) points out, Moormann treats Boeventaal (1c) (the language of criminal elements) as a sub-variety of Bargoens (1b). Obviously the presence of a highly specialized register relating to the

---

1. We have excluded works that are basically restricted to lexical studies.

2. A fifth type, Letter Language (Lettertaal), involving the transpositions of letters or sounds, we will not discuss.

concepts of stealing, doesn't really change the basis of Bargoens, any more than the highly specialized plumbing register of Dutch plumbers alters the basis of Dutch.

The Bargoens elements that have made their way into the vocabulary of popular Amsterdam speech, filtering up into less working-class Amsterdam Dutch, involve normal Bargoens terms, and not in general the more specialized register associated with criminal concepts.

We do not agree with Van der Sijs in her introduction to Moormann (2002) when she identifies popular city speech (Volkstaal) with Bargoens. In short, we do not regard Bargoens as a suitable term for referring to substandard (and often just urban) language just because the latter has adopted a number of lexical items from the former.<sup>3</sup> Moormann (1932) draws a general distinction between *kramertalen*, which (as he adds) contain few originally Romani and Jewish items, on the one hand and *boeventaal*, ‘ruffian’s language’, and *bargoens* in the strict sense, such as e.g. Rotwelsch, the German (and Swiss and Austrian) ambulatory traders and vagrants’ cant (Kluge 1901) on the other. Among the various occupational groups as well as between the various types of secret language there were gradual transitions to be found. As appears from the popular designations of secret languages (e.g. Taal van de jongens van de vlakke<sup>4</sup> (Köster Henke 1906), Gaunersprache<sup>5</sup> (cf. Günther 1919), Gabbertaal<sup>6</sup> (cf. van Bolhuis 1937), and, in connection with Groenstraat Bargoens, boeventaaltje,<sup>7</sup> thieves, vagrants, roaming journeymen, hawkers and other traders, Jews and Gipsies were often lumped together.

In this contribution we will first introduce two Dutch secret languages, which are both located within the geographical space of the Brabant and Limburg dialect areas of Dutch dialects, respectively, and nicely illustrate the differences that may exist among such languages (Sections 2 and 3). While the Brabant secret language, Pottefers Cant (‘Pot-repairers’ language’), has hardly been documented or studied (but cf. Van Hauwermeiren 2018), the Groenstraat Bargoens of Limburg has been well-studied. We will compare the two languages in respect of the lexicon in Section 4, and of morphology in Section 5, paying special attention to hitherto unstudied grammatical aspects. In Section 6, 7 and 8 we will focus on a specific structural similarity in the grammar of these and related cryptolects and its possible origin. In Section 9 we will deal with the small shared Romani lexical element. Finally, in Section 10 we will briefly discuss desiderata for future research.

- 
3. Often with changes in meaning.
  4. “De jongens van de vlakke” (lit. ‘the boys of the flat (area)’) = street criminals.
  5. Ruffians’ language.
  6. Buddies’ language.
  7. Rascals’ slang.



Alas, it is too late for a Dutch equivalent of the German project on Rotwelsch and horse-traders' secret languages described in Siewert (1997; cf. 1991). There are still however apparently a few individual speakers of Bargoens in Belgium (e.g. Van Hauwermeiren 2004: 162–163, and p.c., and in the Netherlands (Haalboom 1995: 16–17), and these should be studied where possible.

## 2. The Pottefers and their hidden secret language

*Pottefers Cant* (literally 'Pot-doers' language') was the secret language employed by travelling pot-repairers, or *Pottefers* (<Fr. "Pots à faire!"), in the Tienen-Hoegaarden area of Belgian Brabant. This area is located some 30 km southeast of Leuven, in the eastern part of the Dutch-speaking area in Belgium. However, the ancestress of the last-known Pottefers Cant-speaking family, the Imbrechts, was born in Dutch Limburg.

As the manuscript article which is the source of our information (Hubert van Nerum ca. 1964)<sup>8</sup> states:

- (2) "Te Wittem in Hollands Limburg werd in 1841 Gertruijde Schaeken geboren, als dochter van schaapherder Franciscus Schaeken en Margriet Schwanen."  
 "Gertruijde Schaeken was born in Wittem in Dutch Limburg in 1841, as the daughter of shepherd Franciscus Schaeken and Margriet Schwanen."

This event appears in a transcript of the official register of births as:

- (3) Details of Gertruijde Schaeken<sup>9</sup>

	Born	Schaken, Gertrudis
Wittem 25-01-1841	Father of the child	Schaken, Jan Francis
Report of birth	Mother of the child	Schwanen, Anna Margaretha

Although none of the names in the manuscript is orthographically identical to the name that appears in the official register, there can be no doubt about the identity of the people involved:

8. A study of the available material is Van Hauwermeiren (2018).

9. <<http://www.archieven.nl/mi/1540/?mivast=1540&miadt=38&mizig=100&miview=tbl&mi-lang=nl&micols=1&mires=0&mip1=Schaken&mip3=Gertrudis>>

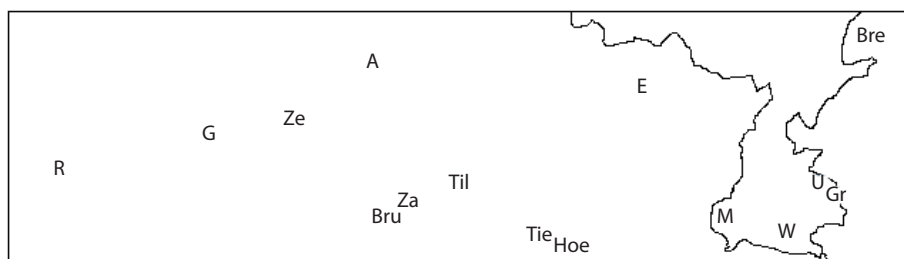
**Table 1.** Gertruijde Schaeken, the ancestress of this Pottefer family

Relationship	Pottefers manuscript		Official transcript	
Child	Gertruijde	Schaeken	Gertrudis	Schaken
Father	Franciscus	Schaeken	Jan Francis	Schaken
Mother	Margriet	Schwanen	Anna Margaretha	Schwanen

The variations in spelling and number of forenames we see here is typical even of Dutch *official* records in the first half of the 19th century.

The map below indicates the approximant relative positions of the places that are of importance in the family history of the ancestress, along with the positions of Maastricht, Brussels and Antwerp as reference points, as well as other locations that are relevant for our article, such as the locations of Bargoens varieties discussed below.

The national borders are indicated. Most of the map is concerned with Belgium (on the left). The Netherlands is located in the right middle, while Germany is on the extreme right.

**Map** of places in Belgium (left), the Netherlands (right middle) and Germany (right)**Table 2.** Legend of abbreviations of places on map

A	Antwerp	Bre	Breyell	Bru	Brussels
E	Eksel (Exel)	G	Gent (Ghent)	Gr	Groenstraat
Hoe	Hoegaarden	M	Maastricht	R	Roeselare
Tie	Tienen	Til	Tildonk	U	Ubach over Worms
W	Wittem	Za	Zaventem	Ze	Zele

Gertruijde Schaeken was married twice. She had four children from her first marriage, ca. 1865, one born in Hoegaarden, only 5 km SW of Tienen, and three born in Zaventem, near Brussels. From her second marriage in Hoegaarden to Jan Baptist Imbrechts (1850–1881) of Tildonk she had two more children, born in Tienen.

From the time of her first marriage the family had become travelling Pottefers, and they continued in this line of business up till nearly 1914. Then Gertruijde went to live with her son Florimond who had opened a lodging-house for pedlars and other travelling tradesmen. In the years following the Second World War Florimond took up the trade of Pottefer again for a number of years. He died about 1966.

The name Pottefer is derived from the French *Street Cry* “Pots à faire.” Whenever Pottefers arrived in a village, they would camp in a place where they could dig out a hollow for a fire. Then they would go along the houses and farms calling out this *Street Cry*, or its Dutch dialect version, “Potte te flikken”<sup>10</sup> People would bring out their possessions that needed repaired.

To repair damaged earthenware the Pottefers would drill holes around the crack or hole, and join these up with iron hooks. Then they would cover up the cracks and hooks with clay, and bake the repaired earthenware vessels in a fire. In later times they would also repair metal objects. After a few days repairing things they would then move on to the next village.

Hoegaarden was the place the Pottefers Cant manuscript was probably recorded in 1964 (Van Hauwermeiren 2009: 149). Van Hauwermeiren is not entirely clear on this, but the most likely recorder of this manuscript seems to have been Hubert van Nerum (1923–1994) who was, among many other things, an amateur ethnologist. At the time when a copy of the manuscript was supplied to the Nijmeegse Centrale voor Dialect- en Naamkunde,<sup>11</sup> Hubert van Nerum was the chairman of the “Museum J. van Nerum”, ‘t Nieuwhuys, Hoegaarden. This museum existed from 1965 until 2001.

In all probability, the Pottefers Cant word-list was obtained from the above-mentioned Florimond Imbrechts, who also served as an informant for Tienen (P145) for the *Woordenboek van de Brabantse dialecten* (‘Dictionary of the Brabantish Dialects’) (Vos & van Lieshout 1990), in particular Part II, Section 5 (1990) (including potters’ terms), produced by the former *Nijmeegse Centrale voor Dialect- en Naamkunde* (‘Nijmegen Centre for Dialectology and Onomastics’), Faculty of Arts, Nijmegen university.

---

10. *Flikken* is a Limburg dialect and Potterfers Cant (Bargoens) word for ‘do, make.’

11. The Nijmegen Centre for Dialectology and Onomastics.

### 3. The “not so well-hidden” secret language of the Groenstraat pedlars

Groenstraat Bargoens (henceforth GrB) had already been touched upon by Winkler (1874), and Van Ginneken (1914), who plagiarized Winkler shamelessly. The first relatively thorough study was Endepols (1924), while Moormann (1932) paid systematic attention to GrB. These linguists mainly concentrated their studies on the typological status of GrB, tracing its various lexical and (to a lesser degree) morphological relationships, and were thus able to identify most of the diverse strata it consists of. They also discussed the relationships which GrB bears to its most important exogenous sources (German and Flemish cants and slang, as well as Latin, French and – to a lesser extent – Judeo-German/Hebrew<sup>12</sup> and Romani), and its functional similarities to, and differences from, these sources.

#### 3.1 Development

After a sketch of the location and the social history of Groenstraat and the surrounding area, we will outline the external and internal history of GrB.

Until the beginning of the 20th century, the region of southeast Limburg was economically mainly oriented towards agriculture and the density of population was relatively low. “Bad and rough roads constituted the poor connections between the small villages and hamlets” (Breuer 1974: 7, our translation, NS&FH). Around the year 1885 Ubach over Worms consisted of the parishes of Waubach (the administrative center) and Rimburg (which was united with Waubach in 1886 to become the municipality of Ubach over Worms), as well as the township of Groenstraat. The majority of the more well-to-do farmers and their families lived in Rimburg and in particular in Waubach. Most of the inhabitants of Groenstraat did not have any fixed source of income and were relatively poor; in the best case they might own a goat. Most inhabitants (not only men) of Groenstraat and the neighbouring village of Nieuwenhagen earned their living in a variety of uncertain manners. Many worked as day labourers and seasonal workers on farms, and in brickfields on the other side of the German border, the so-called *brikkebekkere*, lit. ‘brick-bakers’.

---

12. As reviewer Yaron Matras points out, “there were secret languages in the region that were particular to Jewish traders – usually known as Lekoudesch, Lottegorisch, etc, a euphemism for Loshn-koudesh ‘holy tongue’. They consisted of lexical elements from the liturgical Hebrew scriptures, in Ashkenazi pronunciation, that were embedded into Judeo-German, not Yiddish. Etymologically, of course, the words in question are ultimately Hebrew.”

Peddling was also a very important means of sustenance; small-scale ambulant trade flourished in the period from 1700 to 1900 (Goossens 1981: 317–326). The pedlars' trade circuits were very wide: in their travels they frequently visited central Germany, as well as Luxemburg and the Brabantish Kempen (east of Eindhoven).<sup>13</sup> Their journeys often lasted several months.

In the course of their trade, and through contact with colleagues from these regions, the Groenstraat pedlars had developed a linguistic system which was later labelled Groenstraat Bargoens by Endepols (1924); it was essentially a basic -vocabulary-relexified variety of their mother-tongue, the Groenstraat dialect.<sup>14</sup> The “secret” language items were borrowed from a number of *Krämersprachen*, ‘pedlars’ cants (Endepols 1924: 173,174) – in this connection, one of Hinskens’ informants in his 1982 fieldwork used the term [kri<sup>1</sup>əmərwəlʃ],<sup>15</sup> which can be translated into Dutch as ‘kramerswelsj’ or ‘kramerwaals’, to refer to the unintelligible cant used by pedlars and hawkers (‘kramers’).<sup>16</sup> For the pedlars, GrB functioned partly as a professional language and partly as a secret language. In addition to the economic situation, the presence of the nearby German border motivated the development of a secret language which was spoken in the twilight zone between ambulant trade and smuggling. Stüve (1923: 10) observed that travelling salesmen’s languages occur preferably in border towns; for that reason they are often associated with contraband.

If nowhere else, the Groenstraat traders would meet their colleagues from other places in the lodging houses and barns where they would spend the night. The origins (of the lexical replaceive items) of the GrB lexicon must in the first place be sought in the *Henese Fleck*, which used to be spoken by traders<sup>17</sup> from Breyell (Germany, some 15 kms east of Venlo) which “was said to comprise only some 300 content words” and used “the function words of the Breyell dialect to construct sentences” (Kluge 1901: 447, our translation, NS&FH; cf. Winkler 1874: 411).<sup>18</sup>

---

13. This is the area around Eksel – E on the map – and continuing north across the Dutch border.

14. Located in the dialectal transition zone between Ripuarian and East Limburg dialects (Goossens 1965).

15. Realized with an acute tone (Tone 1), which contrasts with a circumflex tone (Tone 2); Limburg dialects are ‘pitch accent’ languages.

16. Also Scots *cramer*. The word *cramer* in Scots ‘market salesman, packman’ is a Dutch loanword.

17. Known as *Kiepenträger*, lit. ‘basket bearers’, those who carry baskets on their back.

18. This is not strictly true. See our remarks on Henese Fleck below.

The language of the Teuten, the pedlars and travelling craftsmen from the *Kempen* area,<sup>19</sup> was also important for the development of GrB. In the period between 1670 and 1870, the Teuten traded, among other things, in human hair, from which the wholesale buyers made wigs and little portraits (Knippenberg 1974: 166–184). The language of the Teuten, which according to Winkler (1874: 411–416) belonged to the group of the “Westermaassche Roodwaalsch”, i.e. ‘Rotwelsch or Bargoens spoken west of the River Meuse,’<sup>20</sup> “was not a *kringtaal*, ‘local cant’; the other members of the families of the speakers were not even aware of its existence.” The language of the Teuten was used fairly infrequently by those involved” (Van Winkel (1974: 33), our translation, NS&FH).

The Groenstraat pedlars also traded in (women’s) hair. In the second half of the 19th Century, and especially in the period around 1870, the hair-trade was thriving, in the dialect referred to as [ho.əɪʃnɪt], lit. ‘hair cut(ing)’. The hair was sold for the production of wigs and little portraits, as above-stated, but also for watch chains and jewelry. Endepols (1924: 174) ascribed a “primitive character” to GrB; this concerned primarily GrB’s limited functional elaboration, since it was almost exclusively used amongst the relatively small group of hawkers and hair-traders, mainly while bargaining.

### 3.2 Ausbau (expansion) and subsequent functional specialization

In particular in the decades before the turn of the century, GrB underwent a considerable functional growth. While it originated as a lexically modified trade cant, it gradually became an in-group code for the inhabitants of Groenstraat. GrB was no longer exclusively used as a cant/secret language for trade, since it was spoken more and more often in the context of the family. As the usage of GrB became gradually associated with ordinary Groenstraat life, many GrB words were adopted into the dialect (cf. Goossens 1981: 559). By the time the pedlars’ guild was approaching extinction, it had become customary in Groenstraat, and also in the other Ubach over Worms neighborhoods – even among the higher social classes – to use GrB words when speaking the dialect. In addition, inhabitants of Groenstraat would speak GrB among themselves when in the company of those who were not familiar with Bargoens.

---

19. Moormann mentions the villages of Weert, Tungelrooi, Exel and Belgian Hamont; in Belgium, Teuten also came from Lommel, Achel, Overpelt and Neerpelt.

20. Unlike the cryptolects from Breyell, Nieuwenhagen and Groenstraat, which Winkler classified as “Oostermaassch Roodwaalsch,” Rotwelsch spoken east of the Meuse.

The rapid rise of coal mining in the area meant the end of seasonal and peripatetic work<sup>21</sup> and large-scale smuggling, and hence the beginning of the end of GrB. Nevertheless, after 1900 and in the decades of growth of coal mining,<sup>22</sup> GrB underwent an internal expansion. Analogical forms were added to the lexicon and the use of derivational morphology was extended. Examples are *spöaret*, ‘bicycle’, *fietsjanoek*, ‘train conductor’, in Endepols’ (1924: 178) transliterations.

Some of the new words were regarded by the most genuine old Groenstraat dwellers as neologisms. These were not genuine Bargoens words, they had been created by the younger generation on the model of older words. Making a disparaging gesture, one older Bargoens speaker called them miners’ words

(Endepols 1924: 178, our translation, NS&FH)

Endepols based his description among other things on data from conversations he had with “native speakers.” Linguistically significant is the fact that, apart from the extension of the lexicon, GrB underwent a notable structural addition: while the first “layer” did not contain any compounds, this more recent “layer” does. Most neologisms, such as *voonke-knoeker*, ‘flint-stone’; ‘lighter’ (from *voonk*, ‘spark’, and *knoeke* ‘hit’ which, however, is not a common GrB word), are said to have been introduced by the native Groenstraat dialect-speakers among the coal miners, who used GrB deliberately as a secret language, especially in the presence of their betters who did not understand it. In spite of this, and probably because of the relatively small number of coal miners who knew it, GrB never acquired the new function of a coal miners’ work language (a *Bergmannssprache*, one of the *Berufssprachen*, cants sketched by Bischoff (n.d. 3–165). Had this happened, it would have lost that function anyway after 1975, when all the Limburg coal mines were closed.

In short: GrB underwent Ausbau (expansion) as a result of 1. the extension of its usage from a mere secret trade cant to a Groenstraat in-group code; 2. the broadening of the speech community; consequently the speakers were no longer only pedlars, but included more of the well-to-do, and more sedentary groups from all over Ubach over Worms; 3. a fairly short phase of existence as a coal miners’ secret code. At the same time, however, functional shrinkage occurred. As a result of the enormous social-demographic changes brought about by the sudden boom

---

21. Some traders became successful, though. Antoon Jurgens (1805–1887), the grandson of Groenstraat-born pedlar Dionysus Jurgens, founded a company for the international trade in butter; three of his sons developed this into a margarine factory in Oss (Brabant), which would eventually develop into the multinational company, Unilever.

22. In 1902 Mining Laws were passed; in 1927 there were 12 mines in Limburg with a total of some 35,000 employees.

in coal mining,<sup>23</sup> Groenstraat was ‘opened up’, which after World War II had a negative effect on the usage of GrB.

Nowadays GrB is almost exclusively used by members of older generations. By relexifying dialect utterances with GrB words a certain in-group feeling can be created. But in the days of Hinskens’ 1982 fieldwork in Ubach over Worms even the few ‘authentic’ GrB speakers already showed a degree of ‘linguistic insecurity’, which may not only have been caused by the increasing forgetfulness of old age.

### 3.3 Groenstraat Bargoens’ third youth?

To judge by the publication of a dictionary (Bosten et al. 1981), at the end of the 1970’s there was a revival of the interest in GrB, which in Ubach over Worms is encompassed with *Alltagswissen* or ‘implicit common everyday knowledge’, concerning Groenstraat (carnival) traditions. Local lore has it that these traditions are rooted in the [ho.əʃfɪnt] ‘[het] haarsnijden’, the former hair trade.

Some common expressions as well as commercial slogans were jocularly translated into GrB, such as e.g.

- (4) GrB. kwɪz ət mɪt flɔskəl-ə<sup>24</sup>  
 IMT say 3SN WITH flower-PL  
 SD. ‘Zeg het met bloemen.’  
 ET. ‘Say it with flowers.’

Apart from the words which had penetrated into the dialect, GrB was no longer in living use at that time.

In the second half of the 1990’s there was yet another wave of interest in GrB. In 1996 the *Grunsjstroater Bargoensjkloeb Ut Keieskwieëste*, lit. ‘Groenstraat Bargoens Club (for) Street Speaking’, or street language (in an unusual sense of this latter notion), was founded. Its members meet regularly in order to speak GrB and to promote GrB usage.

23. From 1880 to 1930, i.e. in 50 years, the number of inhabitants multiplied by a factor of 5.5, and in 91 years from 1880 to 1971, this multiplied by a factor of 8.2 (Breuer 1974: 8, 20). In other words, there was a huge positive balance of immigration. The newcomers added a range of more or less related dialects of Dutch, as well as varieties of Frisian, German, French, Polish and Czech to the linguistic repertoire.

24. From now on we will make use of IMTs – *interlinear morpheme translations*. G(r)B refers to Groenstraat Bargoens, SD refers to the translation in Standard Dutch, while ET refers to the English Translation.



The following two illustrations are representative of this period. The first is a newspaper cutting entitled “Vers bloed voor Groenstraat-Bargoens” ‘New blood for Groenstraat-Bargoens’. The folkloristic aspect is obvious from the picture.



**Illustration 1.** Newspaper clipping with a picture of ‘third generation’ speakers wearing 19th century clothes (cf. Hobsbawm 1983)

The next illustration is of the greeting and first sentence of a trilingual letter inviting the father of one of the authors of this paper to a retirement party. This demonstrates the local use of Groenstraat Bargoens. The three languages involved are (a) Dutch, (b) the local Groenstraat Limburgish dialect, and (c) Groenstraat Bargoens. We have enlarged the font of the sentence to make it more easily legible. We have also added an English translation.

It only requires a cursory examination to spot the fact that the function words used in the Bargoens version are the same as those of the local dialect. There is only one exception, caused by the fact that the word corresponding to the English “business” differs in gender between the local dialect and Bargoens versions, and therefore requires a different form of the definite article.



## *Dassen Transport b.v.*

*Groenstraat 107 6374 JM Landgraaf*  
*sinds 1928*

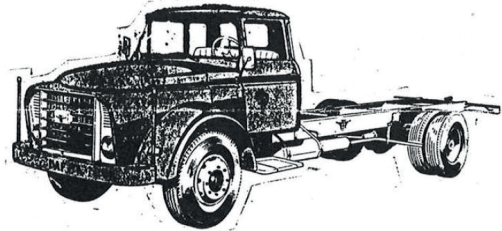
de Heer P. Hinskens  
Nieuwenhagerstraat 8  
6374 XR Landgraaf

Landgraaf 19-04-99

GEACHTE VRIEND

*Geachte vrunk*

Grannige fiese



[Dear friend]

NA MEER DAN 40 JAAR IN ONS BEDRIJF GEWERKT TE HEBBEN HEEFT VOOR SJENG OOK DE KLOK GESLAGEN OM VAN DE VUT TE LEVEN.

*Noa mië wie fietig joar in 't geschef gewirkt te han, had vuur der Sjang òg de klok geslage om va de vut te lève.*

Noa mië wie derès jamkes in der rögel geböegeld te han, had vuur d'r Sjeng òg de klak geboozd om va de vut te lève

[After more than 40 years working in the business, the bell has sounded for Sjeng too, to live from his pre-pension.]

DUTCH

*Groenstraat Dialect*

Groenstraat Bargoens

[English]

### Illustration 2. ‘Trilingual’ letter of invitation for a celebration (part)

In 2002 the chairman of the *Bargoensjkloeb* published a non-specialist book containing stories written in GrB, together with a wordlist (Vek 2002). There is also a CD with songs in GrB, written and performed by Vek. In addition the *Bargoensjkloeb* have set up a website dedicated to GrB: <[www.groenstraat-bargoens.nl/](http://www.groenstraat-bargoens.nl/)>

In the 2002 book there are indications that even insiders no longer know whether certain divergent<sup>25</sup> words (heteronyms) used to be part of the traditional dialect or rather of GrB. The German dialectologist and historical linguist Günter

25. Different from both Standard Dutch and the surrounding Limburgish dialects.

Bellmann (2003) has called a similar type of phenomenon, *ethnolinguale Unbestimmtheit*, ‘ethnolinguistic indeterminacy’. Bellmann discusses two instances of this phenomenon: (i) until a few generations ago, the speakers of certain German dialects in Hungary were not aware that they spoke a German dialect; (ii) certain, partly adapted, originally German loanwords in Sorbian are now considered to be Sorbian (“Wendisch”) by the speakers.

A few generations of Groenstraat inhabitants have maintained GrB as an ‘in-group’ code; today it leads a largely artificial life and possesses a rather mythical status.

#### 4. The lexica of both cryptolects, their origins and overlap

##### 4.1 Pottefers Cant

*Ethnolinguistic indeterminacy* can also be observed in the wordlist in Van Nerum’s undated manuscript (ca. 1964). For example, *snaps*, ‘gin’, Standard Dutch *jenever*, is a dialect word in the Ubach over Worms region, and in colloquial German it is a generic label for ‘liquor’; in GrB it is referred to as [ʏʁi.<sup>1</sup>s]. A similar case is what Van Nerum recorded for standard Dutch “praten; zeggen”, ‘to speak, to say’, Pottefers Cant (henceforth PC) *smoesteren*. In the Ubach over Worms region [ʃmu.<sup>1</sup>zə] is a generally used dialect form for ‘to speak secretly’; in GrB ‘to speak, to say’ is [kwi.<sup>1</sup>əstə]. A number of items (mainly function words) which Van Nerum includes in his alphabetically ordered PC word-list are in fact Brabantish (in some cases Limburgish) dialect variants and not secret language words at all. These include:

(5) PC word-list	Standard Dutch meaning	English meaning
<i>wa</i>	beetje; weinig; wat	a bit; what
<i>da</i>	dat	demonstrative; relativizer ( <i>that</i> )
<i>nie</i>	niet	negative ( <i>not</i> )
<i>dee</i>	die	demonstrative
<i>em</i>	hem	him
<i>he</i>	hij	he
<i>de, ‘t</i>	het	it, neuter def. article; neuter pronoun
<i>ma</i>	maar	but
<i>na</i>	naar	to (locative prep.)
<i>teige</i>	tegen	against
<i>heure</i>	uw	2SG possessive pronoun

To judge by Van Nerum’s manuscript, PC has a considerable number of items in common with what Moormann (1932) referred to as the northwestern group of secret languages of Dutch, i.e. those spoken in the urban centres in Holland:<sup>26</sup> Amsterdam, Rotterdam, The Hague and Utrecht. This subset includes:

(6)	PC <sup>27</sup>	source	Dutch meaning	English meaning
	<i>wout</i>	MD <i>waut</i> <sup>28</sup>	agent	police officer
	<i>leep</i>	? <sup>29</sup>	behendig, sluw	agile, sly
	<i>liep</i>	cf. <i>leep</i>	slim	smart, clever
	<i>dokken</i>	?	betalen	to pay
	<i>maf</i>	?	dom	stupid
	<i>bikken</i>	MD <i>bicken</i> <sup>30</sup>	eten	to eat
	<i>poen</i>	?	geld	money
	<i>tof</i>	Hebrew	goed	good, fine
	<i>joekel</i>	Romani ‘dog’	hond	dog
	<i>tippelen</i>	Hebrew	lopen	to walk
	<i>bink</i>	Romani ‘devil’	man	man
	<i>link</i>	?	gevaarlijk; slim	dangerous; clever
	<i>ploemp</i>	onomatop. <sup>31</sup>	water	water

#### 4.1.1 Combiverbs

Certain secret languages have what we will refer to as a “combiverb,” a verb which can serve both as a *be*-verb and as a *have*-verb in certain functions – including those of auxiliaries, lexical verbs (existential *be*, and possessive *have*), and the copula function of *be*. The (combination of) functions vary in the different secret languages.

In the case of the Pottefers Cant the combiverb is *mozen*, stem *mooz*.<sup>32</sup> *Mozen* is a weak verb, forming its past participle with the circumfix *ge-...-d*.

26. We use this term as a cover-term for the provinces of Noordholland, Zuidholland, and Utrecht.

27. The Standard Dutch meanings of PC items given do not always correspond to the meanings of the corresponding items in urban popular speech.

28. Probably Middle Dutch *waut* ‘authority’.

29. <<http://etymologiebank.nl/trefwoord/leep1>>; Van der Sijs (ed.) 2010.

30. Middle Dutch ‘to pick with a beak’.

31. Compare Early Modern Standard Dutch *plomp* ‘splash’.

32. The Bargoens stratum in the Roermond dialect has *maes*, ‘have, has’ (sg), see Bakkes (2007).

#### 4.1.2 *Personal pronouns*

A number of personal pronouns have special forms in PC.

(7)	<i>mechels</i>	1SG (all cases)
	<i>van mechels</i>	1SG possessive
	<i>hoeg</i>	2SG (all cases)

When these are subjects the special 1st and 2nd person pronouns generally take 3rd person inflection.

#### 4.1.3 *Negation*

PC has a number of special negative forms.

(8)	<i>den drol (van)</i>	nothing, no (quantifier)
	cf. [ <i>de gal</i> ]	[something]
	<i>drol</i>	un-
	<i>noppes</i>	nothing, No!

#### 4.1.4 *Numerals*

Special forms were noted for the first five cardinal numerals.

(9)	<i>enkels</i>	one
	<i>bis</i>	two
	<i>draaiers</i>	three
	<i>verkers</i>	four
	<i>vinken</i>	five

## 4.2 Groenstraat Bargoens

GrB had many etymologically diverse – for example, Hebrew or Judeo-German – elements which had become adapted phonologically and grammatically to the Groenstraat variety of the local dialect. According to Endepols (1924: 178) most of these belong to the older lexical strata.

Endepols (1924: 176) points out that GrB does not contain many “thieves” cant words (“dievenwoorden”), adding that it is thus not a *Bargoens*<sup>33</sup> system in the sense in which this notion is generally used. According to Endepols, GrB has numerous words in common with other traders’ languages. And in contrast to PC, GrB also had a relatively small number of items in common with what Moormann (1932) referred to as the northwestern group of Dutch secret languages, which tended to function in rather more criminal contexts.

---

33. Here we are not in agreement with Endepols in his suggestion that the term *Bargoens* normally implies a secret language used by criminals.

On the basis of etymological and comparative lexicographical research, Hinskens (1985) reconstructed the following geographical sources of the GrB lexicon:

- from the Kempen traders’ cant via Tiöttensprache<sup>34</sup> (or Humpisch), which used to be spoken by salesmen from Mettingen<sup>35</sup> and Tecklenburg (20 km west of Osnabrück, North Rhine-Westphalia);
- from the Romani language via Rotwelsch/Bargoens and via the ambulant traders’ languages Henese Fleck and Tiöttensprache;
- from Hebrew and/or Judeo-German via Rotwelsch/Bargoens, again via the ambulant traders’ languages Henese Fleck and Tiöttensprache. However, according to Moormann (1932: 123), none of the dozen “Jewish elements” of GrB<sup>36</sup> he listed occurred in Henese Fleck and Tiöttensprache and so they must have been more recent.

In the Dutch villages in the Kempen region the traders’ cant disappeared well before 1900. In Belgium, the language seems to have run a different social course, according to Labbé (1892: 25), who stated that it was spoken by members of the lower classes, ..., visitors to fairs as well as hawkers. “But it should be noted that the language is used less and less” (our translation, NS&FH).

34. <Got. *tiuhan*; Old Sax. *tiohan*, cf. modern German *ziehen*, thus Stüve 1923: 16.

35. We have not included Mettingen in the map we give above, for practical reasons. The distance between Maastricht (M) and Breyell (Bre), both on the right-hand side of the map, is about 70 km. Mettingen is a further 170 km on in the same direction. In the area, Mettingen is known as ‘das Tüöttendorf’, the village of the Tüötten or Tiötten. Traditionally, flax and hemp were cultivated; they were used to weave linen, which became an important item of merchandise. Initially the trade was combined with seasonal work (in agricultural work in the Netherlands) and became ambulant trade – the northeast of the Netherlands as well as in the north and northeast of Germany. Only with the rise of the textile industry did this ambulant trade, in which the majority of Mettingen adult males had been involved for a long period of time, gradually disappear. The founders of the clothing store chain C&A, Clemens and August Brenninkmeyer, came from Mettingen.

36. The Hebrew strata in the Bargoens spoken in the Limburg towns of Roermond (Bakkes 2007) and Venlo (Coehorst 1983) appear to be more prominent. In both towns, Bargoens words have also penetrated into regular dialect use. For Venlo at least 18 of the 102 words in the list presented by Coehorst have Hebrew roots; today some 10 of these words (*gajes* ‘bad people’, *gappe* ‘to steal’, *jatte* ‘to steal’ [*<*Jiddish *jad* ‘hand’ *<* Hebrew; Van de Kamp & Van Wijk 2006: 242], *kasjewijle* ‘dead, gone’ [*<*Hebrew *hasiwienie* ‘lead us back’; Van de Kamp & Van Wijk 2006: 214], *lauw* (*loene*) ‘no way!, forget it!’, *mazzel* ‘good luck’, *peigere* ‘to die’, *pleite* ‘to disappear’, *schorem* ‘bad people’, *sjoeche(m)* ‘(no) idea’) are in general use in local colloquial Dutch. To judge from Bakkes (2007), the proportion of originally Hebrew items is very similar for Roermond. Here, too, a considerable part is in general use in present-day local colloquial Dutch, including *gajes* ‘bad people’, *gappe* ‘to steal’, *jatte* ‘to steal’, *jofel* ‘beautiful, pleasant’, *kasjewijle* ‘dead, gone’, *mazzel* ‘good luck’, *peigere* ‘to die’, *pleite* ‘to disappear’, *s(j)moeze* ‘to speak’.

The poorly illustrated Kempen traders' cant has influenced the *Tiöttensprache* strongly. These languages still had, despite this, 32 words in common, 12 of which were "Bargoens of the south-eastern group" (Moormann 1932: 110, our translation, NS&FH). "The similarity between Henese Fleck and *Tiöttensprache* is not as large as that between the Kempen traders' cant and *Tiöttensprache*. The Kempen traders' cant has influenced the *Tiöttensprache* strongly. We only have evidence for 9 words: *heet*, *quinten*, *meles*, *mol*, *murfen*, *nobes*, *poy*, *stiepen*, *zipken* or *ziemen* ['expensive', '(to) earn', belly/stomach, 'dead' (adj), 'mouth', 'no', 'thirst', 'give', 'yes']. These are again the oldest words, and all of them involve simplex forms." (Moormann 1932: 122, our translation NS&FH). With the exception of *murfen*, all these items also used to occur in GrB.

In terms of the size of the shared lexicon, GrB has more in common with both the Kempen traders' cant and *Tiöttensprache* than with Henese Fleck (Moormann 1932: 115,124; Goossens 1981: 557). This leads Moormann (1932: 124) to the tentative conclusion that the Kempen traders' cant seems to be the core of these trader's languages, with which GrB and *Tiöttensprache* are closely related. The question whether GrB was influenced directly or rather through contact with the Mettingen *Tiötten*, is not addressed in any of these studies.

"From all evidence it appears that the Groenstraat secret language was kept alive by a group of tramps, who were in direct contact with Bargoens speakers; this situation is very different from that in Mettingen (*Tiöttensprache*), Breyell (Henese Fleck), and the Kempen area, where it is extinct, or else only known by 'prominent citizens.'<sup>37</sup> This explains the Jewish and pure Bargoens elements in this language, which originated as a travelling traders' language. GrB is a kind of 'Bargoensed' traders' language. All the features of a travelling traders' language are still present, but in addition it has an imprint of genuine Bargoens" (Moormann 1932: 124, our translation NS&FH). The set of elements which are *Bargoens* in the strict sense includes e.g., *fonke*, *fonkerik* ('burn; heater, stove'), *fakkele* ('write'), *grannig* ('big; fat; wealthy; rich' etc.), *knul* ('lad'), *mieëles* ('stomach; belly'), *molle* ('kill'), *prieëmerik* ('chaplain'), *sjpanne* ('see, watch') and *buis*, *buize* ('drunk; drink'). Endepols (1924: 176) mentions GrB *prang*, *gruus*, *talfə*, *knieta*, *hens*, *loəts*, *grannig*, 'farmer', 'woman', 'to beg', 'to write', 'money', 'money', 'big; fat; rich', as indications that GrB is related to "the genuine thieves' cant" (our translation, NS&FH).

During the coal miners' phase in the life of GrB, elements from northwestern cryptolects further enriched the lexicon, such as [bɪkə], 'to eat', a new synonym of older GrB [fistə] and [ve:ʔyə].

Honnen (2000) sketches six different *Rotwelschdialekte*, all of them moribund, which used to be spoken in the western part of the German Rhineland, from

37. Moormann uses the word 'deftige', between quotation marks; it is not clear to us whether this is a case of irony or if there is more to it.

Speicher near Trier up to Breyell (Henese Fleck),<sup>38</sup> north of Mönchen-Gladbach (the others were spoken in Kofferen, Stotzheim, Neroth and Bell). All six secret languages developed during centuries of ambulant trade (household items such as pots and pans, mouse traps, sewing kits, drapery, food etc.), artisanry (the construction of outdoor ovens), and performances by small musical bands (‘Tanzkapellen’) which played at fairs and similar events. Like GrB, all six secret languages have larger or smaller originally Hebrew and Romani lexical strata and derivations with *-art*, *-arık*. In all cases the secret part of the vocabulary (largely content words) is integrated in the grammars of the respective local dialects.

The secret language spoken in Bell is the only one in which most of the deviant words are not traders’ or so-called “criminal” cant items in use in similar groups elsewhere, but rather prosodically altered variants of local dialect forms, as e.g. in the name of the language: *Lepper Talp*, which is derived from *Bell-er Platt*, ‘the Bell dialect’, in the case of *Beller* the derivational ending *-er* is disregarded, giving *Leb*, with final devoicing *Le[p]*. The language is also referred to as *Talp*, i.e. *Platt* reversed. This language thus resembles *Verlan*, by the same procedure derived from *l’Envers*, ‘the other way around, backwards’; in this cryptolect, the order of the syllables of a word is changed, as in e.g. *téci* (< French *cité*, ‘city’), *laisse béton* (< French *laisse tomber*, ‘leave it; forget it’), *Verlan* is now a youth language. The Serbo-Croatian equivalent *Šatrovački*, was originally a secret language of criminals, but presently it refers to a *Verlan*-like slang used by young people in cities like Belgrade, Zagreb and Sarajevo.

Ó hAodha (2002) describes ‘methods of disguise’ in the secret language of the Irish Travellers, known as *Shelta*, *Gammon* or *Cant*. These ‘methods’ include making Irish Gaelic words unrecognizable by e.g. switching segments, as in Ir. *cailín* > Sh. *laicín* ‘girl’, metathesis, as in Ir. *coinneal* > Sh. *niukal* ‘candle’, and adding *sr-* or *gr-*, as in Ir. *oiniún* > Sh. *grithiún* ‘onion’ (cf. Bakker 2006).

#### 4.2.1 *Combiverbs*

The GrB combiverb is *draiə*. The stem is *drai-*;<sup>39</sup> any conjugational and inflexional suffixes were attached to the stem. *draiə* is a weak verb, forming its past participle with the circumfix *ge-...-d*.

The Meertens website (databanken > Nederlandse Dialectenbank > Europa > Nederland > Limburg > Waubach) contains among other things the recording of a radio documentary on GrB, broadcast in the 1950’s. URL: <<http://www.meertens.knaw.nl/ndb/soundbites.php?p=Q117a>>. Part of the documentary is a staged dialogue, featuring three utterances with *draiə*:

38. See map.

39. Moormann notes this as *traí-* following voiceless obstruents. This is a normal feature of Dutch phonology – voicing assimilation in obstruent clusters. The rules in Limburgish are rather different but this is not of importance for this article.



(10) 15.08:

GrB. *Miene heech drei-t famus / fanus*

IMT. 1SG.POSS PERS.PRON CV-3S hunger

ET. 'I have hunger', 'I am hungry'

15.22:

GrB. *Sjoef dich, dat drei-t inne sjpitse*

IMT. be.quiet 2SG DEM.DIST CV-3S INDEF police-officer

ET. 'be quiet you, that is a police-officer (military policeman)'

15.59:

GrB. *Ken,<sup>40</sup> dat drei-t puuëk, Kesperke*

IMT. AFF DIST CV-3S fine Kasper-DIM

ET. 'Yes, that is fine, Kesperke'

This is scripted, but older and more authentic recordings probably do not exist.

*draiə* seems to have been in free variation with the local dialect equivalents, witness e.g. Endepols' (1924: 175) observation that e.g. *dər knöl van dər prang is noppəs* and *dər knöl van dər prang dräjt noppəs* ('the boy/son of the farmer is insane') were both possible.

The etymology of *draiə* is as yet unknown. In the local Groenstraat dialect there is no homophonous verb, nor any member of another word class whose phonological shape is in the slightest degree similar. The dialect equivalents of the verbs corresponding to standard Dutch *draaien*, *maaien*, *zaaien* ('turn', 'mow' and 'sow', respectively) all end in [i<sup>1</sup>ənə] and they are realized with an accute tone; *draiə* has a short vowel in the prominent syllable and therefore does not bear any tone. Moreover, 'draaien', 'maaien', 'zaaien' are lexical verbs.

*draiə* looks like the odd one out, a geographical hapax legomenon. However, we know from Willems (1838: 429) that it used to exist in the Kempen traders' cant (*Kramertaal van de Kempen*), spoken by the Teuten. There are other indications of borrowing from the Kempen traders' cants. These cryptolects have all more or less disappeared, so the only possible lead for further research would be older documents. Any documentation that might turn up might still make it possible to find out whether the combiverb *draiə* was imported into GrB from the Kempen traders' cants – either through direct borrowing, or indirectly via Rhineland traders' cants and/or the Mettingen Tiöttensprache for which very little sentential material seems to be available.

#### 4.2.2 Personal pronouns

The dialect form of the possessive pronoun is combined with a nominal stem, which varies for number. This is *hèèch* /he:²ç/ in the singular, and *hèèchte* /he:²çtə/ in the

---

40. *Ken* has Ashkenazic Hebrew roots.

plural. The form *tuuën* also occurs, with *tuuëne* in the plural. The first person singular pronoun *mechels* also occurs in pronominal and possessive functions.

(11) personal pronouns<sup>41</sup>

<i>miene hèèch</i>	minə he:²ç	1SG (all cases)
<i>miene tuuën</i>	minə tyʷə¹n	1SG (all cases)
<i>mechels</i>	mɛçəls	1SG (all cases)
<i>oos hèèchte</i>	ʊs he:²çtə	1PL (all cases)
<i>diene tuuën</i>	dinə tyʷə¹n	2SG (all cases)
<i>uch tuuëne</i>	ʏç tyʷə¹nə	2PL (all cases)
<i>urte hèèchte</i>	yʷə he:²çtə	2PL (all cases)
<i>ziene hèèch</i>	zinə he:²ç	3SG.MASC. (all cases)
<i>hunne hèèch</i>	hynə he:²ç	3SG.FEM. (all cases)
<i>hunne hèèchtə</i>	hynə he:²çtə	3PL (all cases)

With the exception of the 1st person singular, *mechels*, which can also be used in a possessive sense, possessivity is expressed in pronouns by means of phrasal expressions involving postposed PPs with the structure ‘of personal pronoun.’

(12) possessive pronouns

( <i>mechels</i>	mɛçəls	1SG)
<i>va ziene hèèch</i>	va zinə he:²ç	3SG

(13) a possessive example

GrB. dəʁ	al²pəzə	va	zinə	he:²ç
IMT. DEF.ART.M	father	OF	3S.POSS	PERS.PRON
ET.	‘the father of him; his father’			

#### 4.2.3 Negation

Like PC, GrB has a number of special negative forms.

- (14) *nulles* /nʏləs/ none  
*noëëbes* /nu¹:əbəs/ No! (negative reply)

#### 4.2.4 Numerals

According to Endepols (1924: 175) GrB did not have any replacive variants of numerals other than ‘2’. However Bosten et al. (1981) and Vek (2002) mention the forms in the list below.

41. The synthetic forms always consist of the local dialectal possessive form plus either of the fixed stems, *hèèch* or *tuuën*, in singular or plural forms.

(15)	ɔlf	one
	bɛis	two
	trɔms	three
	tɔləs	four
	ki <sup>1</sup> f	five
	wɔ: <sup>2</sup> f	six
	zɔlf	seven
	sjes	eight
	sɔpɔrkes	nine
	joes	ten

### 4.3 Overlap

Following a brief sketch of the handicraft and life style of the Pottefers, Van Nerum (ca. 1964) presents a concise lexicon and 20 sentences in the cryptolect. Some 30 of the 170 PC lexical items in Van Nerum's list also occur in GrB. A small part of the PC lexicon is recognizable as northwestern Bargoens types (e.g. *bink* 'man', *dokken* 'to pay', *knaak* 2½ guilder coin', *link* 'dangerous', *poen* 'money', *wout* 'policeman'). PC also contained originally French lexical material (e.g. *grandig* 'big', *fenêt* 'window', *travakment*, 'work, duty').

The lexical overlap between PC and GrB includes, despite small phonological or phonetic differences:

#### (16) Lexical overlap/similarities

Pottefers	Groenstraat	Dutch	English
bis	[bɛis]	twee	two
gieze(ke),	[ɣɔy <sup>1</sup> s],	vrouw(tje)	woman (affective)
griezeke	[ɣɔyskə]		
ploemp	[plun <sup>2</sup> f] <sup>42</sup>	water	water
spannen	[ʃpənə], [lu.ə <sup>1</sup> nzə]	kijken, opmerken	to look, notice
knul	[knvl]	jongen	boy
knuls	[knvlə]	jongens	boys
talfen	[tal <sup>ə</sup> fə]	bedelen	to beg
nosteren	[nɔstəvə] <sup>43</sup>	bidden	to pray
mol	[mɔ <sup>2</sup> l] <sup>44</sup>	dood	dead
grandig	[ɣvɑndiç]~[ɣvɑniç]	groot	big
sinus	[si <sup>1</sup> bəs]	ja	yes
kasperen	[kəspəvə]	de liefde bedrijven	to make love

42. Both likely onomatopoeitic.

43. From the name/beginning of the Lord's Prayer in Latin: Pater **noster**.

44. Romani (Hinskens 1985: 39–40).

In the following cases it is less clear whether PC and GrB overlap lexically:

(17) Borderline cases

Pottefers	Groenstraat	Dutch	English
batteren	[ˈbatəɾaf] <sup>45</sup>	vechten	to fight
zweterik	[ʒwatə] <sup>46</sup>	koffie	coffee
beschommelen	[bəʃul˚mə] <sup>47</sup>	betalen	to pay
bazen	[bœizə]	drinken	to drink
leimer	[li.ə˚m˚lʃ]	hemd	vest

The cases in (18) seem to be the result of semantic shift, either specialization or extension/generalization:

(18) Cases of semantic shift

Pottefers	Groenstraat	Dutch	English
kiet ‘jail’	[kicə~kitcə] ‘jail’	spec. keet	hut
respel ‘straw’	[ɾispəl] ‘hair’	exten. stro	straw

In the latter case the semantic extension has a metonymic flavor.

## 5. Morphology: Inflexion and derivation

### 5.1 Pottefers Cant

Compared to GrB, PC had a large number of derivations with the suffix /-əɾik/, such as *sniederik*, ‘knife’ (<*snie(d)e*, ‘to cut’), *staanderik*, ‘penis’ (<*staan*, ‘to stand up’), *mafferik*, ‘dumb person’ (<*maf*, ‘dumb; cuckoo’). Moreover, unlike most of the ambulant traders’ languages discussed above, PC appears to have had compounds, albeit only a few, such as *baasbink*, ‘drunkard’, and *laamkiet*, ‘bedroom’.

### 5.2 GrBargoens

Virtually the only distinction between GrB and the local dialect is in the lexicon, and more particularly content words: “the vocabulary consists of nouns, adjectives, adverbs and verbs”, the last being “nearly all weak” (Endepols p. 176, our translation, NS&FH). As in the case of the Pottefers Cant, GrB was embedded in the

45. GrB and dialect: ‘rogue, knave’.

46. Literally ‘the black one’.

47. Hebrew (Hinskens 1985: 37).

morpho-syntactic system of the local dialect, and hence the (verbal and nominal) inflexional endings do not differ from the dialectal ones; the GrB items were thus embedded in the matrix of the dialect grammar.

There were some GrB-specific derivational suffixes which did not occur in the dialect – or at least not productively. From the list of derivational suffixes presented by Van Hauwermeiren (2012) for a large set of secret languages which were/are spoken in the south of the Dutch language area, GrB only had /-əʃt/ (and its variants /-əʃ/ and /-ət/) and /-əʃk/. These suffixes served to derive nouns – often metonyms (including onomatopoeic expressions). Examples are:<sup>48,49</sup>

- |                                     |  |
|-------------------------------------|--|
| (19) /fi <sup>1</sup> -ək-əʃ/       | thief (from /fi <sup>1</sup> -əkə/ ‘(to) steal’)                               |
| /taləf-əʃ/,                         | beggar (from /tal <sup>ə</sup> fə/ ‘(to) beg’)                                 |
| /knap-əʃt/                          | gunpowder (probably onomatopoeic (cf. /knapə/, ‘to explode’)                   |
| /ki <sup>1</sup> -v-əʃt, ki-v-ət/   | coffee (dry) <sup>48</sup>   |
| /kwi-k-ət, kwi <sup>1</sup> -ək-ət/ | pig (from onomatopoeic verb /kwi-kə/ ‘to scream, screech’)                     |
| /ʃnu <sup>2</sup> :v-ət/            | boy ( <i>pars pro toto</i> , assuming the word refers to ‘nose’) <sup>49</sup> |
| /pri <sup>1</sup> :əm-əʃk/          | chaplain (from /pri <sup>1</sup> :əma/ ‘(to) preach’)                          |

Metonyms without a suffix were also used as nouns:

- |                        |   |
|------------------------|---|
| (20) /ʒwatə, ʃwatə/    | coffee (drink) (probably because of the colour) <sup>50</sup> |
| /plun: <sup>2</sup> ʃ/ | water (probably because of the sound) <sup>51</sup>           |
| /ʏxi <sup>1</sup> -s/  | Dutch gin (probably because of the colour)                    |
| /ʁɔ:ŋ <sup>2</sup> ə/  | moon (probably because of the shape: round)                   |

As was the case in Rotwelsch (Wolf 1956: 7), the grammatical gender of these nouns was generally identical to that of the dialect equivalent.

Until roughly the second quarter of the 20th century, word formation was restricted to derivation; compounding did not occur. Unlike Tiöttensprache (Stüve 1923: 15) and Henese Fleck, in Moormann’s (1932: 109–110, 115, 116) reconstruction of the early phase of Kempen traders’ cant, compounds were not very common. This is another property which the Kempen cryptolect and GrB had in common.

48. It is not clear from which item /ki<sup>1</sup>-vəʃt/ or /ki-vət/ was derived.

49. Compare this usage to the phrase *young sniveller* in English, sometimes used of ‘whining boys’.

50. Dutch *zwart*, ‘black.’

51. Compare Dutch onomatopoeia *plons*, ‘splash;’ cf. English *plunge*.

In the phase in which GrB served as a miner’s secret language, some neologisms were created with a morphological procedure of the first layer, viz. derivation with the suffixes /-əʃ/, /-ə(ʃ)t/ and /-əʃɪk/, which were productive. Examples are:

- (21) /plɛm:<sup>2</sup>p-əʃ/ smith  
 /knɛp-əʃt/ pistol (probably onomatopoeic (cf. /knɔpə/, above)  
 /ʃø:<sup>2</sup>k-ət/ tobacco (<Dutch *roken*, ‘to smoke’)  
 /di<sup>1</sup>.əm-əʃɪk/ evening<sup>52</sup>

As was mentioned above, in this phase compounds also started emerging, including e.g.:

- (22) /kla.uwədə-fʃe:ənz-əʃ/ stone-baker, brick-baker  
 from /klau:<sup>2</sup>wət/ egg; stone  
 + /fʃe<sup>1</sup>:ənzə/ (to) make  
 /ʃu<sup>1</sup>:əs-bɪk-əʃ/ lit. shit-eater (in the sense of ‘sycophant’)  
 from /ʃu<sup>1</sup>:əs/ fecal matter, shit  
 + /bɪkə/ (to) eat<sup>53</sup>

Another, telling example is [keijəs-kwi.<sup>ə</sup>stə], lit. ‘street speaking’, the talk of the (Groen)straat.<sup>54</sup> A remarkable formation is [bɪkə-mɔ:zə] < *bikken* + *manger*, which was created on the basis of northwestern Bargoens and PC [bɪkə] and French *manger*, both of which mean ‘(to) eat’. This is an instance of what Muysken (2016) has coined bilingual compound verbs.

Compounds are a distinguishing feature of the second lexical layer of GrB, as the first layer did not contain any compounds. “The younger formations are generally more transparent [than the older ones] and they sometimes look like facetious, funny word creations” (Endepols 1924: 178, our translation, NS&FH). Moormann quotes this claim and adds: “That is really characteristic of the development of the travelling salesmen’s languages” (1932: 123, our translation, NS&FH).

PC and GrB essentially did not have a separate syntactic module. With one exception (which we will discuss in the following sections), the grammar in the widest sense (hence including phonology, morphology and syntax) of an utterance in Pottefers Cant and GrB was that of the dialect in which the secret language’s items were embedded – as in Muysken’s (2001) ‘congruent lexicalisation’ scenario of bilingual speech.

52. Bosten et al. 1981: 34,103. Cf. archaic Dutch *deemster*, ‘dark; darkness’.

53. GrB of the second layer; the first layer has /fɪstə/ and /ve:<sup>ə</sup>ʃə/.

54. Dutch “straat” ‘street’.

## 6. The *have/be* combiverb

This section is concerned with a common feature of Bargoens, namely the tendency to have a single *have/be* verb. This feature is not however evenly distributed in its manifestations amongst the various forms of Bargoens in its occurrence, and in addition different usages of *have* and *be* are involved.

We have chosen to analyse the particular historical sentential material used here for two reasons. The first is that the sentential material available through these sources is mostly casual. That is to say, it was either offered by informants to illustrate the occurrence of particular lexical items, or else provided as illustrations of the daily life of Bargoens speakers. We hope that this makes these sentences more naturalistic. Our second reason is that the informants had (nearly) all been involved in the trades and/or other contexts in the course of which Bargoens had been used as a living linguistic code.

The material we examine here takes as its starting point material from the Pottefers Cant corpus, examined against the backdrop of those sources provided by the more complete edition of Moormann's 1930s work that appeared in 2002, edited by Nicoline van der Sijs. The main sources are, from oldest to newest:

### (23) Main Moormann material examined

Bron 9. Bargoens of Zele (1840)	[Moormann 2002: 374–391]
Bron 19. Bargoens van Roeselare (1890)	[Moormann 2002: 423–430]
Bron 28. Bargoens van Roeselare (1913)	[Moormann 2002: 460–461]
Bron 23. Bargoens van Maastricht (1917) <sup>55</sup>	[Moormann 2002: 438–444]
Bron 24. Bargoens van Groenstraat (1924)	[Moormann 2002: 444–458]
Bron 46. Henese Fleck (Breyell) (1926)	[Moormann 2002: 546–559]

The page numbers quoted below refer to these sources.

For reasons of insufficient data we restrict our examination to the following syntactic types:

### (24) syntactic types of *have/be* studied

1. Main “verbs”
  - a. *be*-copula:
    - i. adjectival complement
    - ii. nominal complement
    - iii. locative complement
  - b. *have*-lexical verb

---

55. Endepols (1924: 198) criticizes Moormann for his localization of this variant in Maastricht, where he encountered his informant. His opinion is that he must speak a Bargoens dialect from S. E. Limburg. Moormann (2002: 438) states that he was born in Hoensbroek, which is probably nearer the mark. We will therefore refer to this cryptolect as ‘Maastricht’ Bargoens.

2. Bound flexional elements:<sup>56</sup>
  - c. perfect:
    - i. “have”-perfect (*hebben*)
    - ii. “be”-perfect (*zijn*)
  - d. progressive: “*aan het* + infinitive
  - e. passive:
    - i. “be”-passive (present/past) (*worden*)
    - ii. “have been”-passive (perfect/pluperfect) (*zijn*)

We start with a direct comparison of Pottefers Cant and Groenstraat Bargoens (Comparisons 1, 2), and then turn to an examination of three other Bargoens cryptotects that are more closely related to Pottefers Cant (Comparisons 3, 4).

## 6.1 Comparison: “Main verb” usages (PC & GrB)

### 6.1.1 *Be-copula*

Firstly we will compare Pottefers Cant with GrB in terms of (25: 1a. i–iii), the *be-copula* usages.

Table 3. *Be-copula* usages (PC, GrB)

Construction	Pottefers C	Groenstraat B
Expression	<i>mozen</i>	<i>dräjen</i>
<i>Be-COP ADJ</i>	3/3	10/12
<i>Be-COP NP</i>	1/1	6/7
<i>Be-COP LOC-PP</i>	1/1	2/2
<i>Be-COP LOC-ADV</i>	1/1	3/4

The proportions in the table indicate the frequency the Bargoens form is used. In the remaining cases the local dialect form of the copula is employed.

In general the number of locative copula occurrences turns out to be very small, so that we have decided not to distinguish further these two syntactic categories, which are not semantically distinct anyway. Unfortunately Pottefers Cant does not provide us with very many examples. There is a clear distinction in the expression of the copula. Pottefers Cant uses a form *mozen*, while Groenstraat Bargoens uses a form *dräjen*. We will return later to the question of this difference. Table 3 now appears as Table 4.

56. By “bound” we mean “auxiliaries in construction with a main verb”. In other words, periphrastically bound.



Table 4. *Be*-copula usages (PC, GrB)

Construction	Pottefers C	Groenstraat B
Expression	<i>mozen</i>	<i>dräjen</i>
<i>Be</i> -COP ADJ	3/3	10/12
<i>Be</i> -COP NP	1/1	6/7
<i>Be</i> -COP LOC-PP/ADV	2/2	5/6
<i>Be</i> -COP total	6/6	21/25

We think that these results are broadly comparable.

We provide some illustrations in (25 ff.). These will be presented in a four-fold form: first an analysis of the Bargoens form, second an Interlinear Morpheme Translation (IMT) of this Bargoens form, third the Standard (southern) Dutch (SD) translation as given in the source, and fourth the English translation (ET). Bargoens and other secret language abbreviations are explicitedated in the examples. Further abbreviations met with in the text include GT (German translation). Page numbers are from Moormann (2002).

(25) *Be*-COP ADJ (Pottefers Cant)

PC. Dee gieze **moos-t**<sup>57</sup> liep ver poen  
 IMT. DIST woman CV-3S clever FOR money  
 SD. 'Die vrouw is slim voor geld'  
 ET. 'That woman is clever with money'

(26) *Be*-COP ADJ (Groenstraat B, p. 447)

GrB. dər<sup>58</sup> häch van dər prang **dräj-t** noppəs  
 IMT. DEF master<sup>59</sup> POSS DEF farmer CV-3S crazy  
 SD. 'De jongen van de boer is gek'  
 ET. 'The farmer's boy is mad'

(27) *Be*-COP NP (Pottefers Cant)

PC. Da **moos-t** ne bink dee tof kan travakk-en  
 IMT. DIST CV-3S INDEF man REL hard CAN work-INF  
 SD. 'Dat is een man die goed kan werken'  
 ET. 'That's a man that can work hard'

57. Only the verbal stems are highlighted: bold for relevant cryptolectal forms, italics for dialect forms.

58. Corrected from dən.

59. *häch* is translated both as 'master' M: 445 and 'boy' M: 445. We wonder if this is because the boy in question is a farmer's son, and therefore higher in status.

(28) *Be-COP NP* (Groenstraat B, p. 448)

GrB. dat trāj-[t] ənə puəkk-ə kanəs  
 IMT. DIST CV-3S INDEF nice-FL dog  
 SD. ‘Dat is een mooie hond’  
 ET. ‘That’s a nice dog’

(29) *Be-COP LOC* (Pottefers Cant)

PC. De gieze van mechels moos-t in de laam-kiet  
 IMT. DEF woman POSS 1S CV-3S IN DEF sleep-house  
 SD. ‘Mijn vrouw is in de slaapkamer’  
 ET. ‘My wife’s in the bedroom’

(30) *Be-COP LOC* (Groenstraat B, p. 449)

GrB. dər kabəs trāj-t in də sankš<sup>60</sup>  
 IMT. DEF priest CV-3S IN DEF church  
 SD. ‘De pastoor is in de kerk’  
 ET. ‘The priest is in the church’

### 6.1.2 *Have-lexical verb*

Now we will turn to the *have*-lexical verb comparison in Table 5.

Table 5. *Have-NP* usages (PC, GrB)

Construction	Pottefers C	Groenstraat B
Expression	<i>mozen</i>	<i>dräjən</i>
<i>Have-V NP</i>	3/3	9/14

This result is clearly different from that of Table 2. Once again Pottefers Cant appears to have consistently *mozen*, but Groenstraat Bargoens only has *dräjən* in about two-thirds of the cases. This raises the question whether different kinds of possession might affect the results.

One conceivable contrast might be between physical and non-physical possession. Non-physical possession would then relate to abstract properties. Such abstract properties have a close relation with predications as can be seen in (31).

(31) have fear = be frightened  
 have hunger = be hungry  
 have wealth = be wealthy

60. /səns/ from Latin *sanctus*, ‘holy’? According to van der Sijs (1995) *sjank*, *chanke* or *sankse*, ‘church’, has Romani origins and has a large geographical spread in the Dutch language area.

Seven out of eight examples with *dräjen* in Groenstraat Bargoens involve physical possession. An example would be (32).

- (32) Physical *have* (Groenstraat B, p. 451)
- GrB. də fem **dräj**-t ənə granig-ə sjpits  
 IMT. DEF woman CV-3S INDEF big-FL umbrella  
 SD. ‘De vrouw heeft een grote paraplu’  
 ET. ‘The woman has a big umbrella’

Four out of five examples with /hɑ:²n/ in Groenstraat Bargoens involve abstract properties, as in (33).

- (33) Abstract *have* (Groenstraat B, p. 447)
- GrB. mienə<sup>61</sup> häch *hat* poj  
 IMT. 1s.POSS boss have.3s thirst  
 SD. ‘ik heb dorst’  
 ET. ‘I have thirst/I am thirsty’

This distinction appears to be absent in Pottefers Cant. Here we have very few examples, but support can certainly be found in closely related cryptolects, like Zele Bargoens, as we will see below.

- (34) Physical *have* (Pottefers Cant)
- PC. Da mokke **moos**-t toffe loenker-s  
 IMT. DIST girl CV-3S good eye-PL  
 SD. ‘Dat meisje heeft schone (or: mooie) ogen’  
 ET. ‘That girl has beautiful eyes’
- (35) Abstract *have* (Pottefers Cant)
- PC. Da groem-eke **moos**-t grandig-e schrok  
 IMT. DIST child-DIM CV-3S big-FL hunger  
 SD. ‘Dat kindje heeft grote honger’  
 ET. ‘That child is very hungry’

## 6.2 Comparison 2: Periphrastic usages (PC & GrB)

### 6.2.1 *Active perfect*

First of all, we will compare the perfect verb forms. As in Standard Dutch, active verbs are divided into those with a *have*-perfect (*hebben*) and those with a *be*-perfect (*zijn*). We will look at the *have*-perfect. Although there is only a single

---

61. Corrected from mienən.

case in Pottefers Cant, this is supported by data from other similar cryptolects as we will see below. There are a number of cases from Groenstraat Bargoens.

- (36) *Have*-perfect (Pottefers Cant)
- PC. Mechels **moos-t** vinken zakk-en en bis knak-en  
 en draaiers bal-s **bedis-t**
- IMT. 1s CV-3S FIVE “hundred”-PL CNJ TWO “five”-PL  
 CNJ THREE “one”-PL **receive-PART**
- SD. ‘Ik heb 513 f. [5 honderd en 2 vijf en 3 één] **gekregen**’
- ET. ‘I have<sup>62</sup> **received** 5 hundred, 2 fives and 3 ones’

Note that since the Pottefers Cant has uniformly *mozen* for both *have* and *be*, we wouldn’t actually expect to see any difference here.

Groenstraat Bargoens does not seem to have *dräjen* in the case of either perfect type. We get the same distribution of forms as in Dutch, in dialect form.

- (37) *Have*-perfect (Groenstraat B, p. 450)
- GrB. dər pos’tələ-noeək hattə pomz-ə gə-wal-t
- IMT. DEF cook-man PERF potato-PL PART-cook-PART
- SD: ‘De kok heeft de aardappelen gekookt’
- ET. ‘The cook has cooked the potatoes’
- (38) *Be*-perfect (Groenstraat B, p. 453)
- GrB. dər poets is haas gə-sjpoor-t
- IMT. DEF policeman PERF away PART-gone-PART
- SD. ‘De veldwachter is weggegaan’
- ET. ‘The country policeman has gone away’

### 6.2.2 *Progressive*

Surprisingly, only the Pottefers Cant has evidence for this form.

- (39) *Progressive* (Pottefers Cant)
- PC. De bink **moos-t** an ’t baz-e in de baas-kiet
- IMT. DEF man CV-3S ON DEF **drink-INF** IN DEF drink-house
- SD. ‘De man is aan het **drinken** in het café’
- ET. ‘The man is **drinking** in the pub’

---

62. To simplify things we have consistently translated the Dutch perfect forms by English perfect forms, in spite of the fact that these would frequently be translated by English past tense forms.

### 6.2.3 *Passive*

The forms of the passive (*worden, zijn + past participle*) do not occur frequently in such codes. However, with the aid of other cryptolects it should be possible to reconstruct a partial paradigm. Our only example from these two cryptolects is from the Pottefers Cant. In this case we have a case of a *Perfect Passive*. No clear Groenstraat Bargoens examples occur.

- (40) *Perfect Passive* (Pottefers Cant)
- PC. We jall-en de bie, elles **moos-t e** schoep-ement ge-**flik-t**  
 IMT. 1P go-PL away PROX CV-3S INDEF steal-NMLZ PART-**do**-PART  
 SD. ‘We gaan weg, hier is een diefstal **begaan**’  
 ET. ‘We’re off, a robbery **has been committed** here’

### 6.3 Comparison 3: “Main verb” usages (PC & close relatives)

The close relatives of Pottefers Cant that will be considered here will be the Bargoens cryptolects of Roeselare, Zele, and Maastricht. The reason for this was simply the availability of a greater amount of data than that available from other places.

The Pottefers Cant was spoken in Tienen, located between Zele and Maastricht. The rough distances involved on an approximate W-E line are Zele 72 km E of Roeselare; Tienen 75 km ESE of Zele; and ‘Maastricht’ ca. 60 km E of Tienen. Only ‘Maastricht’ Bargoens<sup>63</sup> was spoken in the Netherlands.

#### 6.3.1 *Be-copula*

As can be seen from the table below, the four cryptolects are very similar, with the exception of Zele Bargoens which has a second *have/be* lexical item *modeeren*. We will not look into this fact further here.

Table 6. *Be-copula* usages (MB, PC, ZB, RB)

Construction	‘Maastricht’ B	Pottefers C	Zele B	Roeselare B
Expression	CV*	CV	CV	CV
<i>Be-COP ADJ</i>	16/20	3/3	26/26	6/7
<i>Be-COP NP</i>	5/7	1/1	8/8	5/5
<i>Be-COP LOC-PP/ADV</i>	2/3	2/2	6/6	5/6
<i>Be-COP total</i>	23/30	6/6	40/40	16/18

\* The hyphen, ‘-’, separates the stem from the infinitival marker.

63. Moormann’s informant for ‘Maastricht’ Bargoens, was actually born in Hoensbroek (Zuid-Limburg). For this reason we have noted the place-name in inverted commas.

We repeat the Pottefers Cant *Be*-COP ADJ Example (25) here as (41), and give examples from the other three cryptolects, to illustrate the parallels found in this case.

(41) *Be*-COP ADJ (Pottefers Cant)

PC. Dee gieze **moos-t** liep ver poen  
 IMT. DIST woman CV-3s clever FOR money  
 SD. ‘Die vrouw is slim voor geld’  
 ET. ‘That woman is clever with money’

(42) *Be*-COP ADJ (‘Maastricht’ B, p. 441)

MB. de mos **maas-t** grannig  
 IMT. DEF woman CV-3s big  
 SD. ‘de vrouw is zwanger’  
 ET. ‘The woman is pregnant’

(43) *Be*-COP ADJ (ZeLe B, p. 380)

ZB. ons geule **modeer-t** zu tof as d’ hulder  
 IMT. 1PP mouth COP-3S SO good AS DEF 3PP.POSS  
 SD. ‘Onze mond is zo goed als de hunne’  
 ET. ‘Our mouth is as good as theirs’

(44) *Be*-COP ADJ (ZeLe B, p. 390)

ZB. zijn vieke-noster-ik-sken **mois-t** zu<sup>64</sup> loens  
 IMT. 3SM life-pray-NOMZ-DIM CV-3s SO bad  
 SD. ‘Zijn levensboek is zo vuil’  
 ET. ‘His life-story is so evil’

(45) *Be*-COP ADJ (Roeselare B, p. 426)

RB. den being<sup>65</sup> **maas-t** buis  
 IMT. DEF man CV-3s drunk  
 SD: ‘Hij is dronken’  
 ET. ‘He’s drunk’

### 6.3.2 *Have-lexical verb*

The *have-V NP* cases only display a minor difference from the *be-copula* types. In ZeLe Bargoens we only see one form of *have/be* verb, *moizen*. *Modeeren* does not occur in this role. Once again in ‘Maastricht’ Bargoens we find the least consistency of usage.

64. The text has *zit* here. This is presumably due to a misreading of *zu*.

65. In Moormann (1932), this item is subject to spelling variation – which seems to indicate phonetic variation across the various relevant cryptolects.

Table 7. *Have*-verb usages (MB, PC, ZB, RB)

Construction	'Maastricht' B	Pottefers C	Zeke B	Roeselare B
Expression	CV	CV	CV	CV
<i>Have</i> -V NP	14/16	3/3	9/9	6/6

For our examples we repeat Pottefers Cant Example (34) as (46)

(46) *Have*-lexical verb (Pottefers Cant)

PC. Da mokke moos-t toffe loenker-s  
 IMT. DIST girl CV-3S good eye-PL  
 SD. 'Dat meisje heeft schone [or mooie] ogen'  
 ET. 'That girl has beautiful eyes'

(47) *Have*-lexical verb ('Maastricht' B, p. 441)

MB. er maas-t griekse-n in 't kemsel  
 IMT. 3S CV-3S louse-PL IN DEF shirt  
 SD. 'hij heeft luizen in 't hemd'  
 ET. 'He has lice in his shirt'

(48) *Have*-lexical verb (Zeke B, p. 387)

ZB. me moiz-en een marelse snaps en ne klonkoird roei  
 IMT. 1P CV-PL INDEF bottle gin CNJ INDEF glass beer  
 SD. 'We hebben een fles jenever en een glas bier'  
 ET. 'We have a bottle of gin and a glass of beer'

(49) *Have*-lexical verb (Roeselare B, p. 428)

RB. maas-t den kaffer grandige poen in zijn melis?  
 IMT. CV-3S DEF farmer big money IN 3SP pocket  
 SD. 'Heeft de boer veel geld op zak?'  
 ET. 'Has the farmer much money in his pocket'

## 6.4 Comparison 4: Periphrastic usages (PC & close relatives)

### 6.4.1 *Perfect*

As we have seen above if we find the single *have/be* word in the perfect we do not expect to find the Dutch (and German) distinction between the active *Have* and *Be*<sup>66</sup> perfects in the Bargoens cryptolects closely related to the Pottefers Cant. It is difficult to see patterns in the data because of gaps in incidence.

66. We will indicate the original Dutch/German perfect types as *Have* perfects and *Be* perfects.

For the Pottefers Cant we only had one example of an (active) perfect, in this case (36) a *Have* perfect, repeated here as (50).

- (50) *Have*-perfect (Pottefers Cant)
- |      |   |               |          |                     |     |     |           |
|------|---|---------------|----------|---------------------|-----|-----|-----------|
| PC.  | Mechels   | <b>moos-t</b> | vinken   | zakk-en             | en  | bis | knak-en   |
|      | en  | draaiers      | bal-s    | <b>bedis-t</b>      |     |     |           |
| IMT. | 1s  | CV-3S         | FIVE     | “hundred”-PL        | CNJ | TWO | “five”-PL |
|      | CNJ   | THREE         | “one”-PL | <b>receive-PART</b> |     |     |           |
| SD.  | ‘Ik <b>heb</b> 513 f. [5 honderd en 2 vijf en 3 één] <b>gekregen</b> ’      |               |          |                     |     |     |           |
| ET.  | ‘I <b>have</b> <sup>67</sup> <b>received</b> 5 hundred, 2 fives and 3 ones’ |               |          |                     |     |     |           |

Let us examine the closely-related cryptolects one-by-one, starting with ‘Maastricht’ Bargoens.

From ‘Maastricht’ we have eight examples of *Have*-perfects and none of *Be*-perfects. The majority of these *Have*-perfect cases use the auxiliary *hebben*, but a case with the combi-verb is illustrated in (53).

- (51) *Have*-perfect (‘Maastricht’ B, p. 439)
- |      |   |             |     |         |                      |
|------|---|-------------|-----|---------|----------------------|
| MB.  | Ik  | <i>heb</i>  | m’n | kluft   | <i>verpas-t</i> , .. |
| IMT. | 1s  | <i>PERF</i> | 1SP | clothes | <i>sell-PART</i>     |
| SD.  | ‘Ik <i>heb</i> kleren <i>verkocht</i> , ..’ |             |     |         |                      |
| ET.  | ‘I <i>have</i> sold my clothes, ..’         |             |     |         |                      |
- (52) *Have*-perfect (‘Maastricht’ B, p. 443)
- |      |   |               |       |                        |       |           |
|------|---|---------------|-------|------------------------|-------|-----------|
| MB.  | michels   | <b>maas-t</b> | kwant | ge- <b>pen-d</b>       | op de | rusjert   |
| IMT. | 1s  | CV-3S         | well  | <b>PART-sleep-PART</b> | ON    | DEF straw |
| SD.  | ‘ik <b>heb</b> lekker in ’t stro <b>geslapen</b> ’    |               |       |                        |       |           |
| ET.  | ‘I <b>have</b> <b>slept</b> comfortably in the straw’ |               |       |                        |       |           |

Turning to Zele Bargoens, we may just possibly see a distinction between *Have*-perfects and *Be*-perfects. Unfortunately there is only one case of a *Be*-perfect. Of the eleven cases of *Have*-perfects, all employ dialect forms of Dutch *hebben*.

- (53) *Be*-perfect (Zele B, p. 379)
- |      |  |               |      |     |         |        |              |                       |
|------|--|---------------|------|-----|---------|--------|--------------|-----------------------|
| ZB.  | ge   | <b>mois-t</b> | me   | ou  | fiëm-en | in een | fondoesje    | ge- <b>bommel-t</b>   |
| IMT. | 2s   | CV-2S         | WITH | 2SP | hand-PL | IN     | INDEF window | <b>PART-fall-PART</b> |
| SD.  | ‘Je <b>bent</b> met je hand in een ruit <b>gevallen</b> ’        |               |      |     |         |        |              |                       |
| ET.  | ‘You <b>have</b> <b>fallen</b> with your hands through a window’ |               |      |     |         |        |              |                       |

67. To simplify things we have translated the Dutch perfect forms with English perfect forms, in spite of the fact that these would frequently be translated by English past tense forms.



- (54) *Have*-perfect (Zeile B, p. 383)
- ZB. die knul *heed* een korante *verpas*-t  
 IMT dist fellow *have*.3s indef cow *sell*-part  
 SD. 'die man *heeft* een koe *verkocht*'  
 ET. 'that man *has sold* a cow'

finally we look at Roeselare Bargoens, where again we only find *have*-perfects. here however we only come across cases of the *hav e/be* word.

- (55) *have*-perfect (Roeselare B, p. 429)
- rb. mechels **maas**-t geen tand **ge-zop**-en  
 imt 1s CV-3S NEG.Q tooth PART-**win**-PART  
 SD. 'Ik **heb** volstrekt niets **gewonnen**  
 ET. 'I **have won** absolutely nothing

Now we look at the results in tabular form.

**Table 8.** Active perfect usages (MB, PC, ZB, RT)

Construction	'Maastricht' B	Pottefers C	Zeile B	Roeselare B
Expression	CV	CV	CV	CV
<i>Have</i> -perfect	2/8	1/1	0/11	5/5
<i>Be</i> -perfect	–	–	1/1	–

The strongest tendencies visible for Active Perfects are the following:

- (56) Fairly strong tendencies among *Perfects*
- 'Maastricht' B: Most *Have*-perfects do **not** use *combiverbs*.  
 Zeile B: **No** *Have*-perfects use *combiverbs*.  
 Roeselare B: All *Have*-perfects use *combiverbs*.

In contrast:

- (57) Very weak tendencies among *Perfects*
- 'Maastricht' B: Some *Have*-perfects use *combiverbs*.  
 Pottefers Cant: The only example of a *Have*-perfect uses a *combiverb*.  
 Zeile B: The only example of a *Be*-perfect uses a *combiverb*.

Possible implications might be:

- (58) Possible implications for *Perfects*
- Pottefers Cant: *Be*-perfects would be expected to use *mazen*.  
 Zeile B: *Be*-perfects would be expected to use *mazen*.  
 Roeselare B: *Be*-perfects would be expected to use *mazen*.

A hypothesis as to the expression of perfects might then be the following:

**Table 9.** Hypothetical Active Perfect usage (MB, PC, ZB, RB)

Construction	‘Maastricht’ B	Pottefers C	Zele B	Roeselare B
<i>Have</i> -perfect	<i>hebben</i> > CV	CV	<i>hebben</i>	CV
<i>Be</i> -perfect	??	CV	CV	CV

We will now turn to passives.

#### 6.4.2 *Passives*

The notion of active perfects using *Have* and *Be* as components of the expression of the active perfect in West Germanic languages is probably familiar to most linguists. Even Modern English has the odd remnant of this, as in “they’re gone,” equivalent to “they’ve gone.” However the expression of the passive tenses might be less familiar. To remedy this we’ll give a quick overview of the system in Standard Dutch, using 3rd person singular forms. We take as exemplary for the Aux(iliary) case the future passive *zal*. Note that the *simple* passive “auxiliary” *worden*, when used in isolation, means ‘become,’ while the *perfect* passive “auxiliary” *zijn* is ‘be’ in isolation.<sup>68</sup>

#### (59) A schema of the Dutch Passive

##### *Simple Passive Paradigm:*

Present:	SUBJECT	<i>wordt</i>	<i>past-participle</i>	‘he is X-ed’
Past:	SUBJECT	<i>werd</i>	<i>past-participle</i>	‘he was X-ed’
Present Aux:	SUBJECT	<i>zal worden</i>	<i>past-participle</i>	‘he’ll be X-ed’
Past Aux:	SUBJECT	<i>zou worden</i>	<i>past-participle</i>	‘he’d be X-ed’

##### *Perfect Passive Paradigm:*

Pres. Perfect	SUBJECT	<i>is</i>	<i>past-participle</i>	‘he has been X-ed’
Past Perfect	SUBJECT	<i>was</i>	<i>past-participle</i>	‘he had been X-ed’
Pres. Aux Pf.	SUBJECT	<i>zal zijn</i>	<i>past-participle</i>	‘he’ll have been X-ed’
Past Aux. Pf.	subject	<i>zou zijn</i>	<i>past-participle</i>	‘he’d have been X-ed’

Unfortunately we have only a single example of a *Present Paradigm* form. This is a *Simple Passive Present Aux* form from Roeselare Bargoens.

68. Modern Standard Dutch has generally dropped the use of *geworden* following the past participle in the perfect paradigm, which is regarded as archaic.

- (60) *Simple Present (Aux)* (Roeselare B, p. 425)
- RB. je **zul-t** **befazel-d** **maz-en**  
 IMT. 2s FUT-2s **trick-part** CV-inf  
 SD. ‘gij **zult** **bedrogen worden**’  
 ET. ‘You’ll be **tricked**’

However, comparing this Roeselare Bargoens sentence with a Zele Bargoens sentence which receives an interpretation as a *Perfect Passive Past Aux* sentence, we immediately observe a problem.

- (61) *Perfect Passive (Past Aux)* (Zele B, p. 387)
- ZB. het **zou** agâ deur ons stroet-e **ge-ris-t** **modeeren**  
 IMT. 3SN FUT-PST soon thru 1PP throat-PL part-wash-part CV-inf  
 SD. ‘Het **zou** gauw door onze keel **gespoeld zijn**’  
 ET. ‘It **would** soon **have been** washed down our throats’

If the Bargoens translations of *worden* and *zijn* both involve the unitary *have/be* combiverb, then these Bargoens cryptolects would not have been able to distinguish the Simple Passive and Perfect Passive paradigms from each other.

By chance all the other passive sentences we have receive a “perfect-type” translation in Dutch, i.e. the sentences receive a translation with *zijn* and not with *worden*. These are available from Pottefers Cant, Zele Bargoens and Roeselare Bargoens. However, it is also the case that Perfect Passives may refer to the *results* ensuing from actions rather than the *actions* themselves. Additionally it may be the case that deverbal adjectives are homophonous with the resultative participles, or even that it is difficult to tell the difference. (61) is clearly a Perfect “action passive.”

Two other cases are clearly also action passives. (62) is repeated from (40).

- (62) *Passive Perfect (Present)* (Pottefers Cant)
- PC. .., elles **moos-t** e schoep-ement **ge-fliik-t**  
 IMT. .., HERE CV-3S INDEF steal-NMLZ PART-**do**-PART  
 SD. ‘.., hier **is** een diefstal **begaan**’  
 ET. ‘... , a robbery **has taken place** here’
- (63) *Passive Perfect (Present)* (Roeselare B, p. 429)
- RB. Den being **maas-t** **ge-schoep-t**  
 IMT. DEF man CV-3S PART-**catch**-PART  
 SD. ‘De man **is gesnapt**’  
 ET. ‘The man **has been caught**’

(64) clearly has a resultative interpretation. The bacon “has been fried” and as a result can be described as being in a “fried” (or “done”) state.

(64) *Passive Perfect* (Present) (Zelev B, p. 379)

ZB. de pitjoukri mois-t ge-flik-t  
 IMT. DEF bacon CV-3S PART-do-PART  
 SD. ‘t spek is gebakken’  
 ET. ‘The bacon is done’<sup>69</sup>

A final case is probably, but not certainly, a case of a deverbal adjective.

(65) *Passive Perfect* (Present) (Zelev B, p. 382)

ZB. kiewerrick-en, modeer-t dat ge-permenteer-d-Ø  
 IMT friend-pl, CV-3s dist part-permit-part-adjz  
 SD. ‘vrienden, is dat toegelaten’  
 ET. ‘friends, is that permitted’

On the assumption that the two passive paradigms of Dutch are not distinguished in Bargoens (see above sentences (60), (61)), this gives us the following distribution of forms.

Table 10. Passive usage (mb, pc, zb, rb)

Construction	‘Maastricht’ b	Pottefers c	Zelev b	Roeselare b
expression	??	CV	CV	CV
<i>Passive</i>	??	1/1	5/5*	3/3

\* These include two examples of the “fixed expression,” poss *lichterikken mazen/moizen gekwikt* ‘poss days are numbered’ which we interpret as an archaic passive.

Combining this table with the Active Perfect gives us the following table of periphrastic forms.

Table 11. Combined periphrastic usage (MB, PC, ZB, RB)

Construction	‘Maastricht’ B	Pottefers Cl	Zelev B	Roeselare B
<i>Have-perfect</i>	<i>hebben</i> > CV	CV	<i>hebben</i>	CV
<i>Be-perfect</i>	??	CV	CV	CV
<i>Passive</i>	??	CV	CV	CV

69. As can be seen from the present example, the normal English usage with states and results is with the present tense form *is*.

## 6.5 Bargoens versus Traders' cants (Kramertalen)

In our first comparisons between Pottefers Cant and Groenstraat Bargoens we noticed a substantial difference between the two. Pottefers Cant, and the other Bargoens cryptolects we have just compared it with, showed a widespread tendency to use a unitary *have/be* combiverb-form for all “have/be” forms whether “main verbs” or “auxiliary verbs” used periphrastically, while Groenstraat Bargoens only displayed “main verb” usages of its *have/be* verb. Secondly, the form employed was different. Pottefers and its close relatives used *mozen*, *mazen* and other variants of what is clearly the same root while Groenstraat Bargoens used *dräjə*.

### 6.5.1 A Kempen parallel

The use of this latter form is only recorded elsewhere in the Kempen area in the Belgium-Netherlands borderland, as *dra(e)yen*. Moormann does not mention the following sentences from the Kempen traders' cant (KTC) provided by Willems (1838: 427–431), listing only the lexical items. Willems does not provide a precise localization, mentioning only Lommel, Exel,<sup>70</sup> Kleinbreugel and St-Hubrechtsselle as definite locations where this ambulant traders' cant was used.<sup>71</sup> Moormann refers to this source as *Kramertaal van de Kempen* (KK).

This is not a spontaneous text, being a translation of that passage beloved by 19th century amateur-linguists, the *Parable of the Prodigal Son*.<sup>72</sup> We quote all the forms that seem to us to involve the same item that occurs as the unitary *have/be* verb in Groenstraat Bargoens. The examples presented are limited to a much more restricted context than in Groenstraat Bargoens, however. The only context in which it occurs is in *Be-COP ADJ*, and even then inconsistently. We have no case of *draey* used with the meaning ‘have’. We note the few cases we have as *CV*?

The translation appears to be rather free. We will supply two English translations, the first being a translation the Standard Dutch version, and the second our guess as to the actual translation of the Kempen version.

Table 12. Occurrences of *dra(e)yen*

Construction	Kempen TC
Expression	<i>dra(e)yen</i>
<i>Be-COP ADJ</i>	5/11

70. We have indicated the location of Exel on the map on p. 4 of the article. Exel is now Eksel, Belgium.

71. Three sentences are also recorded from Stramproy (Moormann 1930: 56–57; Creemers 1871: 115) but unfortunately the available material is not relevant for our purposes.

72. Comparable to the 20th century text “The North Wind and the Sun”.

- (66) *Be-COP ADJ* (Kempen TC, p. 429)<sup>73</sup>  
 KK. ik *heb* loens ge-**draey**-t tegen den theuw-en diennes ....  
 IMT. 1s *PERF* bad part-CV?-part against DEF good-F heaven ...  
 SD. ‘ik heb gezondigd tegen de hemel ...’  
 ET1. ‘I have sinned against heaven ...’ [translation of SD]  
 ET2. ‘I *have been* bad towards good heaven ...’ [translation of IMT]
- (67) *Be-COP ADJ* (Kempen TC, p. 430)  
 KK. ... en *heb* (ik) u nooit loens ge-**draey**-t  
 IMT. ... CNJ PERF (1S) 2S NEVER bad part-CV?-part  
 SD. ‘... en nooit ben ik u ongehoorzaam geweest’  
 ET1. ‘... and I’ve never been disobedient to you’  
 ET2. ‘... and I’ve never **been** bad to you’
- (68) *Be-COP ADJ* (Kempen TC, p. 430)  
 KK. ... en ik *zal* uwe fuyk alles theuw **draij**-en  
 IMT. ... CNJ 1S *FUT* 2SP PRON everything good cv?-inf  
 SD. ‘... en alles wat van mij is, is van jou’  
 ET1. ‘... and everything that is mine, is yours’  
 ET2. ‘... and I *will be* good to you’

The words *theuw* and *loens* are clearly adjectives meaning ‘good’ and ‘bad’ respectively. Look at the following sentences to see this.

- (69) *Be-COP ADJ* (Kempen TC, p. 429)  
 KK. Ik *ben niet theuw genoeg* uwe knulle ge-noem-d te word-en  
 IMT. 1S BE NEG **good** enough 2SP boy PART-call-PART INF pass-INF  
 SD. ‘ik ben het niet meer waard uw zoon genoemd te worden’  
 ET1. ‘I’m no longer worthy to be called your son’  
 ET2. ‘I’m *not good enough* to be called your son’
- (70) *Be-COP ADJ* (Kempen TC, p. 429)  
 KK. ik *ben te loens* om uwe knul ge-noem-d te word-en  
 IMT. 1S BE TOO **bad** CMPL 2SP boy PART-call-PART INF PASS-INF  
 SD. ‘ik ben het niet meer waard uw zoon genoemd te worden’  
 ET1. ‘I’m no longer worthy to be called your son’  
 ET2. ‘I’m *too bad* to be called your son’

73. Another sentence is virtually identical to (66): Ik **heb** loens **gedrayt** tegen den theuwen diennes en tegen uwe fuyk.

Very interesting here is the fact that *drayen* in the Kempen traders' cant takes a *have*-perfect rather than a *be*-perfect. Unfortunately there is no parallel example from Groenstraat Bargoens.

### 6.5.2 *Henese Fleck* (Breyell, left-bank of lower Rhine, Germany)

Breyell is only 5km from the Dutch border, near Tegelen in north central Limburg. Henese Fleck is also classified by Moormann under Kramertalen (Traders' cants). Here we find a third verb *hucke*, restricted to *be* "main verb" usages. All three types we have distinguished occur nearly universally.

(71) *Be-COP ADJ* (Henese Fleck, p. 554 (M))

HF. minotes **huck-t** mörrig  
 IMT. 1s BE-3S tired  
 GT. 'Ich **bin** müde'  
 ET. 'I'm tired'

(72) *Be-COP NP* (Henese Fleck, p. 553 (M))

HF. **Huck-t** dot zinotes Limthürken?  
 IMT. BE-3S DIST 2SP darling  
 GT. 'Ist das dein Liebchen?'  
 ET. 'Is that your darling?'

(73) *Be-COP LOC* (Henese Fleck, p. 557 (M))

HF. Dem Benk **huck-e-t** bei de Ville-pretter op dem Bott,  
 IMT. DEF man BE-PST-3S AT DEF town-preacher ON DEF food  
 GT: 'Der Mann war beim Bürgermeister zu Mittag'  
 ET. 'The man was with the mayor for dinner'

This word appears to be related to meanings such as Southern German *hocken* 'sit, seat oneself' (Duden Online).

Some stray examples of *hucken/hocken* appear also to occur in Bargoens cryptolects, for instance (75)–(78). The meaning seems to be restricted to a locative copular one, however.<sup>74</sup>

74. GrB also has a verb <hoeme>, /humə/, with a closely related meaning ('live, reside, stay; sit; sojourn, be located') and a partially similar form. The phonological difference suggests that the two verbs are not etymologically related; /humə/ may result from folk etymology (after /he<sup>2</sup>m/, 'home' adv.). Moreover, Endepols (1924: 192) also mentions <hoeəstə> with the very same meanings, and adds the example *də sjpitsə hoeəstə in də siep*, 'the police-officers are sitting in the inn.'

(74) *Be-COP LOC* (‘Maastricht’ Bargoens, p. 438)

MB. Was dat z'n mos, die efkes hier **huk-te?**

IMT. BE.PST DIST 3SP wife, REL briefly PROX.LOC BE.LOC-PST

SD. ‘Was dat zijn vrouw, die zo-even hier was?’

ET. ‘Was that his wife who was here just now?’

(75) *Be-COP LOC* (Groenstraat Bargoens, p. 452)

GrB. dao **hoeæk**<sup>75</sup> ə fikskə

IMT. DIST.LOC BE.LOC INDEF policeman

SD. ‘daar is een politieagent’

ET. ‘There’s a policeman’

Related is no doubt the following:

(76) *Be-COP LOC?* (Groenstraat Bargoens, p. 454)

GrB. mienə<sup>76</sup> häch hat ach daag mot-ə **hoeæk-ə**

IMT. 1SP PRON HAVE.PST EIGHT day MUST-INF stay-inf

SD. ‘Ik heb acht dagen moeten **verblijven**<sup>77</sup> (in het gevangenis)’

ET. ‘I had to **stay** in prison for eight days’

And in a similar usage to (76):

(77) *Be-Cop Loc?* (‘Maastricht’ Bargoens, p. 441)

GrB. er maast ge-gannef-t en **hok-t** in bajis

IMT. 3SP CV-3S PART-steal-PART CNJ BE.LOC-3S in prison

SD. ‘Hij heeft gestolen en **zit** nu in de gevangenis’

ET. ‘He has stolen and is now in prison’

## 7. Conclusion on *have/be* verbs

What can we conclude from the foregoing? If we attempt to describe the occurrence of replacive forms of *have/be* in a tabular form, we get the following picture. We abstract here from minor variations of orthography, and in the case of Zele Bargoens from a second *have/be* verb *modeeren*. The table reveals that the Pottefers Cant, spoken in Tienen in Belgium, forms part of a complex of similar cryptolects extending across the North of Belgium and the South of the Netherlands.

75. [huæk] from /huæk-t/ after word-final [t] deletion, which is a highly productive postlexical process in Limburg and Ripuarian dialects.

76. Corrected from [mienən] (Moormann 2002: 454).

77. Dutch *zitten* is a frequently used slang word for ‘be in prison.’



Table 13. Summing up

Construction	Pottefers	Roeselaare	Zele	'Maastricht'	Groenstraat	Kempen	HeneseFleck
Expression	<i>mozen</i>	<i>mazen</i>	<i>mazen</i> <i>modeeren</i>	<i>mazen</i>	<i>dräjén</i>	<i>drayen</i>	<i>hucken</i>
Date	1964	1890/1913	1840	1917	1924	1838	1926
<i>be</i> -COP ADJ	<i>mozen</i>	<i>mazen</i>	<i>mazen</i> <i>modeeren</i>	<i>mazen</i>	<i>dräjén</i>	<i>drayen</i>	<i>hucken</i>
<i>be</i> -COP NP	<i>mozen</i>	<i>mazen</i>	<i>mazen</i> <i>modeeren</i>	<i>mazen</i>	<i>dräjén</i>		<i>hucken</i>
<i>be</i> -COP LOC	<i>mozen</i>	<i>mazen</i>	<i>mazen</i> <i>modeeren</i>	<i>mazen</i>	<i>dräjén</i>		<i>hucken</i>
<i>have</i> -V NP	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjén</i>		
<i>have</i> -PERF	<i>mozen</i>	<i>mazen</i>		( <i>mazen</i> )			
<i>be</i> -PERF	<i>mozen</i>	??	<i>mazen</i>	??			
PROG	<i>mozen</i> <i>an't</i>	??	??	??			
PASSIVE	<i>mozen</i>	<i>mazen</i>	<i>mazen</i> <i>modeeren</i>	??			

Groenstraat Bargoens is a borderline case. On the one hand, it has (or had, strictly speaking) a unitary *have/be* form, restricted to “main Verb” occurrences. It does not however display the extension of this to *have/be* used in auxiliary functions.

We hardly know enough about the morphosyntax of the Kempen traders' language to say much at all, except of course that it employed the same form *draeyen* as in Groenstraat Bargoens, and to state that it was used in a more restricted fashion would be unsupported by the present evidence.

Henese Fleck does not have a unitary *have/be* form at all, merely adopting a replacive form for the copular uses of *be*.

## 7.1 The source of *have/be*

The existence of a unified *have/be* combiverb in the secret languages of the Low Countries<sup>78</sup> is a bit of a mystery. We will attempt to provide an answer to the question of where this odd feature might have come from in Section 8.

78. As far as we know it is not used in the more northerly regions.

It is well-known that numerous languages express possession by means of a copular expression. For instance in Scottish Gaelic (SG) the verb used is *tha*,<sup>79</sup> the Gaelic copula.

- (78) SG. *tha Eachann beag*  
 IMT. cv Hugh small  
 ET. ‘Hugh is small’

Possession is expressed in Gaelic by the copula and the preposition *aig* ‘at’.

- (79) SG. *tha tigh aig Mairi*  
 IMT. cv house at Mary  
 ET. ‘Mary has a house’

In the Bargoens cryptolects we have according to Moormann (2002: 149–250) many words of Hebrew origin. And Hebrew also expresses possession by means of the copula. However these Hebrew words entered Bargoens via Judeo-German and/or Yiddish, and Yiddish normally expresses possession with *hop* ‘have’.

Another problem for any hypothesis involving Hebrew influence is the fact that the proportion of Hebrew words is highest in the north of the Netherlands – up to 40% – where the unitary verb *mazen* is not recorded, and lowest in the south where *mazen* is used. So we have to look elsewhere for the source of this feature.

## 8. Romani as the source of *have/be* verbs?

The Sinti variety of Romani is another possibility.<sup>80</sup> Although the number of Romani lexical items in Bargoens is in most cases low to very low, there is a small group of words of ultimate Romani origin such as *bink* ‘man’, *gies* ‘woman’ (Van der Sijs 1995; Moormann 2002: 287). Moormann interprets these as very early borrowings, and suggests that they were borrowed ca. 1500 from the Sinti-Romani-speakers by the Dutch caravan-dwellers<sup>81</sup> and ambulant traders encountered in the Eastern Netherlands at that time.

Lexical borrowings are of course more easily replaced in secret languages. What we are proposing as a hypothesis here is a case of *morphosyntactic* influence from Sinti-Romani, which we would suggest, could well have represented a longer-lasting phenomenon than mere lexical influences. We give a range of illustrative sentences

---

79. Pronounced [ha].

80. Sinti is “German Romani.” It has been spoken in the Netherlands since the 15th century.

81. See (1b).

from the western Sinti dialect from Sowa (1898), and Finck (1903), to illustrate the parallels. Note that the present tense paradigm of the copula is as follows:

(80) Paradigm of the Sinti-Romani present copula

1SG	<i>hom</i>	1PL	<i>ham</i>
2SG	<i>hal</i>	2PL	<i>han</i>
3SG/PL	<i>hi</i>		

## 8.1 Parallels between Bargoens and Sinti-Romani

### 8.1.1

Firstly we demonstrate *copula* usage in Sinti-Romani (82)–(86). We provide examples of the three copular types:

- (81) i. adjectival complement  
 ii. nominal complement  
 iii. locative complement

(82) *Be-COP ADJ* (Finck 1903: p. 35)

SN.	Jow	<b>hi</b>	múlo
IMT.	3s	CV.3	dead
GT.	‘Er <b>ist</b> tot.’		
ET.	‘He <b>is</b> dead.’		

(83) *Be-COP ADJ* (Sowa 1898: p. 9)

SN.	O	boläpen	<b>hi</b>	bāro
IMT.	DEF	heaven	CV.3	high
GT.	‘Der Himmel <b>ist</b> hoch.’			
ET.	‘Heaven <b>is</b> high.’			

(84) *Be-COP NP* (Finck 1903: p. 35)

SN.	Mir-i	tsáwo	<b>hi</b>	baš-əp-as-kər-(i)-ítsa
IMT.	1SP-F	daughter	CV.3	music-ABSTR-OBL-GEN-FEM](=fiddle)-FEM
GT.	‘Meine Tochter <b>ist</b> Musikantin’			
ET.	‘My daughter <b>is</b> a musician’			

(85) *Be-COP NP* (Sowa 1898: p. 16)

SN.	Me	<b>hom</b>	je	čač-o	kālo
IMT.	1S.NOM	CV.1S	INDEF	real-m	Rom
GT.	‘Ich <b>bin</b> ein echter Zigeuner’				
ET.	‘I’m a real Rom’				

- (86) *Be-COP LOC* (Sowa 1898: p. 54)  
 SN. Mər murš hi ano stil-əpen  
 IMT. 1SPM husband CV.3 ON imprison-ABSTR  
 GT. ‘Mein Mann ist im Gefängnis’  
 ET. ‘My husband’s in prison’

In order to indicate possession an *oblique* stem-form may be used with the so-called “logical subject”. This form has an *object* meaning in transitive sentences.

- (87) *Have-V NP* (Finck 1903: p. 17)  
 SN. Man hi o bār-o līl  
 IMT. 1S.OBL CV.3 DEF.M big-M letter  
 GT. ‘Mir ist der grosse brief’ [literal]  
 GT. ‘Ich habe einen Gewerbeschein’  
 ET. ‘At me is the big letter’ [literal]  
 ET. ‘I have a permit’

Alternatively, the “logical subject” may appear in the so-called prepositional or locative<sup>82</sup> case.

- (88) *Have-V NP* (Sowa 1898: p. 23)  
 SN. Man-de hi dār akola kov-es-tər  
 IMT. 1S.OBL-PREP CV.3 fear PROX.OBL.M thing-OBL-ABL  
 GT. ‘Ich habe Scheu vor dieser Sache’  
 ET. ‘At me is fear from this thing [literal]’  
 ET. ‘I have fear of this thing’

Of course Bargoens has no case-endings in nouns, unlike Romani, which results in the parallellism we think exists. All case-marking is removed, as it were. Possession is indicated in Romani on the noun preceding the copula by giving it an oblique or prepositional case-marking. In Bargoens the cases are no longer marked, but the word-order remains the same.

### 8.1.2 *Perfect passive*

The so-called perfect passive is formed periphrastically with the copula<sup>83</sup> + past participle, and agrees with the subject in gender. Semantically, it can slip over into a more adjectival resultative meaning.

82. See Matras (2004: 88)

83. Both Dutch and Sinti also have periphrastic present passives. These employ other verb stems.

(89) *Preterit Passive*

(Finck 1903: p. 38)

- SN. kówa mas hi kér-ə[v]-do  
 IMT. PROX.M meat CV.3 cook-CAUS-PART  
 GT1. '..., dieses Fleisch is gekocht worden' or  
 GT2. '..., dieses Fliesch ist gar'  
 ET1. '..., this meat was cooked' or  
 ET2. '..., this meat is ready'

This semantic ambiguity reminds us of cases (64), (65).

8.1.3 *Conclusion*

We conclude that the Bargoens variants (including older Groenstraat Bargoens) illustrate probable Romani influence in their use of *mazen etc./dräjen* for possession – the *have-V NP* row in Table 14 below.

Table 14. Sinti-Romani – Bargoens parallels with *have/be* combiverbs

Construction	Dutch	Romani	Pottefers	C Roeselaare	Zelev	Groenstraat	Henese Fleck
Expression	<i>zijn</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjen</i>	<i>hucken</i>
<i>be-COP ADJ</i>	<i>zijn</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjen</i>	<i>hucken</i>
<i>be-COP NP</i>	<i>zijn</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjen</i>	<i>hucken</i>
<i>be-COP LOC</i>	<i>zijn</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjen</i>	<i>hucken</i>
<i>have-V NP</i>	<i>hebben</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>	<i>dräjen</i>	
<i>have-PERF*</i>	<i>hebben</i>		<i>mozen</i>	<i>mazen</i>			
<i>be-PERF</i>	<i>zijn</i>		<i>mozen</i>	??	<i>mazen</i>		
PROG	<i>zijn aan 't</i>		<i>mozen an't</i>	??	??		
PERF PASS**	<i>zijn</i>	<i>hi</i>	<i>mozen</i>	<i>mazen</i>	<i>mazen</i>		

\* The preterite/perfect distinction in Dutch is not comparable to the preterite in Sinti. Sinti lacks an explicit active perfect.

\*\* The periphrastic passive with a past participle is the only point of comparison available.

The first examples of Sinti-Romani in the Netherlands were recorded in Groningen by a nobleman, Johan van Ewsum, in about 1570 (see Moormann 2002: 262–265). Strikingly they include cases of the *hi*-copula with a nominal complement.

(90) Copula with nominal complement<sup>84</sup>

	Orthography Sinti-Romani + Dutch	Analysis	Gloss
a.	<i>hiranij issen jüffer</i>	hi rani	'it's a lady'
b.	<i>hirai issen jünker</i>	hi raj	'it's a gentleman'
c.	<i>hirackilo dat issen knegt</i>	hi raklo	'it's a (non-Rom) boy'

84. See 1b.

d.	<i>gerackijlij</i>	<i>dat issen (schone) maegt</i>	hi rakli	‘it’s a (non-Rom) girl’
e.	<i>horom</i>	<i>dat issen man</i>	hi rom	‘it’s a (Rom) man’
f.	<i>hiromenij</i>	<i>dat issen frouwe</i>	hi romni	‘it’s a (Rom) woman’

Cases (90d, e) represent obvious mistakes in the transmission of this list (marked in bold), which, apart from a few cases of non-organic vowels splitting up consonant clusters, suggest that little has changed in Dutch Romani since the 16th century.

The form *hi* for the copula is an important indication of the type of Romani represented by the c.1570 recording. This has to be Sinti-Romani (Matras 2004).

In short, Romani, in the dialect known today as Sinti, was spoken in the Low Countries at that time and has left traces behind in these cryptolects. It follows that the copular and possessive constructions / combiverbs in PC and GB may be modelled on Sinti-Romani.

## 9. An examination of Romani lexical items in Bargoens

We give two diagrams on the following page to illustrate what we mean by the small number of Romani items surviving in Bargoens varieties. We assume that in earlier times the Romani vocabulary was more extensive. This can probably be deduced for the list reproduced in Moormann (2002: 365–372), from a manuscript originally from 1769, with a number of Romani items. This has terms for metal-working that are derived from Romani, as well as others that are not so derived. An earlier “Bargoensch” word-list, appearing as a appendix to Cartouche, of de Gestrafte Booswicht .. (1731) is restricted to *bink* ‘man’, *mol* ‘die, kill’, *pier-* ‘play’, and *koeter* ‘knife’.

Now, lexical items from Romani have diffused widely across all European secret languages and slang forms. So the presence of a number of Romani words does not prove our argument, but it does provide circumstantial evidence.

### 9.1 Some shared Romani lexical items

We will illustrate 13 words of possible/probable Romani origin which appear in the following Bargoens varieties.

- (91) a. Pottefers, Zele, Roeselare, Maastricht  
 b. Winschoten, Weert, Nijmegen, Tilburg (Netherlands)  
 St. Truiden (Belgium)

With the exception of the second row entry ‘woman, girl’ these are all fairly certain Romani items.

The Bargoens items often display a certain “semantic distance” with respect to their Sinti models. So, in the case of the first row entry *bink* ‘man’, the Romani meaning would appear rather to be ‘devil, thief’, a case of semantic bleaching.

We find the avoidance of Sinti-Romani /dʒ/ in various of the Bargoens words. Oddly enough there are three words in the list of 13 items containing this un-Dutch phoneme in initial position. That it is un-Dutch can be seen in the old Dutch loan from English *jam* /dʒæm/. This word appears as /ʃem/ in Dutch, with retention of the English spelling. A variety of replacements appear for Sinti-Romani /dʒ/. These are dealt with in (92).

A varied set of developments is also encountered in the voiceless counterpart to /dʒ/, /tʃ/. This is dealt with in (93).

**Table 15a.** Bargoens equivalents of Romani items<sup>85</sup>

Meaning	Pottefers	Zele	Roeselare	Maastricht	Groenstraat	Romani	English
<i>man</i>	bink	bink	beink, bink being	bink		beng	<i>devil, thief</i>
<i>woman, girl</i>	gieze(ke)	geeze	geeze	giezeke	gruus?	gadʒi?	<i>non-Romani woman</i>
<i>dog</i>	joekel		jokker, tjoeker			dʒukəl	<i>dog</i>
<i>chicken</i>				kachlientje		kaxli/ kaxni	<i>hen</i>
<i>bread</i>						maro	<i>bread</i>
<i>porridge</i>		sjaf	djafte			dʒob	<i>oats</i>
<i>money</i>						lovi	<i>money</i>
<i>door</i>				de val	val	vudar*	<i>door</i>
<i>slice/cut</i>	kotterik		kotterik	köteren	kot-ə**	kotər	<i>piece</i>
<i>play</i>	pieren	pieren pierig	pierig pierement	pieren pierder		perjas	<i>fun</i>
<i>kill, dead</i>	mol	moll(en)	mollen	molle	mol	mul-	<i>dead</i>
<i>steal</i>						tʃor-	<i>steal</i>
<i>ask, beg</i>	mangen		manken	manken		mang-	<i>beg</i>
<i>go</i>	jallen patjalder		jal-	jallen		dʒal(a)	<i>he goes</i>

\* In Bohemian Romani the forms *vudar* and metathesized *duvar* both existed (Ješina 1886: 120). Moorman (2002: 275) considers “de val” (Dutch for ‘the door’) to be a reinterpretation of *duval*

\*\* Is this a back-formed stem /kot-/ from /kotər/, with /ər/ reinterpreted as an instrumental suffix?

85. In Tables 15a and 15b only the Romani forms are transcribed phonemically.





d.	/gadʒi/	‘woman’			
i.	/dʒ/ > /j/	<i>gooi</i>	=	/goji/	1x
		<i>gonje</i>	=	/gõjə, gõji/?	1x
ii.	/dʒ/ > /z/	<i>gieze, geeze</i>	=	/gi(:)zə, ge:zə/?	7x

## (93) Replacements for /tʃ/

a.	/tʃor-/	‘to steal’			
i.	/tʃ/ > /ʃ/	<i>schōr-</i>	=	/ʃo:r-/	1x
		<i>schoer-</i>	=	/ʃu:r-/	1x
ii.	/tʃ/ > /tj/	<i>tjoer-</i>	=	/tju:r-/	1x

We also notice some cases of word-class shifts. So the Romani noun stem /perj-/ ‘fun’ becomes a verb stem /pirr-/ ‘play’. We find among others the following cases:

(94)	<i>pier-en</i>	‘to play’	verb
	<i>pier-ig</i>	‘pleasing’	adjective
	<i>pier-ement</i>	‘fair organ, street organ’	noun
	<i>pier-der</i>	‘player’	noun
	<i>pier-ig-erik</i>	‘pleasure’	noun
	<i>pier-bolle</i>	‘play+man’ > ‘player’	noun
	<i>pier-keete</i>	‘play+house’ > ‘casino’	noun

## 10. Possibilities for future research

Section 8 on the combiverbs involving detailed comparisons of their uses has brought to light subtle distributional differences in the occurrence of these verbs in the various lects. It has also led us to the hypothetical sources of the PC and GrB cv’s, Sinti-Romani. This part of our study would have been more feasible (and possibly more adequate) with an appropriate database.

A desideratum for future research is then the development of a database with word and sentence lists, and grammatical (including phonological) properties of the documented cryptolects. This could be a major instrument for the study of the typological and lexical overlaps between these languages. It could also lead to new insights into the question of the historical relationships among and contact between these cryptolects.

Moreover, in addition to mere relexification of content words, word-formation, grammatical vocabulary (numerals, personal pronouns, possessives, etc.) and phrase structure were also camouflaged. It would be useful to have a typology of the resources employed for this camouflaging by the secret languages under study.

## Acknowledgements

We would like to extend our grateful thanks to Piet Vos, editor-in-chief of the *Woordenboek van de Brabantse Dialecten* (‘Dictionary of the Brabantish Dialects’) produced by the former NCDN (Nijmeegse Centrale voor Dialect- en Naamkunde (‘Nijmegen Centre for Dialectology and Onomastics’)), for making available the Pottefers Cant ms. from the dictionary archive. Also to Lidy Jansen and Diedrik van der Wal of the Meertens Instituut library, who played an important role in acquiring the manuscript text from the Radboud University library, Nijmegen. We are also grateful to Yaron Matras and Nicoline van der Sijs for their insightful comments on this article. We would also like to thank Paul van Hauwermeiren for kindly making available various relevant publications. Lastly we would like to express our appreciation to the participants at the Colloquium on Mixed Languages, September 28–29, 2017, Universität Bremen, for useful questions and suggestions, and in particular to the organisers Eeva Sippola, Ingo Warnke and Maria Mazzoli for organising such a successful event.

## References and sources

- Bakkes, P. 2007. Sjian, kantj ane bóks. Roermondenaren en hun Bargoens. In *Riek van klank*, R. Keulen, T. van de Wijngaard, H. Cromptvoets (eds), 77–84. Sittard: Mediagroep Limburg.
- Bakker, P. 2006. Review of J. Kirk & D. Ó Baoill (eds), *Travellers and their Language*. *Language in Society* 35(3): 429–432.
- Bellmann, G. 2003. Ethnolinguale Unbestimmtheit und Interlingualität. Talk delivered at the Internationale Fachtagung “Niederländisch-Deutsch-Slawische Sprachkontakte im Mittelalter und ihre Spuren in (nieder-)deutschen Mundarten des 20. Jahrhunderts”, Lutherstadt Wittenberg, 25. October 2003.
- Bischoff, E. n.d.. *Wörterbuch der wichtigsten Geheim- und Berufssprachen. Jüdisch-Deutsch, Rotwelsch, Kundensprache; Soldaten-, Seemanns-, Weidmanns-, Bergmanns- und Komödiantensprache*. Leipzig: T. Grieben.
- Bosten, L., Goossens, H., Plum, J., Roemkens, H., Schunken, X. & Driessen, C. 1981. *Groenstraat-Bargoens woordenboek*. Ubach over Worms.
- Breuer, H. 1974. Mijnsluiting; personeelsbeleid in het kader van de sluiting van het steenkolenbedrijf ‘Laura-Julia’ in een tijd van regionale herindustrialisatie. Thesis, Katholieke Sociale Akademie Sittard.
- Coehorst, E. 1983. De plaats van het bargoens in het Venloos dialect. *Veldeke* 1983(2): 8–11.
- Creemers, C. 1871. *Aanteekeningen over het dorp Stamprooy*. Roermond.
- Dewulf, M. 1980. Bargoens te Sint-Niklaas. *Annalen van de Koninklijke Oudheidkundige Kring van het Land van Waals* deel 83, Jaarboek: 5–127.
- Duden Online. “hocken” in Duden Online. <<https://www.duden.de/rechtschreibung/hocken>> (12-7-2018).
- Endepols, H. 1924. Groenstraat-Bargoens. *De Nieuwe Taalgids* 18: 172–198.
- Finck, F. N. 1903. *Lehrbuch des Dialekts der deutschen Zigeuner*. Marburg: N.G. Elwert’sche Verlagsbuchhandlung.
- Goossens, H. M. W. 1981. *Ubach over Worms: Van Overworms kwartier tot gemeente Landgraaf*. Ubach over Worms.

- Goossens, J. 1965. Die Gliederung des Südniederfränkischen. *Rheinische Vierteljahrsblätter* 30: 79–94.
- Günther, L. 1919. *Die deutsche Gaunersprache und verwandte Geheim- und Berufssprachen*. Leipzig: Quelle und Meyer.
- Haalboom, A. 1995. 'Zwartje met flens en soetering'. *Woonwagennieuws* 3 april 1995: 16–17.
- Hermant, P. 1933. Glossaire d'argot Bruxellois (Burgonsch). *Le Folklore Brabançon* 13(73): 53–92.
- Hermant, P. 1933/1934. Woordenlijst van brusselsch Bargoensch (Burgonsch). *De Brabantsche Folklore* 13(73–76): 29–68, 186–218.
- Hinskens, F. 1985. Versjtrunkelt diene tuuën wat e keeäneske oop e huuëvelke drait? Enkele structurele en functionele aspecten van het 'Groenstraat-Bargoensch'. *Mededelingen van de Nijmeegse Centrale voor Dialect- en Naamkunde* XIX: 17–56.
- Hobsbawm, E. & Ranger, T. (eds). 1983. *The Invention of Tradition*. Cambridge: CUP.
- Honnen, P. 2000. *Geheimsprachen im Rheinland. Eine Dokumentation der Rotwelschdialekte in Bell, Breyell, Kofferen, Neroth, Speicher und Stotzheim*. Köln: Rheinland-Verlag (with a CD).
- Ješina, J. 1886. *Románi Čib oder die Zigeuner-Sprache (Grammatik, Wörterbuch, Chrestomathie)*. Leipzig: Verlag von List & Francke.
- Knippenberg, W. H. T. 1974. Haarteuten in Noord-Brabant en het kasboek van een Bergeijkse haarteut. In *De Teuten. Buitengaanders van de Kempen: Een verzameling losse bijdragen*, W. H. T. Knippenberg (ed.), 166–184. Eindhoven: St. Brabants Heem.
- Kluge, F. 1901. *Rotwelsch. Quellen und Wortschatz der Gaunersprache und der verwandten Geheimssprachen. I Rotwelsches Quellenbuch*. Strassburg: Trübner.
- Köster Henke, W. 1906. *De Boeventaal. Zakwoordenboekje van het Bargoens of De Taal van de jongens van de vlakke, in woorden en zinnen alphabetisch gerangschikt*. Dockum: Schaafsma & Brouwer.
- Labbé, J. 1892. Limburgsch Bargoensch. *Limburgsch Jaarboek*, 24–33. Bilsen: Het Genootschap.
- Matras, Y. 2004. *Romani: A Linguistic Introduction*. Cambridge: CUP.
- Moormann, J. G. M. 1930. De kramertalen. Een vergelijkende geheimtaal-studie. *Tijdschrift voor Taal en Letteren* 8: 49–83, 121–149, 233–266.
- Moormann, J. G. M. 1932. *De geheimtalen. Een studie over de geheimtalen in Nederland, Vlaamsch-België, Breyell en Mettingen*. Zutphen: W. J. Thieme & Cie.
- Moormann, J. G. M. 1934. *De geheimtalen. Bronnenboek*. Zutphen: W. J. Thieme & Cie.
- Moormann, J. G. M. 2002. *De geheimtalen: Het Bargoense standaardwerk, met een nieuw, nage-laten deel*, N. van der Sijs (ed.), met een inleiding van Enno Endt. Amsterdam: Uitgeverij L.J. Veen.
- Muysken, P. C. 2001. *Bilingual Speech. A Typology of Code-mixing*. Cambridge: CUP.
- Muysken, P. C. 2016. From Colombo to Athens: Areal and universalist perspectives on bilingual compound verbs. *Languages* 1(2): 1–18.
- Ó hAodha, M. 2002. Travellers' language: Some Irish language perspectives. In *Travellers and Their Language*, J. Kirk & D. Ó Baoill (eds), 47–63. Belfast: Cló Ollscoil na Banríona.
- Siewert, K. 1991. Masematte. Zur Situation einer regionalen Sondersprache. *Zeitschrift für Dialektologie und Linguistik* 58(1): 44–56.
- Siewert, K. 1997. Das Pfdelbacher Jenisch. Mit einem Glossar aus den schriftlichen Quellen. *Zeitschrift für Dialektologie und Linguistik* 64(1): 37–56.
- Stüve, L. 1923. *Die Geheimssprachen, insbesondere: Die Tiöttensprache genannt Bargunsch oder Humpisch; eine Geheimssprache westfälischer Kaufleute aus dem Kreise Tecklenburg. Mit vollständigem Wörterverzeichnis*. Recke: Vorberg.
- van Bolhuis, E. G. 1937. *De Gabbertaal: Woordenlijst van het Bargoens*. Ede: Niessen.

- van de Kamp, J. & van der Wijk, J. 2006. *Koosjer Nederlands. Joodse woorden in de Nederlandse taal*. Amsterdam: Contact.
- van der Sijs, N. 21995. Zigeunerwoorden in het Nederlands. *Trefwoord* 10: 102–110. Den Haag: Sdu Uitgevers.
- van der Sijs, N., (ed.). 2010. *Etymologiebank*. <<http://etymologiebank.nl/>> (27 April 2020).
- van Ginneken, J. 1914. *Handboek der Nederlandse taal, Deel II: De sociologische structuur onzer taal II*. Nijmegen: Malmberg.
- Van Hauwermeiren, P. 2004. Bargoens, een ‘taal’ met vele namen. *Taal en Tongval* 56: 160–186.
- Van Hauwermeiren, P. 2009. De belangstelling voor het Bargoens. *Taal en Tongval* 61: 114–183.
- Van Hauwermeiren, P. 2012. Nominale afleidingen in het Bargoens. *Taal en Tongval* 64: 263–296.
- Van Hauwermeiren, P. 2018. De Pottefers van Tienen en hun drol verknoeiide preuvel: De kramertaal van rondtrekkende handwerkslieden. *Van Mensen en Dingen: Tijdschrift voor Volkscultuur in Vlaanderen* 16(1): 9–20.
- van Nerum, H. n.d. De Pottefers. Ms, 9 pp. [ca 1964].
- van Winkel, F. 1974. De Teuten, benaming en oorsprong. In *De Teuten. Buitengaanders van de Kempen*, W. H. T. Knippenberg (ed.), 29–47. Eindhoven: St. Brabants Heem.
- von Sowa, R. 1898. *Wörterbuch des dialekts der deutschen Zigeuner* [Abhandlungen für die Kunde des Morgenlandes. Band XI, No. 1]. Leipzig: Deutsche Morgenländische Gesellschaft.
- Vek, J. 2002. *Tieske en Niske. Ing gesjichte uever d'r aldaag van de Grunsjtroater lüj um 1860 erum*. Bargoensjkloeb ‘ut Keieskwieëste’.
- Vos, P. & van Lieshout, L. 1990. *Woordenboek van de Brabantse dialecten, Deel II: Niet-agrarische vakterminologieën, Aflevering 5: Kaarsenmaker, sigarenmaker, potten-bakker, diamantbewerker*. Assen: Van Gorcum.
- Willems, J. F. 1838. Het Bourgondsch in de Kempen. *Belgisch Museum* II: 427–431.
- Winkler, J. 1874. *Algemeen Nederduitsch en Friesch Dialecticon II*. 's-Gravenhage: Nijhoff.
- Wolf, S. A. 1956. *Wörterbuch des Rotwelschen. Deutsche Gaunersprache*. Mannheim: Bibliographisches Institut AG.



PART 4

## Sociolinguistic aspects of language contact



# Sociolinguistic enregisterment through languagecultural practices

Leonie Cornips and Vincent de Rooij  
Humanities Cluster (KNAW) & Maastricht University /  
University of Amsterdam

This chapter will explore the effects of the sociolinguistic enregisterment of Heerlen Dutch in the carnivalesque summer song *Naar Talië/Naar Talia* ‘To (I)taly’, performed and uploaded onto YouTube by a band called the Getske Boys. The Getske Boys is a group of three male performers who, by selecting a particular set of linguistic forms from dialect, Dutch, in-betweens, Italian and English, work to enregister these as local to Heerlen-Noord and the speech of the coal miners who once lived there. Their selection of specific co-occurring forms is based on perceived past patterns of co-occurrences: an experiential knowledge, accumulated over the years, of the indexical ties between linguistic forms, specific (groups of) people and a specific place.

**Keywords:** coal mining, enregisterment, Heerlen Dutch, indexicality, languageculture, parodic language use, place-based identities, place-making, stereotyping

## 1. Introduction

This chapter will explore sociolinguistic enregisterment in the language practices of the *Getske Boys*, a popular band from Heerlen, Limburg, the Netherlands.<sup>1</sup> The focus of analysis will be on their language use on Facebook and on the lyrics of their carnivalesque summer song *Naar Talië/Naar Talia* ‘To (I)taly’. Inspired by Johnstone (2016), we will examine enregisterment as the process by which linguistic forms

---

1. The data used in this chapter are also analyzed in Cornips & de Rooij (2015), and Cornips, de Rooij & Stengs (2017) which take ‘belonging’ and ‘languageculture’ as central analytical concepts, respectively. We started working on these data during our stay at the Netherlands Institute of Advanced Study in the Humanities and Social Sciences (NIAS) in 2013–2014. The support of NIAS for this research is hereby gratefully acknowledged.



in this song become linked with social and place-based identities on social media (see also Androutsopoulos 2006; Beal 2009). New media, such as social media, lead to new mediations, new modalities and hence to new language practices (Androutsopoulos 2006; Leppänen 2012). We will focus on place-based identities (Antonsich 2010), and the central research question we will address is: how do actors engage in processes of place-making in a specific locality in the province of Limburg in the Netherlands, i.e., Heerlen, through language practices on social media, creatively drawing on features from what are considered local and regional language varieties? Actors on new media engage in acts of (dis)identification through language practices in many different ways and in relation to many different people, objects and places of attachment (Christensen 2009; Yuval-Davis 2006: 199). The aim of this paper is to reflect on the ways in which new media provide space for changing, locally embedded, *languagecultural*<sup>2</sup> (Cornips et al. 2017) practices.

The focus of this paper is on the relation between the use of linguistic forms that are crucial in the process of place-making and indexing social identities, and the cultural forms through which this awareness is expressed. The overall aim of the paper is to show that people perceived to be in the periphery – such as the Getske Boys – put distinct cultural and linguistic behaviors to use, which may lead to new, original linguistic and cultural forms. The linguistic construction of centralization-cum-peripheralization as exemplified in the performances of the Getske Boys may lead to detailed insights into processes of language variation and change. The languagecultural resources of the Getske Boys will be analyzed as a cultural activity linked to ideology (Hanks 1996, Silverstein, 1985, 1996) and as a materialization of a grammatical system, which offers an insight into why and how different actors (individuals as well as institutions) traverse linguistic variation space.

## 2. The sociolinguistic context

The Getske Boys have been performing in the carnival scene of Heerlen, a city located in the Dutch province of Limburg, since 1986. In order to grasp the how and why of the linguistic performance of the Getske Boys, a basic sketch of the political, sociocultural, and sociolinguistic context of Limburg is necessary. As the map in Figure 1 shows, Limburg shares more border with Belgium (west: 139 kilometres)

---

2. As we noted in an earlier publication (Cornips & de Rooij 2015), it was Agar (1994: 60) who was the inspiration for this word. Agar uses *languaculture*: “I stole the term from linguistic anthropologist Paul Friedrich, who suggested “*linguaculture*,” but that phrase keeps reminding me of pasta.” We believe ‘*languageculture*,’ although distinctly odd to native speakers of English, does justice to the equal value of both ‘*language*’ and ‘*culture*’ in this concept.



**Figure 1.** Map of the Netherlands and Flanders showing the Randstad area in the western part of the Netherlands (light grey) and Limburg in the southeastern part (dark grey)

and Germany (east: 212 kilometres) than with its neighboring Dutch provinces, Noord-Brabant and Gelderland.

The province of Limburg presents an excellent case for contemporary empirical exploration of how language and cultural practices and identity politics are interrelated in center-periphery hierarchies and dynamics (Thissen 2013). In the Netherlands political, economic, cultural and linguistic power are concentrated in the so-called Randstad area (Cornips, de Rooij & Smakman 2017). The province of Limburg is located ‘deep’ in the southeastern part of the Netherlands and is considered peripheral from an economic, sociocultural, and linguistic perspective.

Nationwide, the province of Limburg is known for its very lively, year-round cycle of carnival-related events (Weijers 1995; Thissen 2018). Some of these are relatively closed events for members only – although they may be reported on in local and regional media – but many of them are open to anyone, and what is more, they are highly mediated events.

Limburg is also famous for the prevalent use of ‘dialects’ (Cornips 2013; Hinskens & Taldeman 2013).<sup>3</sup> Around 900,000 people – or 75% of its inhabitants – claim that what they speak is a dialect (Driessen 2006: 103). During carnival and carnivalesque events, local ways of speaking may index and authenticate a specific kind of localness and at the same time denaturalize other kinds of localness (Bucholtz & Hall 2005). Both tendencies are caught up in a dialectic of mutual articulation; that is, in their attempts to center themselves as local and denaturalize others’ localness, actors like the Getske Boys reproduce the center-periphery in the public sphere. We, therefore, consider carnival and carnivalesque public events in Limburg as constituting a political force field engendering exclusion and inclusion on the basis of languagecultural practices (Cornips & de Rooij 2015).

The label ‘dialect,’ as used by both linguists and laypeople, is based on and reproduces specific linguistic, social, cultural and political assumptions. The problem with this use of ‘dialect’ is that it conceptualizes ways of speaking as clear-cut, bounded phenomena, comparable to the widespread use of ‘culture,’ or, rather, ‘cultures,’ as discrete objects (Cornips et al. 2017). A co-occurrence of linguistic forms labeled as dialect is “mapped onto geographical space” in the Germanic dialectological tradition or “onto demographically defined social groups” in the Anglo-Saxon tradition (Johnstone 2011b: 659), leaving little room for speakers’ agency. While the scientific label of ‘dialect’ may be criticized, we realize all too well that what speakers call dialect has psychological reality for them (Watson 2013: 237); that is, dialects are cognitive constructs creating borders between groups of speakers reflecting a shared sense of ‘they speak like us’ or ‘they speak differently from us,’ based on evaluating ways of speaking in relation to the local and social context of its users. Under conditions of globalization, however, local ways of speaking are a resource for speakers for indexing various, often instable and ambiguous, social and place-based meanings. The concept of dialect in this sense is processual; that is, “a set of linguistic forms linked with and constitutive of a context” (Johnstone 2016: 633), a practice that, like speakers, is constantly moving, interacting, and changing. From such a perspective, it would be a serious mistake not to pay attention to translocality and the workings of ideology (Woolard 2008) in languagecultural practices like the ones we are focusing on in this chapter. A brief sketch of the development of the coal mining industry in and around Heerlen shows how mobility and translocality have shaped the languagecultural matrix of this urban environment.

---

3. The ‘dialects’ of Limburg were granted Minority Language recognition in 1997, under the label ‘Limburgish’ by the Netherlands, a signatory to the 1992 European Charter for Regional Languages and Minority Languages. Since then public funds have been made available for promoting the use of Limburgish, as well as protecting the language rights of its speakers.

Cornips (1994, see also Cornips & de Rooij 2019) recounts that in the first decades of the 20th century Heerlen developed into the center of a rapidly expanding coal mining industry. During that period Heerlen witnessed a massive influx of migrants from outside Limburg and the Netherlands. This immigration transformed the sociolinguistic situation of Heerlen in two important ways: (i) the native population that spoke the local dialect became almost a minority within a span of twenty years (see Table 1), and (ii) a new intermediate variety or, rather, a new regional variety of Dutch, Heerlen Dutch, emerged:

**Table 1.** Number and origin of inhabitants of Heerlen between 1899 and 1930 (taken from Cornips 1994)

Year	Number of inhabitants of Heerlen	Born in the province of Limburg %	Born outside the province of Limburg %
1899	6312	87.8	12.1
1920	33014	47.8	51.1
1930	46917	45.3	54.7*

\* 22% of whom were born outside the Netherlands

As Cornips' previous analysis reveals (1994), the language shift was a rapid process. The shift began with the use of forms and structures from the shifters' dialect into their version of Dutch; subsequently, these patterns have spread to Dutch spoken in Heerlen as a whole (cf. Thomason & Kaufman 1988: 38, 51).

### 3. The peripherality of Limburg

The growing significance of celebrating carnival in Limburg seems to parallel the troublesome 'negative integration' (Knotter 2009) of Limburg into the Dutch nation-state starting in the 19th century. What is now called the Dutch province of Limburg was originally a geopolitical anomaly without a common history. Knotter (2011: 1–14) points out that the area of present-day Dutch Limburg has been a border area since Carolingian times, fragmented into an array of small territories and enclaves, even after the Westphalian Treaty of 1648, when the territorial state system of early modern Europe was more or less delineated.<sup>4</sup> As a consequence of the Second Treaty of London (1867), Limburg became a regular province of the Netherlands, thus its integration is relatively recent compared to that of other Dutch provinces. Knotter (2011) analyzes Limburg's integration into the Netherlands as a negative one, since together with its integration into the Dutch nation-state, it

4. The historical outline of Limburg is taken from Knotter (2011: 1–14).

simultaneously developed a strong regional identity by recognizing itself as deviating from the national norm. An echo of how Limburgian languageculture has been constructed in the past is still present in a 2014 brochure put out by the Government of the Province of Limburg:

Over the centuries, the region has been ruled by the Romans, Spanish, Prussians and French, each of whom have left their indelible mark. After the Napoleonic wars in 1815, Limburg was united under the United Kingdom of the Netherlands. King Willem I decided to name the region 'Limburg' after the 'Limbourg' castle in the Belgian Ardennes, the seat of a medieval principality. Following Belgian secession in 1839, Limburg was divided between the Dutch and Belgian crowns. Limburg's vast cultural diversity is expressed in its variety of dialects. But this cultural diversity is also reflected in a rich tapestry of events, the most famous being the annual shooting tournament Oud Limburgs Schuttersfeest (OLS), the World Music Contest (WMC) in Kerkrade, and the European Fine Art Fair (TEFAF) in Maastricht.<sup>5</sup>

People's self-identification in the new province of Limburg in the 19th, and 20th century as well, took place through perceived "common economic interests, religion, habits, character, and culture", which heavily relied/s on the outsider figure of the Hollander, regarded as being everything the Limburger is not (Cornips & Knotter 2017: 18; Knotter 2011). According to Knotter (2011): "especially low culture ... has proved itself to be the most sustainable part of provincial particularism. Taking part in the catholic rites de passage (baptism, first communion, marriage, burial), processions and associated festivities, celebrating carnival, playing brass music, using dialect, being a member of a shooting association, and the like, all are presented as typical manifestations of regional identity. However, they can only be considered 'typical' because they are deviant from a national Dutch norm."

Taken together, the historical context of Limburg may explain why languagecultural actions, especially those that take place in public settings such as celebrating carnival and speaking dialect, are highly susceptible to increased scrutiny and social evaluation.

Due to power asymmetries in center-periphery dynamics between Holland and 'Limburg,' people in Limburg (the (perceived) periphery) may feel marginalized economically, politically, culturally, and linguistically by Holland (the (perceived) center). In Limburg this is often referred to as the *Calimero* effect. Calimero is an Italian/Japanese cartoon about the only black chicken in a family of yellow chickens. He wears half of his eggshell still on his head. Calimero originally appeared on the Italian television show Carosello on July 14, 1963 and soon became a popular icon

---

5. Brochure of the Province of Limburg, May 2017, see <[https://www.limburg.nl/publish/pages/4/the\\_province\\_of\\_limburg.pdf](https://www.limburg.nl/publish/pages/4/the_province_of_limburg.pdf)> (20 Mai 2014).

in Italy.<sup>6</sup> In the Netherlands, Calimero became famous for frequently saying: *Zij zijn groot en ik is klein, en da's niet eerlijk, o nee* 'They are big and I is small, and that's not fair, oh no.'

One clear example of power pressure from the center at the nation-state level is the process of language standardization that is imposed on speakers in the periphery. As pointed out by Wang et al. (2014: 36), "[i]n many countries, one finds a deeply entrenched language ideology that is built on the unquestioned principle that a nation-state should be linguistically uniform" (see also Milroy 2001: 531; Silverstein 1996). The process of language standardization in Limburg commenced right from the start of its integration into the Dutch nation-state. Before then, the (Hollandic) Dutch standard language in spoken form was nearly nonexistent in Limburgian society. It is no coincidence that simultaneous with Limburg's integration into the Dutch nation-state, the codification and promotion of local varieties of labeled dialects took place (Cornips et al. 2016). Both the urban bourgeoisie and the rural clergy in Limburg, especially in the capital of Maastricht, took an interest in the local language as a foundation of authentic local culture and identity. This allowed the dialects to maintain a far stronger social and cultural position than elsewhere in the Netherlands.

Within center-periphery dynamics in Limburg, Maastricht is considered a city with a long history and a rich tradition of cultivating 'its' dialect (Extra 2004; Hagen & Giesbers 1988: 34; Kessels-Van der Heijde 2002; Münstermann & Hagen 1986), which is perceived to be of a higher social status than that of other dialects in Limburg and certainly more prestigious than the one spoken in Heerlen, which was influenced by numerous migrants in the first decades of the 20th century.

The process of standardization works by promoting the varieties spoken and written in the center of political and economic power in the nation-state and by declaring them as the national 'language.' The imposition of a standardized national variety on speakers in the periphery ipso facto devalues 'incorrect' ways of speaking considered dialectal and peripheral, and there are repercussions at both the personal and institutional level for flouting the standard's strictures. Because of the prestige of the national standard variety, speakers in the periphery are perceived as not speaking 'the right way' and are often negatively stereotyped as backward. Preston (2013: 62) claims that this language ideology largely stems from notions of 'correctness' and 'pleasantness.' Speakers of 'correct' varieties do not believe they speak non-standard varieties or 'dialect.' In response to this, as pointed out by Preston (2013: 177), "speakers of devalued varieties (like discriminated groups in general) derive solidarity from their distinctive behaviors, in this case linguistic

---

6. <<http://en.wikipedia.org/wiki/Calimero>>

ones.” Languagecultural differences between the perceived center and periphery are thus never socially neutral but constitute rich resources for social meaning making and identity work for people. Center-periphery dynamics are not only a question of inequality but also about who is in control and who is not (Massey 1993: 62): it is about who is considered to have the legitimate right to produce, own, market and distribute the ‘right’ kind of language and culture and who is not. In addition, the periphery and center are not fixed notions that can be defined but are part of a dynamic that is “ongoing, always shifting, multidimensional, heterogeneous and ambiguous” (cf. Pietikäinen & Kelly-Holmes 2013: 1–3).

#### 4. The Getske Boys and their audiences

Globalizing forces and advances in media technologies have made people in the periphery more aware than ever before of languagecultural differences between themselves and those occupying the center. New media and the growing commodification of what is perceived as sounding local and authentic have created a stage for public events in which various individual and institutional actors (re)produce their own linguistic versions of (local) uniqueness. There are many, widely differing ways of doing identity through dialect practices due to their vitality and prestige in Limburg. For example, there is a popular Facebook website, *Nine-gag Op zn Limburgs* ‘Nine-gag In Limburgian,’ where people write in dialect, and play out the opposition between Holland and Limburg on an almost daily basis. A famous topic is the ‘Hollander’ who doesn’t know how to celebrate carnival, unlike Limburgians who have carnival in their ‘blood.’<sup>7</sup>

The Getske Boys have uploaded their songs on YouTube (see fn 1) and maintain an active Facebook page.<sup>8</sup> The noun *gats* and its diminutive *getske* mean a narrow passage through houses or meadows. One was able to move quickly from one place to another through such a *gats*. Moreover, a *gats* was most often very filthy due to the garbage that was flung out there. A compound noun is *koetegats*, where *koete* has the meaning of ‘mucus’ (lit. ‘mucus alley’).

The *Getske Boys* consist of two men and their brother-in-law – Wim, Ad en Jan (Cornips & de Rooij 2015; Cornips et al. 2017). They are from an upper middle class milieu and perform the roles of Takkie (‘small branch’), Freddie and Klets (‘Chatty’). As part of Heerlen’s carnival scene, the Getske Boys were already locally known as

7. <<https://www.facebook.com/9gaglb?fref=ts>> (50.463 likes on 29 April 2014).

8. <<https://www.facebook.com/pages/Demi-Sec/134392459981850?fref=ts>> (2.612 likes, 6 June 2014).

*Demi Sec*: three so-called ‘dry’ figures usually dressed in black with whitened faces, who have taken part in the annual Heerlen carnival parade since 1986. The members of the Getske Boys, in contrast, while performing ‘naar Talië’, are loaded up with bottles, they sing in deviant Dutch about farts and scantily dressed women who can be conquered in a second, their glasses are filled with beer, they wear ill-fitting suits, show their naked bellies, smoke cigarettes, wear tacky hairdos and are decked out with far-too-large golden chains; they are one big parody: but whose parody exactly, and why do they do it like that (cf. Cornips & de Rooij 2015; Cornips et al. 2017)?

The Getske Boys’ performances and the linguistic forms they use employ can mean different things to different members of the audience, who “may draw on different cultural frameworks to make these forms meaningful” (Johnstone 2011a: 658). For the outsider unfamiliar with Limburg’s carnival scene, the Getske Boys may parody the figure of the plebeian as the personification of bad taste, or the far-too-large golden chains may refer to hip hop, or to lower working class people, since, as Wim, Jan and Ad inform us: “people who live their lives at the bottom of society show everything they have.”<sup>9</sup> People in the Randstad will undoubtedly interpret their way of speaking as deviating from Dutch and probably as incomprehensible, and will likely label it as dialect (cf. Cornips 1994). For the Limburger, the Getske Boys may personify the figure of the ‘Hollander’ who speaks no dialect, drinks beer all the time while in Limburg, exhibits rude behavior and makes a lot of noise (Mathijssen 2011). Alternatively, the Getske Boys portray the inhabitants of Heerlen ‘as the Hollander,’ that is, as the people who are the descendants of former immigrants, (re)producing Heerlen and its inhabitants as not being part of an ‘authentic’ Limburg.

For the insider of carnival and the carnivalesque, the Getske Boys personify a mode of carnival that is in contradiction to the official, institutionalized and disciplining one (Cornips & de Rooij 2015): “The beige costumes of the Getske Boys and their simple hats without peacock feathers are the opposite of the official carnival uniforms, their showy ‘golden’ medals parody the official orders that are ritual gifts in the performance of the Blue Vessel [a prestigious, local Heerlen carnival association], and their mullets, [a hairstyle characterized by long hair at the back with the rest of the hair cut short] their rowdy behaviour and exaggerated facial expressions are carnivalesque comments on the neatly coiffured well-behaved dignitaries of regular carnival associations.” (Cornips & de Rooij 2015: 93). For the Limburger, their way of singing and speaking is neither Dutch nor dialect but something in-between (see below). For the insider of the Heerlen carnival scene, the Getske Boys parody the local disciplining mode of carnival by the official carnival dignitaries such as

---

9. Interview with the Getske Boys (11 May, 2013, in Café Pelt on Pancratiusplein in Heerlen: “Personen aan de onderkant laten zien wat je hebt”.



the Blauw Sjuut ‘Blue Vessel’<sup>10</sup> and Winkbülle<sup>11</sup> ‘windbags’ (Cornips & de Rooij 2015). The language they use is immediately recognizable as being typical of those Heerlen neighborhoods that once formed the heart of the linguistic melting pot of coal mining migrants (Cornips 1994); this language variety contrasts sharply with the ‘pure’ dialect spoken and sung by the official Heerlen and other Limburg carnival associations and used in the official annual carnival song context.

The social significance of and the oppositions at work in center-periphery dynamics are thus a scalar phenomenon taking place at the national level, and also within the province, cities, neighborhoods and smaller local entities (cf. Eckert 2011; Gal & Irvine 1995).<sup>12</sup> The process of this layered and ambiguous meaning-making by embedded oppositions all combine in one way or another to construct linguistic ideologies and orientations, and give meaning to linguistic variation (Eckert 2011). Consequently, center-periphery dynamics may lead to the emergence of multiple new norms that may change depending on specific political and economic processes in the periphery (Pietikäinen & Kelly-Holmes 2013: 9).

The Getske Boys rebel against dominant (languagecultural) norms at local and national levels, revealing fractal recursivity (Gal & Irvine 1995). Local oppositions in Heerlen are embedded within larger oppositions in Limburg, and again at the larger nation-state level, with the combined effect of a social and linguistic continuum. The Getske Boys, situated in and imagined as being in the periphery, find ways to (re)shape and strengthen place-based and social identities through languagecultural practices, i.e., the process by which linguistic forms become ideologically linked with social and place-based identities.

The imagined audiences of the Getske Boys are not so much the speakers of the “right kind” of standard language or people from the Randstad or other provinces. As local celebrities, the Getske Boys’ primary audience is to be found in Heerlen in the first place. A good example of the linguistic uptake of the Getske Boys’ register by a local audience, restricted to inhabitants of Heerlen, is the use of the word ‘ram’ (see Section 7.2 under Lexicon). However, the “absent others” are crucial as invisible categorical referents for the Getske Boys to make their point. With their songs and appearance, the Getske Boys “question dominant ideas in society about what is correct and what is wrong, real or fake, pure or polluted” (cf. Cornips, de Rooij & Stengs 2017: 84). These are the kinds of ideas that usually remain implicit, but which are at the same time responsible for mobilizing processes on the basis of which people are judged and consequently included or excluded.

10. <[http://www.blauwsjuut.nl/nieuws\\_li.html](http://www.blauwsjuut.nl/nieuws_li.html)> (28 April 2020).

11. <<http://www.winkbulle.nl/>> (28 April 2020).

12. Gal & Irvine’s concept of ‘fractal recursivity’ that is relevant here can be traced back to Evans-Pritchard (1940).

## 5. The Getske Boys, social type and place-making on Facebook: Heerle-Noord Ouwhoer

The audience has to be local to be able to make sense of the self-presentation on the Facebook page of Demi-Sec aka The Getske Boys and their hometown, which is *Heerle-Noord Ouwhoer*.<sup>13</sup> Etymologically, *ouwhoer* means ‘old whore,’ but the Getske Boys use it as a frequent tag in sentence-final position, for example, the song *Naar Talia* closes off with *Mille grazie ouwhoer*. Here they draw on the perceived traits of people living nowadays in the former koloniés, i.e., coal mining neighborhoods such as Vrieheide, Hoensbroek, Passart and Nieuw-Einde near Brunssum.<sup>14,15</sup>

‘Het is een persiflage op mensen uit de volksbuurt, uit de (koel)kolonie, die vrijgevochten van aard zijn, een grote bek maar een klein hartje hebben. Waar zij vandaan komen heersen de wetten van de straat. In Heerlen waren dat doorgaans de wijken aan de verkeerde kant van spoor zoals dit in de volksmond genoemd werd.’ *It’s a parody on people from the popular neighborhood, from the (pit) colony, who are unruly, have a rough tongue (big mouth), yet are soft-hearted (have a small heart). Where they come from, the law of the street rules. In Heerlen, these were usually the neighborhoods on the ‘wrong side of the tracks,’ as they are popularly called.*

The self-presentation shows that the Getske Boys are preoccupied with persuasive place-making and construction of a social type on Facebook. The coal miners’ neighborhoods – koloniés – to which the Getske Boys refer, were planned and built by the mining companies in an arrangement similar to that of a garden city. Virtually all the men earned their livelihoods in the pit, leading to a situation in which the entire regional economic structure was related to it, in which the work itself was an almost exclusively male domain. As a consequence of the exclusion of women from mining work, the public space of the kolonié was entirely under male dominance (Stuyck et al. 2008).

The neighborhoods on the ‘wrong side of the railway tracks’ are and were peripheral vis-à-vis the neighborhoods in the south of Heerlen that included the city center as well. The executives and pit-engineers migrated from the Randstad area to Heerlen-South and, because of their origins, spoke the ‘right’ kind of standard Dutch. In consequence, the segregation of Heerlen’s public space – south versus north – mirrors/mirrored the strict functional segregation between ‘high’ (above ground) and ‘low’ (underground) laborers in the pit.

13. They spell *Heerle-Noord* without a final *n* (see § 4).

14. Interview with Getske Boys, May 2013.

15. <<https://www.facebook.com/pages/Demi-Sec/134392459981850?fref=ts>> (28 April 2020).

The self-presentation on Facebook also shows (see below) that the Getske Boys link language ideologically with social and place-based identities. The people in the former koloniés in Heerlen-Noord use:

De taal die gesproken wordt is een combinatie van Nederlands met dialect, een taal met knoebelen, maar erg herkenbaar.

*The language spoken is a combination of Dutch and dialect, a language with 'lumps,' but a very recognizable one.*

Within Limburg (south), Heerlen is indeed well known and stereotyped for its 'language with lumps.' In 1952 Dutch dialectologist Kats (1952: 318) wrote:

ieder die zijn oor wel eens te luisteren legt op straat, in tram of bus, zullen evenals ons wel eens de taalkundige haren te berge zijn gerezen bij het horen van het Nederlands dat daar (in Heerlen/ LC) vaak wordt uitgebraakt.

*Everyone who listens in the street, on the tram or bus will have their linguistic hairs stood on end by listening to the kind of Dutch that is often spat out there (in Heerlen/LC)*

The language spoken in Heerlen has acquired many epithets emphasizing its incorrect Dutchness – or maybe better, its non-Limburgishness – during the 20th century (cf. Cornips 1994). Stereotypes used to denote the people in Heerlen are: *In Heerlen is Limburg gestorven* 'Limburg died in Heerlen,' *Hollanders* 'Dutch people, i.e., western people' and for the kind of Dutch they speak: *Misjmasj* 'mixed language,' *Hollesj mit knoebele* 'Dutch with lumps,' *Steenkolennederlands* 'coal mining Dutch' and *Huillands* (lit. Cry-landish) 'a kind of Dutch that makes you cry.' (Cornips et al. 2017; Cornips & de Rooij 2019). These epithets show that what is spoken in Heerlen is 'from the margins,' that is, a language that falls outside the existing dominant order, which is neatly divided into these two categories: a language is either the standard or a dialect. Due to the influx of many immigrants from 'Holland' to Heerlen-South, Dutch was introduced into Heerlen much earlier (i.e., the early twenties of the previous century) than was the case in other localities in Limburg, where it was introduced around the sixties (Cornips 1994; Weijnen 1967: 20). Therefore, *Hollesj mit knoebele* 'Dutch with lumps,' as Heerlen Dutch is described (Cornips 1994), may be regarded as an anomaly which finds itself on the margins, as "betwixt and between" (Cornips et al. 2017: 84). This position on the margins helps to explain the extra potential of Heerlen Dutch for parody and for carnivalesque performances, especially in an environment (Limburg) where the celebration of carnival and the ability to speak dialect are important, politically charged dimensions of what it means to be a Limburger. In sum, Heerlen is construed as different, as 'Holland' within Limburg. In their performance of *Naar Talia/ Talië* in 2010, the Getske Boys exaggerate and magnify this *Hollesj mit knoebele* 'Dutch with lumps,' and the carnivalesque cultural form in which this takes place is

a vehicle to interrogate stereotypes and prejudices, and hence significant dominant normative languagecultural norms (carnival, standard Dutch and dialect) within national and local levels.

## 6. The lyrics of *Naar Talia/Talië* (2010)

In their song *Naar Talia/Talië*, the protagonists go to Italy in a fast car ‘to steal small songs’ (*liedjes klauwe* in line 2). *Liedjes klauwe* (line 2) is a jibe at local artists in Maastricht, the capital of Limburg, who, in the eyes of Wim, Ad and Jan, ‘steal songs,’ since they make conventional carnival songs in the prestigious Maastricht dialect for a commercial corporate CD-producer and, hence, ‘go for the money.’ ‘*Liedjes klauwe*,’ i.e., singing a conventional song in a perceived ‘pure’ dialect, is thus a languagecultural practice that establishes borders between Heerlen (periphery) and Maastricht (center). The protagonists in *Naar Talia/Talië* tell us that Genaro Clementaro has invited them to come to Rome where he lives. Before moving to Rome, Genaro worked at the oldest and most famous Italian ice cream salon, *La Veneziana* in the Heerlen city center, founded by an Italian immigrant family in the thirties of the previous century, emphasizing Heerlen’s past as an immigrant city. Here the Getske Boys, in mentioning *La Veneziana*, explicitly connect place and practice, linking the spatial with a lifestyle (cf. Taylor n.d.) that unfolds in the song to be discussed below. While in Italy, they eat pasta instead of French fries and mashed potatoes, and they drink a lot of beer that they have brought with them because the Italian beer has no taste. They lie on the beach and sit with their balls in the sea watching scantily dressed women with only a shoelace between their buttocks. They can’t help but sweat, the pasta gives them diarrhea, and they have to fart. They try to talk to a beautiful woman, but she doesn’t understand Dutch. They go to Rome to the Vatican to light a candle for one of their fathers who is ill. The priests close the doors upon seeing them arrive, and they consequently burn the Vatican down with petrol.

In order to communicate a sense of rootedness in Heerle-Noord, the Getske Boys employ what Taylor (n.d.) terms *metastylistic discourse*, that is, speech about style. By referring to contextually-bound stylistic practices such as looking at scantily dressed women and trying to seduce them; farting, sweating and having diarrhea; singing about their balls;<sup>16</sup> drinking their own beer; not being

---

16. These themes immediately bring to mind Bakhtin’s ‘material bodily lower stratum’ central to what he calls ‘grotesque realism’ (Bakhtin 1984[1968]). A Bakhtinian analysis of the song would certainly be interesting but, given the focus of this chapter, we will not engage in such an analysis here.

able to speak Italian; showing off by driving in a fast American car; and using *ouwhoer* all the time, they evoke a particular masculinity, characterized by what they (and their audience) *assume* to be lower social class characterological traits that come naturally to those living in these neighborhoods. In their view, these traits are indexical of Heerlen-Noord as a lower-classed and male-gendered experience of place. The production of meaning in the song *Naar Talia* is thus based on the accumulation of experiential knowledge of specific co-occurring linguistic forms, specific (groups of) people and a specific place. The selection of linguistic forms and hence deselection of others on the basis of past and current recontextualizations evoke sociolinguistic enregisterment of how males of a low social class (probably descendants of former immigrants working as coal miners who originated this way of talking (cf Cornips 1994) in Heerlen-Noord speak and write, according to the Getske Boys.

Let us examine the unexpected linguistic forms in *Naar Talia*. **Words in bold-faced print** in the song lyrics (taken from their video presentation on YouTube) reveal linguistic forms associated with the local dialect of Heerlen at the level of lexicon, phonology, morpho-syntax and semantics; *italicized words* reveal 'non-standard/unconventional/marked use of Dutch including orthography'; underlined words show hybrid use between English, dialect, Dutch and Italian; words underlined, in italics and bold are associated with German, and words in normal font indicate standard Dutch (nonstandard Dutch in italics in English translation) (see also Cornips & de Rooij 2015):<sup>17</sup>

---

1 He Klets, wat gaan jullie doen in Italië	Hey, Klets, what are you going to do in Italy?
2 <i>Zeik</i> bouwe, liedjes <i>klauwe</i>	Fool around, nicking songs [zeik 'urine,' klauwe 'to steal']
3 Talië, <u>heer we cum</u>	To Talia, to Talia, Talia here we cum
4 Ken je Genaro Clementaro,	D'ya know Genaro Clementaro,
5 Die <i>tot voor</i> 'n jaar of twee	Who till a year or two ago
6 Bij de la Veneziana	At the Veneziana [Italian ice cream shop in Heerlen]
7 Alle ijsjes <i>draaje dee</i>	Made all the ice cream
8 Die <i>doet woone</i> in Rome,	He lives in Rome
9 vroeg ons te koome	Asked us to come
10 met ze drie of met ze twee	With all three of us or just the two
11 <b>Doe</b> de spaghetti maar <b>vas make</b> ..	Get the spaghetti ready already ...
12 Maar we <i>neemeeeeeeee</i> ....	But we'll bringiiiiing ...
13 Ons <i>eige</i> bier mee!!!	Our own brew!!!

---

17. See: <<http://www.youtube.com/watch?v=QToKL91O8ro>> uploaded on 1 July 2010 (23 July 2014; 52.693 reactions).

*Refrain*

14. Naar Talië	To Talië
15 Met de Camaro naar Genaro	With the Camaro to Genaro
16 <b>doen</b> we we <b>blaaze</b>	We will blow
17 voor een weekendje of twee	For a weekend or two
18 Naar Talië	To Talië
19 lekker <i>zitte</i> met chianti, <i>wijve kijke</i>	Sitting comfy with chianti, watching chickies
20 met je <i>klitse</i> in de zee	With your balls in the sea
21 Naar Talië	To Talië
22 Pasta <i>vreete</i> , <i>knoflooksjeete</i> ,	Gobbling pasta, garlic farts
23 <b>hendig</b> <i>zjweete</i> , aan <i>de dunne op de plee</i>	Sweating a lot, with the runs on the loo
24 Ja we gaan daar <i>ram geniete</i> ,	Yep, we'll have fun there
25 zonder <i>friete</i> , zonder puree	No French fries, no mashed potatoes
26 Maar dat bier is niet te <i>zuipe</i>	But that beer is not for boozing
27 Dat neeme we <i>zellevers</i> mee!!!	We'd rather bring that ourselves!!!

*End of ref.*

28 Aan de <u>labia Maggiore</u> , ken je vet <i>sjcore</i>	On the labia Maggiore [pun on Lago Maggiore], it's easy to score
29 <b>Hei</b> je zo 'n "Bella" mee	You can get a 'Bella' in a jiffy
30 <u>Buona note</u> , <sup>18</sup> lekkere <b>votte</b>	Buona notte, hot booty
31 met zo'n veeter in de <i>sjnee</i>	With a shoelace in her crack
32 Vroeg aan Sophia:	Ask for Sophia
33 <i>O cara mia</i>	cara mia
34 Of ze nederlands	If she did
35 <i>sjpreeke dee</i>	speak Dutch
36 " <u>Mi scusiare</u> ,	'Mi scusiare, <sup>19</sup>
37 <u>non parlare</u> "	non parlare'
38 dus daar <b>ken</b>	thus that is beyond
39 jeeeeeeeeeee...	
40 Dus <u>niente</u> mee!	all practicalities

*Refr. 2x*

41 en in Rome <i>aagekoomee eeve</i> naar ut Vaticaan	and once in Rome, on to the Vatican
42 want <i>me ouwe</i> die lag <i>sjlech</i>	'cause my old man's ill
43 tus ik denk ik <i>sjteek</i>	so I think I'll light
44 een kaarsje aan	a candle
45 maar toen we bij de kerk <i>aankwaame</i>	but when we got to the church
46 <i>hebbe</i> die <i>paatere dich</i> gedaan	those priests had shut up shop

---

18. The correct spelling is 'notte.'

19. du alte nutte 'you old whore' is German.

47 en toen is die hele santekraam	and then the whole shebang
48 tus in de fik gegaan	thus went up in flames
49 zittur altijd 'n kaarsje bij...	always a candle with it ...
lalalalalalalalala	lalalalalalalalala

Refr. 3x

50 <u>Angelo Bombrini</u>	Angelo Bombrini
51 <u>Salute, du alte nutte!</u> <sup>20</sup>	Salute, du alte nutte!
52 Mille grazie <i>ouwhoer</i>	Mille grazie ‘old pro’

## 7. The linguistic forms

### 7.1 Parodying how they write in Heerlen-Noord

The spelling of the lyrics of *Naar Talia* is indexical of semi-literate people with little formal education. Unexpected ways of writing are:

*The deletion of word final -n.* The *-n* (although not produced in standard spoken Dutch) should be written according to the standard spelling norm. The *-n* is omitted in almost all potential contexts ( $n = 25$ ):

*infinitives:* *bouwe(n)* (line 2) ‘build’, *klauwe(n)* (line 2) ‘steal’, *woone(n)* (line 8) ‘live’, *koome(n)* (line 9) ‘come’, *make* (line 11) ‘make’, *blaaze(n)* (line 15) ‘blow’, *zitte(n)* (line 19) ‘to sit’, *kijke(n)* (line 20) ‘watch’, *knoflookscheete* (line 22) ‘garlic farts’, *vreete(n)* (line 22) ‘gobble’, *zjweete(n)* (line 23) ‘sweat’, *geniete(n)* (line 24) ‘enjoy’, *zuipe(n)* (line 26) ‘booze’, *sjcore(n)* (line 28) ‘score’, *aagekoo mee(n)* (line 41) ‘arrived’, *aankwaame(n)* (line 45) ‘arrive’;

*inflected verbs* (3ps, pl, presens): *neemeeeeeeee(n)* (lines 12, 27) ‘bring’ and *hebbe(n)* (line 46) ‘have’;

*adjective:* *eige(n)* (line 13) ‘own’;

*plural nouns:* *wijve(n)* (line 19) ‘chickies’, *klitse(n)* (line 20) ‘balls’, *friete(n)* (line 25) ‘fries’, *votte(n)* (line 30) ‘booties’;

*adverb:* *eeve(n)* (line 41) ‘for a moment’;

*preposition:* *aa(n)*(gekoo mee) (line 41) ‘on/at’.

*Long vowels* [o], [e] and [a] in open syllables are spelled as double ‘oo’, ‘ee’ and ‘aa’ in all potential contexts ( $n = 12$ ) instead of standard single ‘o’, ‘e’ and ‘a.’ Thus, *woone* ‘live’ (line 8) is spelled with ‘oo’ instead of standard *wonen*; *koome* ‘come’ (line 9) instead of *komen*; *neemeeeeeeee* ‘to take’ (line 12) instead of *nemen*; *vreete* ‘to chow’ (line 22) instead of *vreten*; *knoflookscheete* ‘garlic farts’ (line 22) instead of *knoflookscheten*; *zjweete* ‘to sweat’ (line 23) instead of

20. The correct Italian form is ‘scusare’; it is unclear whether ‘scusiare’ is a deliberate ‘mistake.’

zweten; *veeter* 'shoelace' (line 31) instead of *veter*; *sjpreeke* 'to speak' (line 35) instead of *spreken*; *aagekoomee* 'to arrive' (line 41) instead of *aangekomen*; *eeve* 'for a moment' (line 41) instead of *even*; *aankwaame* 'to arrive' (line 45) instead of *aankwamen*; *paatere* 'fathers, Catholic priests' (line 46) instead of *paters*.

This spelling is indexically ambiguous: it may express the specific and conspicuous pitch contours of speakers in Limburg both in their dialect and standard variety, which are generally referred to as a 'melodious' or 'singsong' accent in the Netherlands (Hagen & Giesbers 1988), and it indexes first grade pupils' writing lessons in elementary schools. The spelling of double 'oo,' 'aa,' 'ee' versus 'o,' 'a,' 'e' in open syllables is a well-known beginner's misspelling.

*Spelling denotes 'how it sounds'*

- schwa -ə spelled as 'u' instead of standard 'e': definite neuter determiner *ut* 'the' (line 41) instead of standard *het*; *zittur* 'sit there' (line 49) instead of *zit er* and the combining of the locational verb *zit* 'sit' and adverb 'er' as one lemma instead of standard two;
- possessive pronoun *me* 'me' (line 42) instead of *mijn*; *w* instead of *d*: *ouwe* 'old' (line 42) instead of standard *oude* and *d* spelled as *t*: *tus* 'thus' (line 43,48) instead of *dus*;
- *i* spelled as *j*: *draaje* 'turn' (line 7) instead of standard *draaien*.

Name of a language without a capital: *Nederlands* 'Dutch as *nederlands* (line 34).

## 7.2 Parodying how they speak by Naar Talia

Sociolinguistic enregisterment takes place through the following co-occurring linguistic forms within and between the level(s) of:

### Lexicon

The vocabulary used might easily be interpreted as obscene by many, such as *zeik* (line 2) 'urine,' *klauwe* (line 2) 'to steal,' *wijve* (line 19) 'chickies,' *klitse* (line 20) 'balls,' *vreete* (line 22) 'to gobble,' *sjcheete* (line 22) 'farts,' *dunne op de plee* (line 23) 'runs (diarrhea) on the loo,' *zuipe* (line 26) 'booze,' *votte* (line 30) 'booty,' *sjnee* (line 31) 'crack' and *me ouwe* (line 42) 'my oldie/my father';

Language jokes with sexual allusions such as in '*heer wie cum*' (English syntax 'here we come' but dialect phonology) and *labia Maggiore* (line 28);

The intensifier *ram* (line 24) 'lit. 'completely,'<sup>21</sup> is clearly part of the linguistic uptake by the local Heerlen audience of the Getske Boys' register. An article in the regional daily newspaper 'De Limburger' of 22 June 2018, titled

21. See <<http://www.mestreechtaol.nl/dictionair/mst/ram/>> (28 April 2020).



“Heerlens woord in Kruidvat-advertentie zorgt voor verwarring” (*‘Heerlen’ word in Kruidvat advertisement causes confusion*), reports on an advertising slogan for the Kruidvat drugstore in the Metro, a free tabloid newspaper distributed nationwide at railway stations and other public places such as schools and libraries: “Iedere week ram veel voordeel” (lit.: ‘Every week *ram* much discount’). While many Dutch readers were at a loss as to how to understand the word ‘ram,’ the regional newspaper De Limburger reported that reactions to the Kruidvat Facebook page show that ‘ram’ is familiar to Heerleners.

Among the reactions, people suspect that a Heerleener is part of the Kruidvat advertising department. “This must have been a Heerleener who came up with this. The word ‘ram’ is common here in Heerlen. It stands for ‘very much,’ ‘huge,’” noted a Facebook user.<sup>22</sup>

The article concludes that “[o]n social media, it is also jokingly suggested that the Heerlen music band Demi Sec is involved in the [making of] the advertisement.”<sup>23</sup>

Another lexeme is indexical for ways of speaking by youth: the intensifier *vet* (line 28) lit. ‘fat,’ very much (possible borrowing: English: ‘phat’).

The use of *slech* in the meaning of ‘ill’ is definitely nonstandard Dutch (lit. bad) but comes from the local dialect

### Phonology

Hyperfrequent use ( $n = 7$ ) of palatalized /s/ as /ʃ/ instead of standard Dutch /s/ at the beginning of a word before the consonants /t/, /l/, /m/, /n/, and /w/. Traditionally, the use of /ʃ/ in this context is a characteristic of the dialects in the eastern part of Limburg (including Heerlen but excluding Maastricht). The Getske Boys produce /ʃ/ in standard Dutch lexemes *sjpaghetti* (line 11), *sjcheete* (line 22) ‘to fart,’ *zjwete* (line 23) ‘to sweat,’ in the English borrowing *sjcore* (line 28), *sjnee* (line 31) ‘crack,’ *sjlech* (line 42) ‘bad, ill,’ *sjteek* (line 43) ‘to light.’ The lexemes ‘sjcheete’ and ‘zjwete’ are part of the refrain (lines 14–27) and hence repetitively used, causing a strengthening effect.

22. “Bij de reacties vermoeden mensen dat er een Heerleenaar in het advertentieteam van Kruidvat werkt. “Dit moet een Heerleenaar zijn geweest die dit bedacht heeft. Het woord ‘ram’ is hier in Heerlen gebruikelijk. Het staat voor ‘heel erg’ of ‘enorm’, reageert een facebook-gebruiker” (De Limburger, 22 June 2018).

23. “Ook wordt er op social media gekscherend gesuggereerd dat de Heerlense muziekgroep Demi Sec bij de advertentie betrokken is.” (De Limburger, 22 June 2018). As noted above (Section 4), Demi Sec and the Getske Boys consist of the same three men.

The linguistics forms *sjcore* and *sjcheete* are unexpected, since the local dialects have no phonological rule that /ʃ/ precedes a /k/ and the fricative /ç/, as in *sjcore* and *sjcheete*, respectively.

Frequent use of word-final /t/-deletion ( $n = 4$ ) following the fricatives [ç] and [s], which is a characteristic of the dialects in the eastern part of Limburg (including Heerlen but excluding Maastricht):<sup>24</sup> *sjlech(t)* (line 42) ‘bad, ill,’ *dich(t)* (line 46) ‘closed’ and *vas(t)* (line 8) ‘already’ and in *dee(d)* (line 7, 35) ‘did.’ The phenomenon of word-final *t*-deletion is widespread; it occurs in the provinces of Gelderland, Brabant, South-Holland, Overijssel and in ‘Holland’ cities such as Utrecht. It is a linguistic form that is ambiguous (cf. Johnstone 2011b); it is a verbal inflection (3 ps, present), but it also indexes place, namely the typical accent of the dialect of Heerlen and elsewhere in Limburg, and/or social class, for it is perceived as uneducated and sloppy.

Svarabhakti vowel in *zellevers* instead of standard *zelf* (line 27) ‘self’.

### Morphology

Adding *-ers* *zellevers* instead of standard *zelf* (line 27)

Plural formation *-e* instead of standard *-s*: *paatere* ‘fathers, Catholic priests’ instead of *paters* (line 46).

### Syntax

#### Copula

In the lyrics, the construction in (1) is a stereotype used to index dialect speakers who are not able to speak proper Dutch, since it is a literal translation from the dialect. In dialect the verb *liggen* ‘to lie’ can be used as a copula in contrast to standard Dutch (RES = resumptive pronoun):

- (1) me ouwe die lag sjlech (line 42)  
 my old one RES lied bad  
 ‘My father was ill’

#### Declarative *do+infinitive*

As shown in Cornips, de Rooij & Stengs (2017), the Getske Boys frequently use the declarative *do+infinitive* construction with the frame NP<sub>1</sub>-*doen*-(NP<sub>2</sub>)-V in which *do* carries all agreement and tense features, as in (2) (see line 7):

24. Phonetically salient characteristics are the maintenance of the difference between [f]/[v] and [s]/[z] (analogous to the [ç]/[j] case) in the south, whereas many northern speakers do not have this phonological opposition (Hagen, A. & H. Giesbers 1988).

- (2) Ken je Genaro Clementaro die (...) ijsjes draaien dee (lines 4–7)  
 know you Genaro Clementaro RES ice-creams turn did  
 ‘Do you know Genaro Clementaro who made ice cream’

The use of this type of construction is heavily stigmatized at the national and local level. The *do*+infinitive was never part of standard Dutch, although it occurs in adult dialect and regional varieties in the Netherlands, with the exception of the northern provinces (cf. Cornips 1998; Barbiers et al. 2008: 41). The absence of *doen*+infinitive in standard Dutch (and in German) is very likely “due to a conscious process of purification of the language in which it was purged of its supposedly illogical or superfluous aspects (...). So (...) is kept from spreading into the most prestigious spoken variant by a deliberate act” (Auer 2004: 74).

Thus, the Getske Boys exaggerate and ridicule standard Dutch in their repetitive use of *doen*+infinitive, and they transgress local norms in using it in an unexpected manner, i.e., not only expressing habitual aspect. In their song, the Getske Boys make a parody of the use of *doen*, and by doing so they make a parody of those who live and speak in Heerlen, and of Heerlen as a Limburgian locality.

### Hybrid use

#### *Dutch and dialect*

The Getske Boys sing in Dutch with massive interference from the local dialect. For instance, they insert local dialect forms: *klitse* (line 20) ‘balls,’ *hendig* (line 23), *zellevers*, *hei* (line 29) ‘have,’ and *votte* (line 30) ‘booty.’

#### *English and Dutch*

The name the Getske Boys is a hybrid between dialect and English. More hybrid use between English and Dutch is ‘*heer wie cum*’ with English syntax, since Dutch would require the finite verb ‘cum’ to follow the adverb ‘heer’ as in: ‘heer cum wie.’ The sentence shows dialect phonology [he. wi. kʏ m]; in Dutch it would be [hi. <sup>R</sup>wə ko.mə].

#### *Italian, Dutch and dialect*

In (3) forms associated with dialect, i.e., *votte* (in bold) ‘booty’ co-occur with forms associated with Italian *buona not[t]e* ‘good night’ (underlined) and Dutch *lekkere* ‘hot’. *Not[t]e* rhymes with *votte*:

- (3) Buona not[t]e, lekkere **votte** (line 30)

#### *Italian and Dutch*

- (4) Vroeg aan Sophia: O cara mia (lines 32/33)  
 ‘(I) asked Sophia: oh, my darling’

In (5) the infinitives *scusiare* (see fn 19) and *parlare* are sung instead of the expected inflected *mi scusi*, *non parlo*. The use of infinitives instead of inflected verbs resembles the phenomenon of foreigner talk. (5) also with Italian associated *niente* ‘nothing’ shows up:

- (5) “Mi scusiare, non parlare” dus daar ken jeeeeeeeeeeee...Dus niente mee!  
 ‘I apologize, I don’t speak thus this is beyond all practicalities’

(lines 36 through 41)

### 7.3 Co-occurrence of unexpected forms and repetition

From the above, it is clear that sociolinguistic enregisterment takes place at and within all possible linguistic levels. The same holds for unexpected forms, the most striking of which are the phonological palatalized /ʃ/ in *sjchete* ‘farts’ and *sjcore* ‘score’ in contexts that are unusual in the eastern Limburg dialects, word final /t/-deletion after a long vowel in *dee(d)* ‘do’ instead of after fricatives, which is the phonological rule in the eastern dialects, the use of *do*+infinitive construction not expressing habitual aspect, the ‘illiterate’ way of spelling, and using adverbs such as *ram* ‘a lot’ as intensifiers, and the hybrid use of linguistic forms associated with different varieties such as Dutch, dialect, Italian and English. Moreover, the co-occurrence of these features all comes together in the refrain that is repeated three times. The features are used so often that it is clear that one is listening to an exaggeration that works to establish a caricature of people in Heerlen-Noord.

## 8. Conclusion

What is happening in the Getske Boys’ performances shows striking similarities to Johnstone’s analysis of re-enregisterment of linguistic features as Pittsburghese (Johnstone 2011b):

In Pittsburgh, speech forms that had previously been enregistered as incorrect and/or working class were re-enregistered, starting in the 1960s, with place (Johnstone 2013), as Pittsburghers became mobile in new ways and economic change meant that they needed to find new identities.

In a comparable way, our case study has shown how particular linguistic forms are being re-enregistered as local Heerlen-Noord speech. The Getske Boys’ selection of these specific co-occurring forms is based on perceived past patterns of co-occurrences: an experiential knowledge, accumulated over the years, of the indexical ties between linguistic forms, specific (groups of) people and a specific place (Taylor n.d.). The stylized language use of the Getske Boys thus builds on earlier enregisterments, bringing into being a more focused register characterized not only by specific linguistic forms but by co-occurring topics, behaviors, and ideological positionings. Through their stylized and highly marked language use, they also project a highly marked persona, or characterological figure (Agha 2005), allowing the

performers and their audience to rebel against a mainstream which they consider inauthentic and to identify with what is experienced as authentically local.

Such processes of (re-)enregisterment are to be considered as constantly shifting depending on the “particular junctures in the lives of individuals and the histories of communities” (Johnstone 2011b: 641). Popular artists such as the Getske Boys can, through their presence in new media and their performances in public events such as carnival, have a strong influence on processes of enregisterment and, hence, in the creation of new local identities. Given the ever-increasing use of new media, the study of enregisterment in and through these media should become a priority in sociolinguistic research.

## References

- Agar, M. 1994. *Language Shock: Understanding the Culture of Conversation*. New York NY: William Morrow and Company.
- Agha, A. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1): 38–59.
- Auer, P. 2004. Non-standard evidence in syntactic typology: Methodological remarks on the use of dialect data vs. spoken language data. In *Dialectology Meets Typology: Dialect Grammar from a Cross-linguistic Perspective*, B. Kortmann (ed.), 69–92. Berlin: De Gruyter.
- Antonsich, M. 2010. Searching for belonging – An analytical framework. *Geography Compass* 4(6): 644–659.
- Androutsopoulos, J. 2006. Multilingualism, diaspora, and the internet: Codes and identities on German-based diaspora websites. *Journal of Sociolinguistics* 10(4): 520–547.
- Bakhtin, M. 1984 [1968]. *Rabelais and his World*. Bloomington IN: Indiana University Press.
- Barbiers, S., van der Auwera, J., Bennis, H. J., Boef, E., De Vogelaer, G. & van der Ham, M. 2008. *Syntactische atlas van de Nederlandse dialecten Deel II* (Syntactic atlas of the Dutch dialects Vol. II). Amsterdam: Amsterdam University Press.
- Beal, J. 2009. “You’re Not from New York City, You’re from Rotherham” Dialect and identity in British Indie music. *Journal of English Linguistics* 37(3): 223–240.
- Bucholtz, M. & Hall, K. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies* 7(4–5): 585–614.
- Christensen, A.-D. 2009. Belonging and unbelonging from an intersectional perspective. *Gender, Technology and Development* 13(21): 21–41.
- Cornips, L. 1994. *De Syntactische Variatie in het Algemeen Nederlands van Heerlen*. PhD dissertation, University of Amsterdam.
- Cornips, L. 2013. Recent developments in the Limburg dialect region. In *Language and Space – Dutch: An International Handbook*, F. Hinskens & J. Taeldeman (eds), 378–399. Berlin: De Gruyter.
- Cornips, L. & de Rooij, V. 2015. Belonging through languagecultural practices in the periphery: The politics of carnival in the Dutch province of Limburg. *Anthropological Journal of European Cultures* 24(1): 83–101.
- Cornips, L. & de Rooij, V. 2019. Katanga Swahili and Heerlen Dutch: A sociohistorical and linguistic comparison of contact varieties in mining regions. *International Journal for the Sociology of Language* 258: 35–70.

- Cornips, L., de Rooij, V. & Smakman, D. 2017. The Randstad area in The Netherlands. Emergent and fluid identity-locality production through language in use. In *Urban Sociolinguistics: The City as a Linguistic Process and Experience*, D. Smakman & P. Heine (eds), 162–180. London: Routledge.
- Cornips, L., de Rooij, V., & Stengs, I. 2017. Carnavalesque language use and the construction of local identities: A plea for languageculture as a field of research. *Jahrbuch für Europäische Ethnologie/Yearbook of European Ethnology* 12: 61–89.
- Cornips, L., de Rooij, V., Stengs, I., & Thissen, L. 2016. Dialect and local media: Reproducing the multi-dialectal hierarchical space in Limburg (the Netherlands). In *Style, Media and Language Ideologies*, J. Thøgersen, N. Coupland & J. Mortensen (eds), 189–216. Oslo: Novus Press.
- Cornips, L. & Knotter, A. 2017. Inventing Limburg (The Netherlands): Territory, history, language, and identity. In *Räume, Grenzen, Übergänge, Akten des 5. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD)* [ZDL. Beihefte. 170], H. Christen, P. Gilles & C. Purschke (eds), 71–91. Stuttgart: Steiner.
- Driessen, G. 2006. Ontwikkelingen in het gebruik van streektaalen en dialecten in de periode 1995–2003. *Toegepaste Taalwetenschap in Artikelen* 75: 103–113.  
<https://doi.org/10.1075/ttwia.75.10dri>
- Eckert, P. 2011. *Fractals all the way down*. Lecture at NIAS symposium ‘Constructie van lokale identiteit door taal en cultuur’, Wassenaar, 1 December.
- Evans-Pritchard, E. E. 1940. *The Nuer: A Description of the Modes of Livelihood and Political Institutions of a Nilotic People*. Oxford: Clarendon Press.
- Extra, G. 2004. Taal en identiteit. De vitaliteit van minderheidstalen in vergelijkend perspectief. *Taal en Tongval* 56(17): 109–134.
- Gal, S., & Irvine, J. 1995. The boundaries of languages and disciplines: How ideologies construct difference. *Social Research* 62(4): 967–1001.
- Hagen, A. & Giesbers, H. 1988. Dutch sociolinguistic studies. *International Journal of the Sociology of Language* 73: 29–44.
- Hanks, W. 1996. *Language and Communicative Practices*. Boulder CO: Westview Press.
- Hinskens, F. & Taeldeman, J. 2013. Introduction. In *Language and Space: Dutch: An International Handbook*, F. Hinskens & J. Taeldeman (eds), 1–12. Berlin: De Gruyter.
- Johnstone, B. 2011a. Language and place. In *The Cambridge Handbook of Sociolinguistics*, R. Mesthrie (ed.), 203–217. Cambridge: CUP.
- Johnstone, B. 2011b. Dialect enregisterment in performance. *Journal of Sociolinguistics* 15(5): 657–679.
- Johnstone, B. 2016. Enregisterment: How linguistic items become linked with ways of speaking. *Language and Linguistics Compass* 10: 632–643.
- Kats, J. 1952. De dialecten van de mijnstreek. In *Mijn en spoor in goud. Jubileumuitgave*, M. Kemp (ed.), 316–318. Maastricht.
- Kessels-van der Heijde, M. 2002. *Maastricht, Maestricht, Mestreech* [Maaslandse Monografieën 65]. Hilversum: Verloren.
- Knotter, A. 2009. Limburg bestaat niet. Paradoxen van een sterke identiteit. In *Dit is Limburg! Opstellen over de Limburgse Identiteit*, A. Knotter (ed.), 263–277. Maastricht: Sociaal Historisch Centrum voor Limburg & Zwolle: Waanders.
- Knotter, A. 2011. *Regional character and regional stereotypes: Some thoughts about their construction in Dutch Limburg*. Lecture at NIAS symposium ‘Constructie van lokale identiteit door taal en cultuur’. Wassenaar, 1 December.

- Leppänen, S. 2012. Linguistic and generic hybridity in Web writing: The case of fan Fiction. In *Language Mixing and Code-switching in Writing: Approaches to Mixed-language Written Discourse*, M. Sebba, S. Mahootian & C. Jonsson (eds), 233–254. London: Routledge.
- Massey, D. 1993. Power geometry and a progressive sense of place. In *Mapping the Futures: Local Cultures, Global Change*, J. Bird, B. Curtis, T. Putnam & L. Tickner (eds), 59–70. London: Routledge.
- Mathijssen, M. 2011. *De Limburger als de ander*. Echt: Letterkundig Centrum Limburg.
- Milroy, J. 2001. Language ideology and the consequences of standardization. *Journal of Sociolinguistics* 5: 530–555.
- Münstermann, H. & Hagen, A. 1986. Functional and structural aspects of dialect loss: A research plan and some first results. In *Language Attrition in Progress*, B. Weltens, K. de Bot & T. van Els (eds), 75–96. Dordrecht: Foris.
- Pietikäinen, S. & Kelly-Holmes, H. 2013. Multilingualism and the periphery. In *Multilingualism and the Periphery*, S. Pietikäinen & H. Kelly-Holmes (eds), 1–16. Oxford: OUP.
- Preston, D. R. 2013. Language with an attitude. In *The Handbook of Language Variation and Change*, J. K. Chambers & N. Schilling (eds), 157–182. Malden, MA: Wiley-Blackwell.
- Silverstein, M. 1985. Language and the culture of gender. In *Semiotic Mediation*, E. Mertz & R. Parmentier (eds), 219–259. New York NY: Academic Press.
- Silverstein, M. 1996. Monoglot 'Standard' in America: Standardization and metaphors of linguistic hegemony. In *The Matrix of Language: Contemporary Linguistic Anthropology*, D. Brenneis & R. K. S. MacAulay (eds), 284–306. Boulder CO: Westview Press.
- Stuyck, K., Luyten, S., Kesteloot, C., Meert, H. & Peleman, K. 2008. A geography of gender relations role Patterns in the context of different regional industrial development. *Regional Studies* 42(1): 69–82.
- Taylor, C. n.d. Hip Hop parody as veiled critique. Unpublished paper (to appear in *African American Language in Pop Culture*). <[http://www.ruf.rice.edu/~sastre/HipHopParodyAs-VeiledCritique\\_chris-taylor.pdf](http://www.ruf.rice.edu/~sastre/HipHopParodyAs-VeiledCritique_chris-taylor.pdf)> (21 February 2017).
- Thissen, L. 2013. The ambiguities of Limburgerness: Language, place, and belonging in Limburg, the Netherlands. *Etnofoor* 25(2): 119–143.
- Thissen, L. 2018. Talking in and out of Place. Ethnographic Reflections on Language, Place, and (Un)belonging in Limburg, the Netherlands. PhD dissertation, Maastricht University.
- Thomason, S. G. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- Wang, X., Spotti, M., Juffermans, K., Cornips, L., Kroon, S. & Blommaert, J. 2014. Globalisation in the margins: Toward a re-evaluation of language and mobility. *Applied Linguistics Review* 5(1): 23–44.
- Watson, J. 2013 [1994]. On the definition of dialect with reference to Yemeni dialects of Arabic. In *Arabic Sociolinguistics. Issues and Perspectives*, Y. Suleiman (ed.), 237–250. New York NY: Routledge.
- Weijnen, A. 1967. Sociodialectische onderzoekingen in Limburg. *Taalsociologie* 32: 16–31.
- Wijers, C. 1995. *Prinsen & clowns in het Limburgse narrenrijk: Het carnaval in Simpelveld en Roermond 1945–1992*. Amsterdam: P.J. Meertens-Instituut.
- Woolard, K. 2008. Why dat now? Linguistic-anthropological contributions to the explanation of sociolinguistic icons and change. *Journal of Sociolinguistics* 12(4): 432–452.
- Yuval-Davis, N. 2006. Belonging and the politics of belonging. *Patterns of Prejudice* 40(3): 197–214.

# Snow on the Danish Antilles?

## Referee design in Virgin Island Dutch Creole

Cefas van Rossem

Meertens Institute, Royal Netherlands Academy  
of Arts and Sciences<sup>1</sup>

One of the things one does not want to hear when working on a large corpus, is that the content is very artificial, and should be ignored in your research because of the unnatural elements it contains. This is what happened with the Clarin-NEHOL-Corpus of Virgin Islands Dutch Creole. The contents, mainly eighteenth-century missionary texts were considered by some people as ‘just’ a missionary variety which seemed very unlikely to have been used in daily life. Clearly, a theoretical basis was needed to analyse this variety in order to establish the authenticity of these texts. Unexpectedly, Bell’s 1984 Audience Design Model, originally based on spoken language situations, turned out to be ideal for the treatment of older written material. One element of this model, referee design, seemed at first to stand somewhat separate from the other aspects of the theory. However, it enabled us to understand the communication situation which missionaries and their pupils participated in. This article focuses then on referee design as a tool to study eighteenth century Virgin Islands Dutch Creole in particular, and historical Creole texts in general.

**Keywords:** audience design, missionary linguistics, historical sociolinguistics, Virgin Islands Dutch Creole, Clarin-NEHOL-Corpus

### 1. Authenticity: Reliable written representation of spoken language<sup>2</sup>

In 1991, a large number of copies of German microfilms arrived in the Institute of General Linguistics of the University in Amsterdam. Hein van der Voort and I started to enter these texts into a digital database, under the supervision of Pieter

---

1. Everywhere where *he*, *him* or *his* was used to refer to people in general, you can also read *she* or *her*.

2. In this text *authenticity* is used to indicate whether the written Creole texts are a reliable representation of the pure eighteenth century Creole vernacular (see for instance Hernández-Campoy



Muysken and Hans den Besten, financed by both Netherlands Organisation for Scientific Research (NWO), and the Royal Netherlands Academy of Arts and Sciences (KNAW). This database has the potential to become the largest database of historical Creole texts. At the moment, it contains about 3500 pages of mainly missionary translations, i.e. an Old Testament, four Gospel Harmonies, several hymn books and sermons.<sup>3</sup> As Jacques Arends and Adrienne Bruyn had already worked with historical Sranan missionary texts at our Institute in 1991, we had the opportunity of discussing methodology and manuscripts from both the Danish Antilles and Surinam. I will return to the Surinamese texts later on in this article.

While taking every aspect of these texts into consideration, one thought also crossed our minds. Who were the actual writers of these texts? How much did these people know about the language which they were writing in and translating to? However, since there was a large number of pages in Virgin Islands Dutch Creole, it was hard to imagine that these missionaries were working in a language which was alien to their audience, or which their audience found difficult to understand. We kept on typing and encoding, and gained interesting insights into these earliest sources of this Dutch-based Creole language, which have been studied since the dawn of Creole linguistics.

Nevertheless, we received some critical comments from our peers. Some comments left an opening for discussion. John McWhorter (1997: 174–178) sketched an ironic image of both the Sranan and the Virgin Islands Dutch Creole corpus by showing that the use of very different, incomplete snippets of a language could only lead to an incomplete idea of the early stages of these languages. In her extensive work on Virgin Islands Dutch Creole, Robin Sabino (2012) stated that these texts were of interest, but should not be used to study this Creole in a diachronic fashion. The texts were written down by missionaries of whom we are ignorant of how they learned and used the language, and of whom it is sometimes said that their only goal was to convert the enslaved Africans as quickly as possible. See for instance the considerations of Labov (1994: 11). In order to prove the authenticity of our corpus, in other words to answer the question of whether the written language used in the corpus was a good representation of the Creole actually spoken in the Danish Antilles in the eighteenth century, we needed to find convincing arguments.

---

& Schilling 2012: 68, in Hernández-Campoy & Conde-Silvestre 2012: 65–81). I do not refer to the concept *linguistic authenticity* which is discussed in (historical) sociolinguistics (for instance in Lacoste, Leimgruber & Breyer 2014).

3. The digital material is included in the Clarin-NEHOL database, and is available in The Language Archive of the Max Planck Institute. All written material is archived in the Meertens Institute (KNAW). Lists and introductions to the corpus can be found in Van Rossem & Van der Voort (1996) and Van Rossem (2017).

In our anthology *Die Creol Taal* (Van Rossem & Van der Voort 1996: 31) we presented three questions related to the alleged artificiality of the language of these early texts:

- a. Audience design: was the material meant for a predominantly white urban population, or for the plantation slaves?
- b. Linguistic competence and procedure: how well did the translator know Virgin Islands Dutch Creole, and which variety did he know? How were native speakers involved in the translation process?
- c. Translation practice and style: it is clear that sometimes no attempt was made to adjust to the spoken language, e.g. when the Latin accusative *Jesum* appears in the texts. Did the missionaries attempt to create a separate liturgical register, suitable for the conveyance of religious feelings and ideas?

The first two questions are related to a situation in which a translator focuses on his audience, on the people close to him, to convert, to teach. In this case the translator needs to connect with his audience in order to get his message across and encode his message in a comprehensible way. The third question, however, illustrates another perspective. In terms of several elements found in our texts, the message also seems to require a liturgical encoding or style. In fact, some elements in Virgin Islands Dutch Creole texts must have been absolutely unclear and incomprehensible to the audience of enslaved people. This missionary jargon or ‘liturgical register’ can hardly be derived from their vernacular, from spoken language or even from the variety of Dutch used on the Danish Antilles. Take for instance the use of *Jesum* quoted above, which will be discussed more fully in a later paragraph.

In the nineteenth century the missionary variety was already under discussion. Sabino (2012: 93) quotes Pontoppidan (1887: 297):

When compared to *Negerhollands*,<sup>4</sup> the evangelical variety was artificial “so that the Negroes themselves barely understood and in some places [it] seemed more like a bad Dutch.” It seems as if it excludes African elements, it lacks vitality and authenticity, it very little if any ... is the slaves’ language and Mrs. Stevens<sup>5</sup> more than once lamented her unfamiliarity with this variety. (Sabino 2012: 85)

However, a multi-stage study, in which emendations in the Creole manuscripts are studied in relation to the different groups in the audience of the Moravian translators, as is described in my dissertation (Van Rossem 2017), will show that

---

4. Sabino (2012) uses the name *Negerhollands* to indicate the spoken variety of Virgin Islands Dutch Creole, which was spoken until the last speaker, Mrs. Alice Stevens, passed away in 1987.

5. In interviews, historical texts or stories from De Josselin de Jong (1926) were presented to Mrs. Stevens.

a philological approach to these texts provides us with a more complete picture of eighteenth-century Virgin Islands Dutch Creole.

The first step is to determine which language, or more to the point, which variety of Dutch or Dutch Creole was spoken in the early years of the Danish colony. The early sources of Virgin Islands Dutch Creole seem to be too Dutch-like, and do not resemble actual spoken Creole as attested in twentieth-century sources. However, this may in fact not have been due to the influence of the missionaries on the first written texts. The demography of the early Danish colony shows how important influence of Dutch speakers was, using this language not only because of their ancestry, but also as a lingua franca or even koine (Van Rossem 2013). The more equal the ratio of slaves to colonists was, the more unlikely was the rapid emergence of a Creole, and the longer it would take to integrate a Creole into the communicative situation of a colony (Chaudenson 1992; Van Rossem 2017: 50). Besides, a colonists' version of the Creole is already mentioned in the earliest published study of the language, Magens (1770). This means, that in this particular colony three Dutch-related languages were used: Dutch as a lingua franca/koine, a Dutch-based colonists' version of the Creole (*Hoch Kreol* as it is called by Pontoppidan 1881), and a Dutch-based version of the Creole which was spoken by the population of slaves and their families. These groups all influenced the language that was written down in the eighteenth-century material we study nowadays.

The next step is the study of the use of metalinguistic comments related to the use of Dutch and Dutch Creole in the Danish Antilles in order to determine which variety was considered the most useful for missionaries to use with their pupils, the slaves of African heritage. In the early years of the Moravian Brethren's mission standard Dutch was used. The first missionaries used a Dutch New Testament and Dutch primers, probably Sewel's *Nederduytsche Spraakkonst* (1708).<sup>6</sup> A few years later, metalinguistic comments show that Creole is used, and is considered to be better suited to relate to the audience of slaves. Slave letters, the earliest texts containing Virgin Islands Dutch Creole elements were initially in Dutch. Although we know of Creole texts dating back as early as 1739/1742, the first missionary texts were not written in Virgin Islands Dutch Creole until 1749–1753 (Isles & Weber 1749–1753).

The next step is focusing on the Creole texts themselves in relation to their audience. At first glance, the variety which is used in these texts hardly differs from the missionary texts which were written and translated at the beginning of the nineteenth century. Why? Did these texts connect well with their audience? Could the congregation understand their preachers? Metalinguistic comments show that

---

6. Perhaps also Francis Harrison, *De Engelsche en Nederduytsche School-Meester* (...) New York: 1730. See Van Rossem (2017: 69).

the texts were too difficult; instructions and expressions were unclear, words were interpreted differently by the slaves and hymns were too long to be truly comprehensible,<sup>7</sup> the community of missionaries, however, was doing its best to connect with its audience. Unfortunately, we do not have texts which were written by native Creole speakers without the apparent influence of their missionary teachers.

The missionaries who were willing to improve their texts to suit the Creole speakers better, provide the sole source of information in determining the authenticity of the texts. Luckily, all textual emendations were encoded while digitalizing the NEHOL-corpus (Van Rossem & Van der Voort 1996: XII–XIII). However, how can we distinguish these improvements from other changes in the texts? The corpus contains several varieties of the same text, for example from the Gospel Harmony, or from hymns. When comparing these texts, it appears that the earliest of these texts are more ‘Creole-like’.<sup>8</sup> Strangely enough, it is not just the number of English-related elements that seems to be increasing – which is only to be expected from a historical point of view, since English began to replace Dutch and Dutch Creole on the Danish Antilles at the end of the eighteenth century – but also the number of Dutch-like missionary words and expressions. One would expect this number to decrease. At first glance there seems to be no system: Creole words are changed into Dutch ones; some sentences are changed into svo-order, but not all; Dutch expressions are changed into Creole, etc. Which variant should be considered the authentic one in such cases? To create order in this apparent state of chaos, a clear model was needed, one in which all the various linguistic influences and styles within the texts could be analysed separately.

## 2. One model to clarify all

In 1984, Allan Bell presented his *Audience Design model* in which he clearly describes the composition of the audience, and the language which is used by an author to address that audience. This model was further elaborated in Bell 1992, 1997 and 2001. Although initially based on spoken situations, it can be used to reconstruct the communication situation in earlier days, based on written texts. In short, we can distinguish the following participants (see Figure 1 on p. 375):

---

7. In a letter by translator Johann Böhner to the eldest of the Moravian mission (St. Croix, May 11th, 1773).

8. Whatever this may be. Characteristics of Creole languages in the Atlantic which are generally taken into account discussing the ‘Creoleness’ of a text are: svo-word order, the use of particles for tense, mood and aspect, preverbal negation, serial verb constructions, the use of *fo(r)* ‘in order to’, the use of the multipurpose preposition *na* and the use of paragogic vowels/vowel harmony.

- a. **SPEAKER/AUTHOR.** This is the first person and, according to Bell (1984: 159) 'Primary participant at the moment of speech [...] However, speakers design their style for their audience. Differences within the speech of a single speaker can be accounted for by the influence of a second person, and some third persons, who together compose the audience to a speaker.' In written language we see no direct two-way communication.<sup>9</sup> The author has the opportunity to edit older texts, and to take the comments of his audience into account in subsequent texts and translations. In the case of eighteenth-century Virgin Islands Dutch Creole, the authors are the writers/translators of the Moravian Brethren, mainly Johann Christoph Auerbach, Johann Böhner, Samuel Isles, Johannes Loretz and Georg Weber.<sup>10</sup>
- b. **ADDRESSEE.** Bell calls this second person in his model the main character in the audience (Bell 1984: 159). This person is known, ratified and addressed. This addressing becomes visible when the names of these men and women are used in letters, historical records or even in lists of names. In the eighteenth century community of the Moravian Brethren of the Danish Antilles, some African slaves were the so-called *helpers* of the missionaries, intermediaries between the brethren and the other slaves who were not yet part of the community.<sup>11</sup> There is some biographical background available on some of these people, partly because they were mentioned by Oldendorp in his extensive history of the Moravian mission on the Danish Antilles (Oldendorp 2000, 2002), but also because they actually were the most important writers of the so-called *slave letters*. During the early mission, there is for instance Pitrus (the writer of the first known slave letter, 1738), Domingo Gesoe (who also translated texts into Creole for Count von Zinzendorf) and Abraham. In the second half of the eighteenth century, the helper Cornelius plays an important role. He not only writes a lot of letters, he also introduces common Virgin Islands Dutch Creole grammatical elements which are later also used in texts written by other slaves, like perfective *ka* (Stein 1985: 448–452). The most interesting feedback from a helper is referred to in a letter of Johann Böhner (August 2, 1781):

---

9. Metalinguistic comments in letters and notes show that linguistic feedback from both *helpers* and missionaries was presented to the writers.

10. All these are mentioned, accompanied by biographical information, in Oldendorp (2000, 2002).

11. Extremely interesting in this regard is ms. 335 in our Clarin-NEHOL-Corpus, in which the tasks of these helpers are described, and the list of names of the helpers (ms. 127) in which even their linguistic competence is mentioned.

So a local gentleman, who is a native of St. Thomas, and [therefore] a Creole, but who has fallen into extreme poverty, has translated the New Testament into the Creole language, which was printed in Copenhagen in the first quarter of 1781, [...] It was not well-received, it was too much like Danish, and otherwise very inaccurate, and *our helper Brother Cornelius does not like it at all, and he wishes that we could at least have the New Testament printed in their language, because they [think] my translation is clear and appropriate to speaking their dialect.* [my italics, CvR] So everyone who is able to read and who owns a hymn booklet, can buy a New Testament to it. It would be a great pleasure for those who are able to read.

Sources like Oldendorp (2002) and the slave letters provide information as to which languages these addressees used. Their knowledge of languages was often not limited to Creole, but extended to Dutch, German and Danish.<sup>12</sup>

- c. AUDITORS. The third person is the group of *auditors* (Bell 1984: 159), who are known and ratified by the community, but are not directly addressed. These people read and listen to the texts; however, the author is not fully aware of which variety connects best with them. In my opinion this group did not just consist of slaves, but also of brethren of European descent, who we know were present at community meetings. In the early days of the mission Dutch was used, even by the German missionaries who had to learn it during Atlantic crossings. There is no evidence of Creole missionary texts dated before 1749, and we suspect these did not exist. Several metalinguistic comments suggest that only a few brethren were sufficiently capable of preaching in Creole, and that Creole was best suited to connect with the audience of slaves. It is therefore likely that the Creole which was used, had to be understandable for both people of African and of European descent. Since we know a missionary variety of Sranan existed, namely Church Sranan (Voorhoeve 1957, 1971), there is a distinct possibility of a comparable Dutch Creole having existed on the Danish Antilles.
- d. OVERHEARERS. According to Bell, the next group of third persons consists of the *overhearers*. These are the ones who are known, but not (yet) ratified as members of the community of Moravian Brethren. This group may have consisted of enslaved African people and Europeans, who were outside of the community of Moravian Brethren, but who were probably within their range of influence. Bell (1984: 176) mentions the following characteristics to distinguish the style used towards the overhearers from other third persons: “Overhearer design can

---

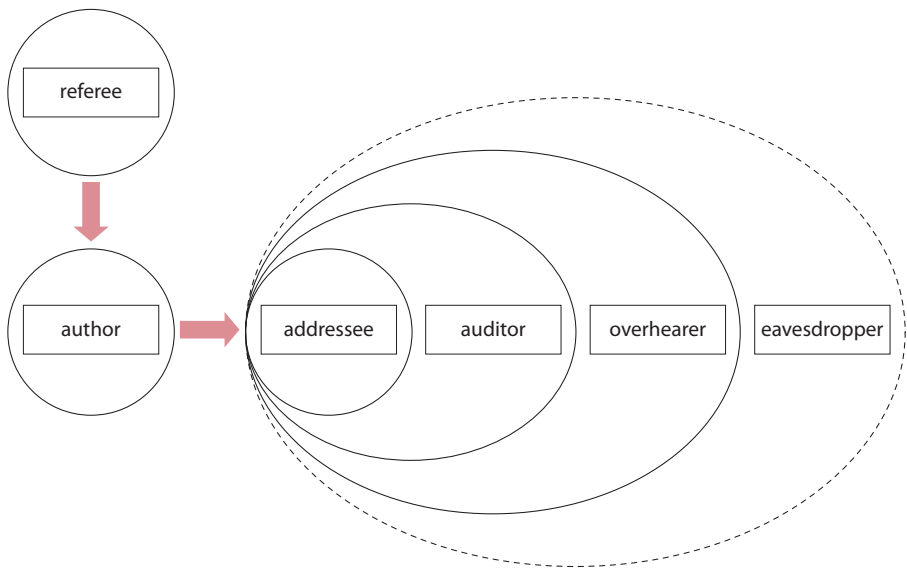
12. Only recently I received a copy of a list of helpers (1760) which is preserved in the Unitäts Archiv in Herrnhut. It presents biographical information, including the status of the slaves. Some were for instance bombas, ‘enslaved overseers’ or freed people of African descent. One helper’s competence in using Creole is even discussed!

still be manifested in qualitative language choices such as politeness-marked pronoun selection, speech act design, and bilingual language shift.” Since pronouns of politeness are absent in Virgin Islands Dutch Creole, and since the corpus consists of written material, it can only be used to study bilingual language shifts. Further research should focus on searching for Creole words or expressions which are accompanied by a Dutch, or even clearer, a German or English translation. Such examples are however very difficult to find, since additions which do not seem to be Creole, often, if not always, refer to liturgical or missionary jargon.

- e. **EAVESDROPPERS.** The final group of third persons consists of the *eavesdroppers*. However, it is quite hard to consider this group part of the audience. Eavesdroppers are not ratified and not personally known. According to Bell, the speaker/author is not even aware of their presence at the moment of communication. With regard to the corpus, the local authorities, and planters who are not related to, and possibly even opposed to the Moravian Mission, can be considered to be eavesdroppers. These were probably absent in meetings, but their opinions on proper language and content may have been of influence on the contents of religious instruction during meetings and during the process of writing (Van Rossem 2017: 133–134). An example of eavesdroppers’ design might be the following text from one of the Gospel Harmonies: ‘I will skip the section about fasting, since this is a subject which is not common among the Negro people, just like among the Copts in Egypt.’ (321: 52).

Bell introduces his model in a concentric fashion (Bell 1984: 159), the author being the middle and the eavesdroppers being the outermost sphere. Ladegaard (1995: 99–100) presents an alternative model in which the speaker is in the middle of a triangle of which addressee, auditor and overhearer are the points. According to Ladegaard, the process of accommodation is a more dynamic process than is suggested by Bell’s model. During the period of communication, the speaker continuously determines who the main character in the audience is and accommodates to that person. One way of indicating that the author is not part of the audience is depicted in Figure 1, which shows the author standing next to his audience directing his texts towards it. Since the source material consists of written material, and therefore the interaction is not as direct as in conversations, I prefer to represent the audience in a slightly remodeled concentric model.

One element in Bell’s model that has not been mentioned yet, is the referee, which does not seem to be exclusively a part of the audience, but appears to influence the style of the author from beyond the speech situation. In referee design the author does not accommodate to his audience, but takes the initiative to use language related to a social group outside of the audience, which is considered to be



**Figure 1.** Bell's 1984 Audience Design Model adapted by Van Rossem (2017: 124)

of importance for (future) correct language use with the audience. Bell (1992: 328) writes: 'In referee design, speakers diverge away from the style appropriate to their addressee and towards that of a third party, a reference group or model. Referees are third persons not physically present at an interaction but possessing such salience for a speaker that they influence language choice even in their absence'.

With regard to the communication situation of eighteenth-century Virgin Islands Dutch Creole, Bell's referee design needs some addition, since it looks as if the initiative not to accommodate to the audience, but to a social group outside the audience, is not only the author's. It seems as if the language and jargon used in translations and other missionary texts is also determined by a translation *tradition*, in this case that of the Moravian Brethren, or of the 'ingroup' of local missionaries. In other words, the referee is not so much a person or institution as such, but rather a tradition in which the translator needs and/or wants to work. Since the use of the jargon of the referee and of other missionary aspects (like references to biblical verses and the selection of texts to be translated) are alien to the Creole addressees, I refer to Bell's *ingroup* referee design. The suggestion by Hinskens (p.c.) to distinguish theological referees (translation tradition) from linguistic referees (who are present in the communication situation) is very interesting, and should be studied in the future.

In my graphic representation I put the referee next to the author and not in the audience, for the following reasons: a referee is not (always) a person, but should be



regarded as a tradition which influences the author. It is not a physical part of the audience. Moreover, I suppose the influence of the referee on the author is much more dominant than the reverse. However, an important writer/translator can of course be a guiding influence for his colleagues.

In the search for authentic Virgin Islands Dutch Creole, the *addressees*, the Creole helpers of the missionaries, provide metalinguistic comments. The texts are edited to connect best with *auditors*, the Creole speaking community, and notes in a second language are added for the group of *overhearers*, who are not yet ratified as members of the Moravian community and may not be fluent in Creole yet, and thematical items are included or excluded to connect with *eavesdroppers*, local authorities or planters who might be critical about language and contents.

The influence of the *referee* differs. As far as we can see at the moment, all the above-mentioned changes lead to a better understanding of the texts, while changes made because of referee design seem to have an educational or referential purpose. In case of addressee or auditor design, the text is altered in the direction of the Creole speaking community, and many typical Creole elements are added. The elements which were included in the texts because of referee design seem to have made them more European, more in line with missionary jargon and, here it is more artificial, and obscures the natural eighteenth-century Danish Antilles communication situation of the congregation of Moravian Brethren.

### 3. Referee design

A study of the use of the Virgin Islands Dutch Creole texts by colonists or slaves, (a) in Van Rossem & Van der Voort (1996: 31), shows that metalinguistic comments in particular reveal that the major Virgin Islands Dutch Creole missionary texts were written to be read to an audience of slaves. None of the related texts contained a reference to an audience of white urban population. A clear reference to the use of printed work to be read on their own by people of African origin can be found in Anonymous (1836: 34):

The receipt of the edition of 2000 copies of the *Creole Harmony of the Gospels*, published by the Society, is thus gratefully acknowledged by Rev. J. Boenhof, Moravian Missionary in the Danish West India Islands: “With the fair and handsome printing of this beautiful little work [he says] we are more delighted than I can express. Now not only can the missionary read in public with far greater convenience from the printed page, but what is far more important, we are able to give a free copy to all those of our colored population who have been taught to read, and they may peruse it at home for their edification (...).”

For further information on the linguistic competence of the Virgin Islands Dutch Creole translator, and on the role of native speakers, (b) in Van Rossem & Van der Voort (1996: 31), see Van Rossem (2017, Chapter 5 *Metalinguistic comments*).

The remarks on translation practice and style, (c) in Van Rossem & Van der Voort (1996: 31), however, touch upon the main question of this article: why did translators sometimes not even attempt to connect to the vernacular of their audience? Was the use of a missionary jargon of such importance that even case endings of proper names (*Jesum*) were left unchanged in translations? This last example leads us back to the *Referee design* in Bell's model.

The main influence on the language used by the missionaries, was that of their *referees* (European members of the Moravian Brethren/Lutheran missionaries who were familiar with the religious context and conventions of missionary translations). Several emendations which were made in the eighteenth-century Virgin Islands Dutch Creole, but also in Saramaccan/Sranan translations of the Moravian Brethren, seem to be of no help to the addressees and auditors of African/Creole descent. For instance: why would a translator add a reference to another biblical verse to a text, which is read aloud and is not even available in the Creole language of the audience? See Example (1):

- (1) doop. <door si Jüngers, vid: Joh:4, V. 2.> (ms. 3231: 46)  
baptize <by his pupils. see: John: 4, verse 2.>

The use of the reference seems unnecessary for an audience of recently baptized people, who could hardly read themselves, and did not have the possibility of using other written biblical sources. However, the use of Latin *vid.* (*vide* 'see') can only be seen as referee design. It is used because in this tradition a correct reference is made by using Latin *vid.* I will show other examples in the following section.

The missionary author composed his texts outside of the speech situation. The texts were used in the congregation, and missionaries tried to connect with their audience as much as possible. See for instance the case of Johann Böhner who writes that he started translating large texts after 38–40 years of living in the Creole community (see the following preface of ms. 322), and the choice of Jochum Melchior Magens to translate the New Testament, as he was born on the Danish Antilles, and therefore knew the language well (see the following introduction to Magens 1781). External to this situation we have the traditions of respectively the Moravian Brethren and the Lutheran church.

Which elements of the translations display the influence of these traditions?

1. *Selection of the texts.* Already in the early period of the Moravian mission there was a desire to start writing and translating a hymn book, followed by the larger religious texts. Even the order of translation was regularly mentioned by Johann Böhner. He also stresses the importance of the translation of the *Idea Fidei Fratrum* (ms. 326) for use in the community.
2. *Related metalinguistic comments.* In the introductions to their major translations both Magens and Böhner indicate that certain religious terminology had to be borrowed from, for instance Dutch (as the lexifier of Virgin Islands Dutch Creole). Not only were these words included in the texts to connect to the source texts, and their Christian tradition, they were also used to educate new members of the congregation, and to teach them Christian jargon.
3. *Church/missionary jargon.* In all the texts we can find words and expressions which were commonly used in the Christian community, for instance among missionaries and within biblical texts. However, these items should be accompanied by synonyms or explanations for an audience of listeners and readers who were unfamiliar with Christian tradition or were new to this community.

The *selection of texts*, (1), needs further study of ego-documents. It is however interesting to see that metalinguistic comments point to the strategy of first translating hymns into the Creole language, before turning to translating large texts like the Gospel, the Old Testament and the *Idea Fidei Fratrum*. It is also remarkable that Lieberkühn's 1769 Gospel Harmony was translated into the languages of the first overseas congregations of the Moravian Brethren, instead of the New Testament, because of its usefulness for new members of the communion.

In *metalinguistic comments*, (2), translators and writers show that the language they use in their texts is highly colored by church jargon. Both Magens, in his 1781 New Testament, and Böhner, in the preface of manuscript 322 and in the preface of ms. 326, write that words to express concepts unknown in Creole, should be borrowed from Dutch. Böhner (1780) writes in the preface of his translation of the *Idea Fidei Fratrum* (ms. 326):

I had to use and keep a German or Dutch word here and there, and I consider it not a bad thing if we at the same time improve the Creole language with words which it lacks, which can easily be explained to the Creoles.<sup>13</sup>

---

13. "Ich habe auch hie u. da ein hoch deut\*sh\* u. Holländisch Woort gebracht u. by behaten müssen, u. ich achte [-es] es nicht für <ol.s>chaden wenn wir auch die Creol Sp. mit Wöter die si nicht hat, etwas ver bessern die man doch den Creolen kan Ver [-s\*..\*d]<ul.staend>lich machen." (diplomatic edition, abbreviations and symbols, Van Rossem (2017: ix-x))

Magens (1781: Voorberigt) writes:

Although the author of the Creole Grammar has given a warning following the dialogue between a catechist and a pagan, and has stated there that it is necessary for religious matters to follow the Dutch language, as the true origin of the Creole language, so I am obliged to give a warning too, that I have followed the same principle in this translation of the New Testament. I have followed the Creole way of speaking everywhere but **I did not want to use ordinary words and expressions, because these are not fitting in this clerical matter.** And I myself have heard that the whites, as well as the negroes, were disturbed and offended when they heard the one or the other in some sermon or spiritual discourse, where such ordinary words were used, which they also use in daily intercourse.

In the extensive preface in ms. 322, Böhner writes:

‘And although I didn’t find it to be a very accurate translation into the Creole language as far every word is concerned, it is still not the case that one who is born a Creole, will not be able to learn to understand the words which are not very familiar to him already. But the context of the language, provides also the meaning of some words which I have taken from the Dutch or High German language, for example: *onberispelik*; which is, when people behave so righteously, that others have no reason to reproach them, or to find mistakes in their behavior.’<sup>14</sup> Manuscript 322, *Die Handelingen of Geskiedenis van ons HEER en Heiland JESus Christus ut die Vier Evangelisten na een te\ssammenhang gefoegt* (Gospel Harmony, about 1780)

With regard to a *missionary jargon*, (3), these metalinguistic comments are clear: it was the referee, the ingroup of missionaries or the congregation of Moravian Brethren (and in the case of Magens the Danish Lutheran church), who required the use of a specific jargon to teach the bible to the slaves in a clear fashion.<sup>15</sup> The necessary words and expressions were not available in the Creole language and had therefore to be borrowed from another language. Since Dutch resembles Dutch Creole most, it was preferable to borrow words from this language, although in several cases German was used. With regard to Church Sranan, Voorhoeve (1957, 1971) shows that the missionaries had to introduce for instance the words *tolnaar* ‘publican’, *gnade* ‘mercy’ and *discipel* ‘disciple’ in their missionary texts, since no Creole words for these concepts were available.

14. Remarkable is that the word *onberispelik* was not used in Virgin Islands Dutch Creole Gospel Harmonies. It appeared however in the Old Testament (ms. 325c), so ms. 322 must have been written after the Old Testament translation.

15. See Schilling-Estes (2004: 384–386), who places the role of the referee in the initiative dimension. It seems to me the author not only seeks the most useful variety to connect with his auditors, but also uses elements to enrich the text, even when the audience feels no immediate need for them to understand the message.

It is also mentioned that these new words were introduced to educate new members of the community, and indeed, not all borrowed terms are accompanied by a Creole annotation. The corpus is too large to search for differences between the characteristic elements of all participants in the communication situation. Although searching has become easy since the digital Clarin-NEHOL-Corpus was improved by the addition of glosses by Robbert van Sluijs, it is still very hard to pinpoint specific characteristics of texts where these comments are made by a writer or editor inside the texts without a reason. Neither is it easy to distinguish differences made by translators, writers or editors between variants of the same text. That is why this philological search for authentic eighteenth-century Virgin Islands Dutch Creole, focused on the following emendations in the manuscripts:

- a. an item is added without discarding another element. In this case the writer/editor supposes that the original item must be used, however a clarifying item or sentence must be added, for instance between brackets, or as a gloss, to clarify the original item to the audience or in order to connect with the referee. See the following explanation and examples (see Van Rossem 2017: Chapter 12).
- b. an item is replaced by another one. In this case the writer supposes it is unclear what was meant by the first item and it should therefore be replaced by another in order to connect with the referee (see Van Rossem 2017: Chapter 10).
- c. an alternative version of an item is presented. In this case the writer presents two items which are synonyms or alternatives from different lects, and the choice has to be made by the reader. In our manuscript the writers present these items vertically or as alternatives using the conjunction *of* ‘or’ or even *oder* (German ‘or’) (see Van Rossem 2017: Chapter 8).

In this article I focus on items mentioned under (a). In the four available Gospel Harmonies, ms. 321, 322, 3231 and 3232, we find about 420 of these additions, of which about 77 can be related to questios of referee design. In this text I focus on a qualitative analysis. In Van Rossem (2017: Chapter 12) some quantitative results can be found; however further study is necessary.

A beautiful introductory example of educating the auditors is the following. In the German source text of the Gospel Harmony (Lieberkühn 1769/1820: 14, Section 7) we find the following note: *Weise <Gelehrte Heiden>*<sup>16</sup> ‘Wise men <learned heathen>’. In the oldest Virgin Islands Dutch Creole translation, ms. 321 (before 1780), we see it literally translated: *Wies mann sender <geleerde Heiden.>* ‘wise man 3PL <learned heathen>’. The most recent manuscript, ms. 3232 (about

---

16. See for explanation of diplomatic symbols Van Rossem & Van der Voort (1996: XII–XIII) and Van Rossem (2017: 25–38). For instance < a > means that *a* was added, [-a] means *a* was deleted and [G+]<g> means *G* was overwritten by *g*.

1795), shows a comparable translation. However, ms. 3231 (before 1792) explains it as follows: *wies Mann* <van die Heiden die a wees geleerd Volk> ‘wise man <of the heathen who were learned people>’. The term *wise man* is thus not only translated, but explained for the audience. There is no reason to assume the note was originally made to connect with the Creole speaking auditors, since it was already present in the German source text. Ms. 322 (before 1780) however, contains a much longer addition:

- (2) *Wiese Mann* <sender a wees geleerde Heyden en Ster onder soeker, Wise man 3PL PST be learned heathen and astrologer en a verwacht die Sterr van d\*..\* Balaam a ka see wies, die and PST expect the star of the Balaam PST PRF say wise who sender a ka lees, 4 Mos\*22\* 12.> (322: Section 7: 22)  
3PL PST PRF read 4 Moses\*22\* 12  
‘wise men <who were wise heathens and astrologers, and expected the star which Balaam had foretold, which they had read, Fourth book of Moses (Numbers) \*22\*: 12>

The note is related to the German source text; however, it is much more explanatory and educational in terms of its content. This function can be seen as auditor design. The reason why the reference *4 Mos\*22\* 12*. was added is however referee design: it is not helpful to the audience of the Creole Gospel Harmony, but helpful for the ingroup of translators/missionaries, for example to find related texts with which to educate the audience. These references appear regularly. In the four Gospel Harmonies (321, 322, 3231 and 3231) about 10% of all additions consist of these references, which can be categorized as follows:

1. reference to a scriptural passage: <5. Mos. 18,15>: fifth book of Moses (Deuteronomy), Section 18, verse 15.
2. reference to a liturgical anniversary: <Evang am Neu Jahr> ‘Gospel at New Year’: section to be read at New Year, in German (of the missionary) and not in Virgin Islands Dutch Creole (of the audience)
3. reference to one or more bible verses: <20–39>, <No 27>

For the study of Virgin Islands Dutch Creole, it is more interesting when these additions are related to language. In the following examples we see that the presentation of alternatives is in both ways:

1. missionary jargon is added next to Creole word: *Stoel* <Troon> ‘chair <throne>’, *Tobbo*. <met reten.> ‘bucket <metrete>’, *Feest* <Pingster> ‘celebration <Pentecost>’
2. addition according to source text: *tien Stadt sender*, <Decapolis>

3. explanation of action, Creole into jargon: *skriev sender op* <vor lat *taxeer sender.*> ‘write them down <to let them be counted>’, *hoppo* <*staan op*> ‘(get) up <stand up>’<sup>17</sup>

With regard to 2, the addition was already placed in the source text. Since the word *Decapolis* is only used once, the only need to add it must have been the teaching of a place name. With regard to 1 and 3, the additions are placed after the Creole words which were already present in the manuscript. The translator must have felt the need to teach the audience new words with meanings slightly different from the first mentioned ones, but common in missionary jargon.

The following examples may look like cases in which the reverse happens: Creole explanation is added next to missionary jargon. However, these additions were all already present in the German source text, and are therefore not exclusively made for a Creole speaking audience.

1. synonym: *Messias* <*die gesalvde.*>, addition in source text: (*welches ist verdolmetschet, der Gesalbte*) ‘which is translated, ‘anointed’; *Godt* <*die HEer*>, in source text: *Gott der Herr* ‘the Lord God’
2. explanation: *Donders-Kinders.* <*Machtige Getugen, glik die Stem van Donder.*>, *Werld* <*die [G+]<g>eheelee Roman-Rik*>, *seste Uur.* <*Meddag, omtrent twalf Uur*>

See (3) and (4):

(3) (Gospel Harmony, Section 25)

Lieberkühn 1769/1820):	<i>Donnerskinder</i> < <i>donnernde gewaltige Zeugen</i> > (German)
Lieberkühn (1771):	<i>The sons of thunder</i> < <i>powerful witnesses</i> > (English)
321	<i>Donners Kinder</i> < <i>Donnerde Machtige getuigen.</i> >
322	<i>Donners-Kinder:</i> < <i>Machtige Getiegens met donner Woorden</i> >
3231	<i>Donders-Kinders.</i> < <i>Machtige Getugen, glik die Stem van Donder.</i> >
3232	<i>Kinders van Donder;</i> < <i>Krachtvolle machtige Getugen;</i> >
Saramaccan:	<i>dondrominini</i> ‘thunder children’, no addition.

---

17. Bob Marley?

## (4) (Gospel Harmony, Section 6)

Lieberkühn 1769/1820):	<i>Welt</i> <Das ganze römische Reich> (German)
Lieberkühn (1771):	<i>World</i> <the whole Roman empire> (English)
321	<i>Werld</i> <die geheele romanisch Riek.>
322	<i>Werlt</i> <die geheele Romisch Rik.>
3231	<i>Werld</i> <die [G+]<g>eheelee Roman-Rik>
3232	<i>Werld</i> <die geheele Roman Rik.>
Saramaccan:	<i>Tulu kondre va hem</i> ‘his whole country’, no addition.

In these two Examples (3) and (4) the explanation seems to have originally been inserted to connect with the Creole audience, whereas in fact only Examples (3), texts 322/3231, provide an explanation for an audience of laymen, while all others resemble the original addition in the source text. In Wietz (1792), a comparable Gospel Harmony in Saramaccan/Sranan, the addition is absent.

The Example (5) shows the choice of the translator to change the original text.

## (5) (Gospel Harmony, Section 19)

Lieberkühn (1769/1820):	<i>die sechste Stunde</i> <Mittags um zwölf Uhr> (German)
Lieberkühn (1771):	<i>the sixth hour</i> <At noon, about twelve o’clock.> (English)
321	<i>tegen twaelf uhr.</i> <meddag> ‘afternoon’
322	<i>die Twaelfde uur tegen</i> (middag)
3231	<i>die seste Uur.</i> <Meddag, omtrent twalf Uur>
3232	<i>die seste Uur</i> <Middag na twalf Uur.>
Saramaccan	<i>teh son kissi dina kaba</i> ‘until sun had got lunch’, no addition.

Although the note is available in both the German source text and the English translation which might have been used for translations 3231 and 3232,<sup>18</sup> both of the oldest texts ignore the original biblical expression. I think this can be seen as a sign of connecting with the audience and not with the referee. It seems as if the later texts consider referee design more important than auditor design.

Another case, in which the replacement of an item plays a role, we see the same tendency, as a Creole word is replaced by missionary jargon. See for instance the Examples (6), where the use of Creole *tobo* ‘bucket’ and German/Dutch *meet/maat* ‘bucket to measure dry goods, bushel’ changes over time, and (7) in which both Creole *tobo* and Dutch *schepelmaat* ‘bushel’ are used.

18. See for instance the resemblance in Example (4) between the use of *Roman* in the English translation and 3231 and 3232, while 321 and 322 use German forms ending on *-isch*.



## (6) (Gospel Harmony, Section 15)

Lieberkühn (1769/1820):	<i>zwei bis drey Maass</i> (German)
Lieberkühn (1771):	<i>Two or three firkins apiece</i> (English)
321	<i>twee tee drie Tobo</i> : only the Creole word is used, without remark
322	<i>twee of drie [-Tobo] &lt;meet&gt;</i> : the Creole word is discarded and replaced by missionary jargon
3231	<i>twee of drie Tobbo. &lt;met reten.&gt;</i> : the Creole word is 'clarified' by missionary jargon
3232	<i>twee of drie Maat</i> : only missionary jargon is used
Saramaccan:	<i>tu effi dri kanetjes</i> 'two or three jugs' (>Dutch <i>kannetje</i> 'jug?'): no biblical jargon

In Section 25 of the Gospel Harmony *tobbo* and *meet* are again presented as alternatives, although here they are related to other words in the source texts.

## (7) (Gospel Harmony, Section 25)

Lieberkühn (1769/1820):	<i>Scheffel</i> (German)
Lieberkühn (1771):	<i>bushel</i> (English)
321	<i>(Tobo) Skoepel-Meet</i> the Creole word <i>Tobo</i> was written first, immediately followed by missionary jargon <i>Skoepel-Meet</i> , however the Creole word was not erased, but by adding brackets changed into synonym between brackets.
322	<i>&lt;een&gt; Tobo (of Skeepel-Meed)</i> Creole word was written first, missionary jargon was added as a synonym, but Creole word remained the preferred word, since the missionary jargon word was placed between brackets.
3231	<i>na onder een Tobbo</i> Creole word was the only one, no missionary alternative was added.
3232	<i>Skepel-Maat</i> No Creole word, missionary jargon is only one used.
Saramaccan	<i>tobbo</i> : no notes, nor missionary jargon.

In several cases we see that the translator did not relate to the language of the auditors, the group of Creole speaking people who were only recently introduced to Christian tradition and jargon. Typical biblical terms and expressions were used, and sometimes explained.

A typical example of referee design is the above-mentioned use of *Jesum* (Van Rossem & Van der Voort 1996: 31, c). It seemed as if no attempt was made to change an example of written language in one which must have been common in the vernacular, for instance *Jesus*. Lieberkühn (1769/1820) uses *Jesum* in several cases. Since this work cannot be searched digitally yet, I have no clear figures. The English translation of Lieberkühn's Gospel Harmony (1771) only uses *Jesus*. The first 35 sections of the Virgin Islands Dutch Creole versions of the Gospel Harmony can easily be compared:

321 1–35:<sup>19</sup> *Jesus* 95 – *Jesum* 6 – *Jesu* 22  
 322 1–35: *Jesus* 135 – *Jesum* 0 – *Jesu* 4  
 3231 1–35: *Jesus* 151 – *Jesum* 0 – *Jesu* 0  
 3232 1–35: *Jesus* 151 – *Jesum* 0 – *Jesu* 0<sup>20</sup>

Only in the oldest translation do we see a relatively large number of forms declined in the accusative case (correct for the object in Latin), which is a clear example of referee design. From 1790 on, only *Jesus* is used, which may be due to a better connection with the auditors, or to the use of the English translation of Lieberkühn (1769/1820). Magens (1781) only uses *Jesus*.

This study also examined what happened when the gospel text uses concepts which are not available on the Danish Antilles. The words *wolf* 'wolf' and *kameel* 'camel' appear a few times in the New Testament and are both adopted in all Virgin Islands Dutch Creole versions. Wietz (1792) also uses *Kameel* and not a Sranan/Saramaccan alternative. The word for *locust* might well have been in use in Virgin Islands Dutch Creole. Lieberkühn (1769/1820, Section 10) uses *Heuschrecke*, however, in all the Creole texts *locust* (<E. *locust*) (321, 322, 3231) or *spring-haan* (<D. *sprinkhaan*) was used. These words do not appear in Virgin Islands Dutch Creole wordlists.

The word *Wüste* 'desert' (Lieberkühn 1769/1820: Section 5) was in chronological order translated as: *Weldnes* (<D. *wildernis*) 'wilderness' (321, 322), *die onbewoonde Plek* 'the uninhabited place' (3231), *Woestyne* (<D. *woestijn*) 'desert' (3232, 3110). The oldest texts show an alternative which can be placed on the Danish Antilles, and 3231 presents a description which connects with the auditors. However, the youngest texts and Magens (1781) use the European/biblical term, which is an example of referee design. At another place Lieberkühn (1769/1820: Section 30) used

19. The four Gospel Harmony manuscript variants only have the first 35 sections in common. These texts are available in Clarin-NEHOL as separate texts to make variance analysis possible.

20. In Wietz (1792): *Jesus* – 127, *Jesum* – 0, *Jesu* – 13,

*wüste Stätte* ('desolate land, wasteland'), While the later texts have *woestyne* again, and the oldest text had *een Plek waar die no hab Volk woon* 'a place where they don't have people living' (321) or *een onbewoonde plek* 'an uninhabited place' (322, 3231). Wietz (1792) uses the Saramaccan term *matu kondre* 'jungle country' and connects with his audience of Creole speaking maroons.

The most beautiful example of using a biblical, European concept, without connecting with the auditor, only with the referee, is the following. Lieberkühn (1769/1820: Section 57) has the following comparison: *weiss, wie der Schnee* (English translation had *White as snow*). In ms. 321 Böhner did his best to connect with his audience: *witt, as gebleekte cottoen* 'white as bleached cotton'. In later translations, however, we find *witt, glik die Snee* (>G. *Schnee*/Da. *sne*) (322). Magens (1781) uses *wit, als Sneew* (>D. *sneeuw*) and the printed version of the Gospel Harmony (3110) also uses the Dutch-related term: *wett, as die Sneuw*. Although this concept must have been explained to the audience, the word for 'snow' never entered one of the word lists. Wietz (1792) has *weti leki Maulu* 'white like cotton', which parallels ms. 321.<sup>21</sup> Snow does not fall in Suriname or on the Danish Antilles. The referee demands 'snow', the auditor needs 'cotton' to understand the concept of intense whiteness.

## Final remarks

Eighteenth-century Virgin Islands Dutch Creole manuscripts do look artificial. The orthography is that of the languages of the translators and is etymological rather than based on pronunciation. The texts were written by L2-speakers, translated from formal religious texts, and reflect a Christian world which is Eurocentric, and strange to those who were not familiar with Christian traditions. However, when peeling off these blurring layers, at least some aspects of the natural vernacular become visible.

It is therefore important to recognize this aspect of referee design in situations of historical language contact and historical sociolinguistics. First, delete references which are only meant for missionaries to link to related liturgical texts. Ignore the etymological spelling of nouns, and look for representations which were not influenced by the traditional European orthography, but which are based on pronunciation. Next, identify church jargon, and search for the Creole forms which were used in annotations to clarify this jargon. Separate the annotations which were already added in the source texts from those which were needed according

---

21. I suppose Saramaccan *Maulu* is an eighteenth-century orthography of Saramaccan *maáun* 'cotton'.

to the translator or editor. The hardest part is to distinguish the fixed biblical constructions from creative work of the translator. And then, focus on what translators and writers did in order to really connect with their addressees and auditors: the presentation of alternatives and synonyms, the word order changes from Germanic SOV to Creole SVO, and the replacement of superstrate elements by those which are obviously Creole.

It would be interesting to see which elements, introduced in Virgin Islands Dutch Creole by referee design in the eighteenth century, were still in use in spoken Virgin Islands Dutch Creole in the twentieth century. In that case words or constructions introduced by missionaries would have prove to have entered the vernacular. A first glance in the texts of Schuchardt (1914), De Josselin de Jong (1926), Nelson (1936) and Sabino (2012) is disappointing. On the other hand, the term *Groot zondag* ‘great Sunday’, which was introduced by missionaries to present the audience of Virgin Islands Dutch Creole speakers with a clear word for the most important Christian holidays, can be found in all vocabularies as the Creole word for ‘Christmas’. Even the glossary of Virgin Islands English Creole (Valls 1981: 52) contains *grossondag* ‘Christmas’, though marked as obsolete. *Groot Zondag*, without *Sneew* – Christmas in the US Virgin Islands.

## Acknowledgements

I owe thanks to Harriet Adriaans, Frans Hinskens, Gijsbert Rutten, Robbert van Sluijs, the editors and reviewers for their comments which helped me to improve this text.

## Funding

My research project was supported by the Netherlands Organisation for Scientific Research (NWO) under project number: 023.001.023.

## Abbreviations

The codes of the manuscripts are from Stein (1986).

## Sources

<[www.diecreoltaal.com](http://www.diecreoltaal.com)>

References<sup>22</sup>

- Anonymous. 1836. Moravian Brethren's Tract Operations. *Twenty-second Annual Report of the American Tract Society, Boston, presented at Boston, May 25, 1836, Showing the Facilities Enjoyed for Enlarged Operations in Foreign and Pagan Lands, and in our own Country. Together with Lists of Auxiliaries, Benefactors, Depositories, Publications, &c.*, 34. Boston MA: Perkins & Marvin.
- Bell, A. 1984. Language style as audience design. *Language in Society* 13: 145–204.
- Bell, A. 1992. Hit and miss: Referee design in the dialects of New Zealand television advertisements. *Language & Communication* 12(3–4): 327–340.
- Bell, A. 1997. 19. Language style as audience design. In *Sociolinguistics. A Reader and Coursebook*, N. Coupland & A. Jaworski (eds), 240–250. London: Macmillan.
- Bell, A. 2001. Back in style: Reworking audience design. In *Style and Sociolinguistic Variation*, P. Eckert & J. R. Rickford (eds), 139–169. Cambridge: CUP.
- Böhner, J. 1780. *Korte Begriep van die Christlike Leer, nabin die evangelische Broedergemeenten daer geleeft van August Gottlieb Spangenberg*. Niw Herrnhuth. Manuscript, Unitätsarchiv, Herrnhut; NBVII, R36i1/6i2s.
- Böhner, J. 1781, August 2. *Letter from Neu Herrnhut in St. Thomas to Joseph Spangenberg*. Barby, Germany.
- Chaudenson, R. 1992. *Des îles, des hommes, des langues*. Paris: L'Harmattan.
- De Josselin de Jong, J. P. B. 1926. *Het huidige Negerhollandsch (teksten en woordenlijst)* [Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam, Nieuwe Reeks, Deel 26, no. 1]. Amsterdam: KNAW.
- Hernández-Campoy, J. M. & Conde-Silvestre, J. C. 2012. *The Handbook of Historical Sociolinguistics*. Malden MA: Wiley-Blackwell.
- Hernández-Campoy, J. M. & Schilling, N. 2012. The application of the quantitative paradigm to historical sociolinguistics: Problems with the Generalizability Principle. In *The Handbook of Historical Sociolinguistics*, J. M. Hernández-Campoy & J. C. Conde-Silvestre (eds.), 65–81. Malden MA: Wiley-Blackwell.
- Isles, S. & Weber, G. 1749–1753. *Criol Leedekin Boekje voor gebrieck Van de Neger broer gemeente Na St. Thomas St Crux Overzet üt de Hoog deutsche taal door Broer Samy Isles en George Weber, en een deel mee Assistentie Broer Johannes Van de Jaar 1749 tot Jaar 1753*. Ms, Bethlehem PA, Moravian Archives.
- Labov, W. 1994. *Principles of Linguistic Change, I: Internal Factors*. Oxford: Blackwell.
- Lacoste, V., Leimgruber, J. & Breyer, T. 2014. *Indexing authenticity: Sociolinguistic perspectives*. Berlin: De Gruyter.
- Ladegaard, H. J. 1995. Audience design revisited: Persons, roles and power relations in speech interactions. *Language & Communication* 15(1): 89–101.
- Lieberkühn, S. 1769. *Die Geschichte unsers Herrn und Heilandes Jesu Christi aus den vier Evangelien zusammen gezogen*. Barby: Heynrich Detlef Ebers.
- Lieberkühn, S. 1771. *The harmony of the four gospels, or the history of our lord and saviour Jesus Christ (...)*. London: Lewis.

---

22. Extensive versions of references to texts related to Virgin islands Dutch Creole can be found in the Comprehensive Bibliography of in or about Virgin Islands Dutch Creole, which is regularly updated on <[www.diecreoltaal.com](http://www.diecreoltaal.com)>.

- Lieberkühn, S. 1820. *Die Geschichte unsers Herrn und Heilandes Jesu Christi aus den vier Evangelien zusammen gezogen*. Gnadau: Buchhandlung der Evangelischen Brüder-Unität, Chr. Ernst Genft.
- Magens, J. M. 1770. *Grammatica over der Creolske sprog, som bruges paa de trende Danske Eilande, St. Croix, St. Thomas og St. Jans i Amerika. Sammenskrevet og opsat af en paa St. Thomas indföd Mand*. Copenhagen: Gerhard Giese Salikath.
- Magens, J. M. 1781. *Die Nywe Testament van ons Heer Jesus Christus ka set over in die Creols tael en ka giev na die ligt tot dienst van Die Deen Mission in America [Gedrykt in Copenhagen, 1781. Bie Die Erfgenamen van Godiche]*. Copenhagen: Godiche.
- McWhorter, John. 1997. Review of *The Early Stages of Creolization*, Jacques Arends (ed.). Amsterdam/Philadelphia: John Benjamins, 1995. xv + 297 pp. *New West Indian Guide/Nieuwe West Indische Gids* 71(1–2): 174–178.
- Nelson, F. G. 1936. Words and short texts in Negerhollands, gathered in St. Thomas, June 1936. Ms.
- Oldendorp, C. G. A. 2000. *Historie der caribischen Inseln Sanct Thomas, Sanct Crux und Sanct Jan, insbesondere der dasigen Neger und der Mission der evangelischen Brüder under derselben. Erster Teil. Kommentierte Ausgabe des vollständigen Manuscriptes aus dem Archiv der Evangelischen Brüder-Unität Herrnhut*, H. Beck, G. Meier, S. Palmié, P. Stein & H. Ulbricht (eds). Berlin: VWB, Verlag für Wissenschaft und Bildung.
- Oldendorp, C. G. A. 2002. *Historie der caribischen Inseln Sanct Thomas, Sanct Crux und Sanct Jan, insbesondere der dasigen Neger und der Mission der evangelischen Brüder under derselben. Zweiter Teil: Die Missionsgeschichte, Teilband 1: 1732–1748, Teilband 2: 1749–1753, Teilband 3: 1759–1768, Anhang, Register, Glossar, Wortverzeichnis, Abbildungen*, H. Beck, G. Meier, S. Palmié, Aart H. van Soest, P. Stein & H. Ulbricht (eds). Berlin: VWB, Verlag für Wissenschaft und Bildung.
- Pontoppidan, E. 1881. Einige Notizen über die Kreolensprache der dänisch-westindischen Inseln. *Zeitschrift für Ethnologie* 13: 130–138.
- Pontoppidan, E. 1887. Det dansk-vestindisk Kreolsprog. *Tilskueren* 4: 295–303.
- Sabino, R. 2012. *Language Contact in the Danish West Indies: Giving Jack his Jacket*. Leiden: Brill.
- Schilling, N. 2004. Investigating stylistic variation. In *The Handbook of Language Variation and Change*, J. K. Chambers, P. Trudgill & N. Schilling (eds). Oxford: Blackwell. Blackwell Reference Online, 2008, 21 January.
- Schuchardt, H. 1914. Zum Negerholländischen von St. Thomas. *Tijdschrift voor Nederlandsche Taal- en Letterkunde* 33: 123–135.
- Sewel, W. 1708. *Nederduytsche spraakkonst, waarin de gronden der Hollandsche taale naauwkeuriglyk opgedólven , en zelfs voor geringe verstanden, zo ten aanzien der spellinge als bewoordinge, duydelyk aangewezen zyn*. Amsterdam.
- Stein, P. 1985. Die Anfänge der Verschriftung einer Kreolsprache: das Negerhollands im 18. Jahrhundert. In *Entstehung von Sprachen und Völkern, Glotto- und ethnogenetische Aspekte europäischer Sprachen, Akten des 6. Symposions über Sprachkontakt in Europa, Mannheim 1984*, P. Sture Ureland (ed.), 437–457. Tübingen: Max Niemeyer.
- Stein, P. 1986. The documents concerning the Negro-Dutch language of the Danish Virgin Islands, St. Thomas, St. Croix, and St. John – Negerhollands -, in the Unitäts-Archiv (Archives of the Moravian Brethren) at Herrnhut. A commented bibliography. In *Papers on Negerhollands, the Dutch Creole of the Virgin Islands. Amsterdam Creole Studies* 51 [Publikaties van het Instituut voor Algemene Taalwetenschap IX], 19–31. Hans den Besten (ed.). Amsterdam: University of Amsterdam.

- Valls, L. 1981. *What a Pistarckle! A Glossary of Virgin Islands English Creole*. St. John, U.S.V.I.
- van Rossem, C. 2017. *The Virgin Islands Dutch Creole Textual Heritage: Philological Perspectives on Authenticity and Audience Design* [LOT Dissertations 477]. Utrecht: Netherlands Graduate School of Linguistics.
- van Rossem, C. & van der Voort, H. (eds.). 1996. *Die Creol Taal, 250 years of Negerhollands texts*. Amsterdam: Amsterdam University Press.
- Voorhoeve, J. 1957. Missionary linguistics in Surinam. *The Bible Translator* 8: 179–190.
- Voorhoeve, J. 1971. Varieties of Creole in Suriname, Church Creole and Pagan cult Languages. In *Pidginization and Creolization of Languages*, D. Hymes (ed.) 303–315. London: CUP.
- Wietz, I. L. 1792. *Die Geschichte unseres Herrn und Heilandes Jesu Christi aus den Vier Evangelisten zusammenzogen*. Manuscript. Moravian Archives, Paramaribo; MS H III A 8(2).

# Subject index

## A

- Amazonia 85–106, 179–216
  - (North)western Amazon 185–190
  - Southern Amazon 190–201
- Boivian Andean foothills 201–202
- Andean 202–205
- applicatives (*see* expression of)
- areal diffusion 179–216
- areal linguistics (*see* linguistic areas)
- audience design model (*see* sociolinguistics of Audience Design)

## B

- bilingualism 15–17, 241
- borrowing 22, 224–229, 232–233
- Breyel 324

## C

- Clarín-NEHOL-Corpus 374–384
- coal mining 292–293, 344–345, 350–354
- code-switching/code-mixing 17–18, 237–260
- code-switching (types of) 243
  - alternations 246–249, 254
  - backflagging 252–253, 274
  - congruent lexicalisation 249–252, 254
  - insertions 243–246, 254
  - non-constituent switching 239–240
  - ragged switching (*see* non-constituent switching)
  - variability in code-switching 240–241

- combiverb (*see* cryptolectal combiverbs)
  - combiverb source (*see* Romani source of combiverbs)
  - computational approaches 137–153
  - constraint-free approach (to code-switching) 238, 254
  - contact-induced change (*see* language contact-induced change) 38
  - contact language 61–83, 87–91
  - contact zones 161–178
  - conventionalisation 255, 274
  - creole exceptionalism 95–101, 108–110
  - creoles, English-lexifier 37–60, 61–83
  - cryptolect (*see* secret languages)
  - creole typology 95–101, 107–158
  - creolization 85–106, 266–270
  - cryptolectal lexica 296–305
  - cryptolectal morphology 305–307
  - cryptolectal combiverbs 308–325
- ## D
- data gaps 142–143, 147
  - dialect 344–348
- ## E
- East Africa 217–233
  - elicitation techniques 39–40
  - English-lexifier creoles (*see* creoles, English-lexifier)
  - enregisterment (*see* sociolinguistic enregisterment)
  - ethnolects (*see* languages of immigrants)
  - Equivalence Constraint 238

- executive functions 241–242
- expression of applicatives 179–216
- expression of motion 37–60
- expression of *until* 217–233

## F

- foreign workers' second language (*see* languages of immigrants)

## G

- grammaticalization 218–224
- Groenstraat 289–296

## H

- heritage languages (*see* languages of immigrants)

## I

- indexicality 342–344, 354–359, 361

## L

- language contact (structural features of) 171–175
- language contact (communicative practices as a factor in) 170–171
- language contact (geography as a factor in) 168–169
- language contact (socio-cultural history as a factor in) 169–170
- language contact-induced change 24, 38–60, 85–95, 161–178
- language contact studies 12–15
- language ideology 342–348
- language policy 61–83
- language switching 241, 256
- languageculture 344–350



- languages of immigrants  
 ethnolects 21  
 foreign workers' second  
 language 20–21  
 heritage languages 21,  
 252–254  
 lexical replacement 296–307  
 linguistic areas 19–20, 161–178,  
 179–260
- M**  
 missionary jargon 367–385  
 missionary linguistics 365–388  
 mixed languages 18–19,  
 261–282  
 split/reverse split mixed  
 languages 272  
 multilingualism 15–17  
 multiple regression analysis  
 116–117
- N**  
 non-relatedness 91–95
- P**  
 parody 349–351  
 parodic language use 352,  
 356–360  
 pattern replication 38  
 phylogenetics 107–108, 111–137,  
 144  
 pidgin & creole studies 8–12  
 place-based identities  
 342–344, 350–353  
 place-making 342, 351–353  
 Pottfers 286–288
- R**  
 ragged switching *see* code-  
 switching (types of)  
 referee design 365–388  
 regional studies 22  
 reliability of data 119–124,  
 142–153, 365–369  
 Romani lexicon 331–334  
 Romani source of combiverbs  
 327–331
- S**  
 secret languages 283–337  
 semantic transparency  
 274–282  
 uniformity 275–277  
 universality 277–278  
 simplicity 278  
 sociolinguistic domains 61–83  
 sociolinguistic enregisterment  
 341–364  
 sociolinguistics (historical)  
 365–369, 384–385  
 sociolinguistics of Audience  
 Design 365–388  
 Sprachbund (*see* linguistic areas)  
 stereotyping 347, 351–353, 359  
 substrate influence 38
- T**  
 Teuten 291, 302, 322–324  
 Tiöten 299–300, 302  
 transparency principle  
 274–275  
 translational motion (*see*  
 expression of motion)
- travellers 284  
 Turkish-German code-switching  
 237–260  
 typology 87–91  
 typology of a complex-motion  
 event 40–43  
 Figure 40  
 Ground 40  
 Motion 40–43  
 Path 40–43
- U**  
*until* (*see* expression of *until*)
- V**  
 valency-changing morphemes  
 184  
 variability in code-switching *see*  
 code-switching (types of)
- W**  
 West Africa 61–83  
 woonwagenbewoners (Dutch/  
 Belgian travelling people)  
 284
- 4-M model 238

# Author index

## A

Aalberse, S. 17, 21  
Aboh, E. O. 6, 45, 56–57, 123, 144  
Adegbija, E. 64–65, 67  
Adejumobi, S. 66  
Adelaar, W. F. H. 203–205  
Agar, M. 342  
Agha, A. 361  
Agheyisi, R. N. 67  
Aikhenvald, A. Y. 133, 170, 175, 180–181  
Al-Jallad, A. 219  
Alnet, A. 233  
Alobwede D'Epie, C. 68  
Alves, P. M. 181, 201  
Ameka, F. 42, 45, 48  
Anchimbe, E. A. 68  
Androutopoulos, J. 342  
Angenot de Lima, G. 181  
Anonymous (van Rossem) 69, 374  
Anonymous (Yakpo) 69, 374  
Ansaldò, U. 109–110  
Appel, R. 9, 15, 17, 20–21  
Araujo, G. A. de 181  
Arends, J. T. G. 9–11, 39, 63, 141, 267, 366  
Arias, L. 169  
Asociación Cristiana de Traducciones Bíblicas 74  
Atkinson, Q. 136–137  
Auer, P. 18–19, 360  
Ayafor, M. 68

## B

Bacelar, L. N. 181, 195–196  
Backus, A. 21, 240, 243, 245–251, 254  
Badiola, L. 241, 244, 256  
Bailey, C.-J. N. 272  
Baker, M. C. 185  
Bakhtin, M. 353

Bakker, P. 6, 18, 25, 63, 85, 88, 90, 96, 98–100, 107–122, 124–125, 127, 140–154, 271, 301  
Bakkes, P. 297, 299  
Baldauf, R. B. 76  
Baptista, M. 116  
Baraza la Kiswahili la Taifa (Kenya) 219  
Barbançon, F. 137  
Barbiers, S. 360  
Barbosa Lemos, A. 99  
Bass, T. 123, 144  
Beal, J. 342  
Bell, A. 32, 365, 369–373, 375  
Bellmann, G. 296  
Bennis, H. J. 16  
Berry, J. 69  
Berwick, R. 123, 144  
Besha, R. M. 233  
Bible Society in Sierra Leone 69  
Bickel, B. 172–175  
Binger, L.-G. 120  
Birchall, J. T. R. 20, 183  
Bird, C. S. 120  
Bischoff, E. 292  
Blokzijl, J. 243  
Blommaert, J. 8  
Boeschoten, H. 240, 243, 253  
Böhner, J. 369–370, 375–377, 384  
Bolekia Boleká, J. 73  
Borges, R. D. 11–12, 15, 38, 52–53  
Bosten, L. 293, 303, 307  
Boswood, J. 201  
Bowern, C. 136–137, 164  
Brandão, A. P. 200  
Breuer, H. 289, 293  
Breyer, T. 366  
Brock-Utne, B. 76  
Brousseau, A.-M. 48

Bruyn, A. 11, 23–24, 32, 37, 40, 52, 55–56, 366  
Bucholtz, M. 344  
Busse, J. 225  
Butler, L. 187  
Bybee, J. L. 185

## C

Cabral, A. S. de A. C. 86–87, 89–92, 94–97, 100  
Cacique Coronado, L. 188  
Cammenga, J. 225  
Campbell, L. 14, 162–164, 166, 168, 173, 175, 180, 183, 208  
Carvajal Carvajal, J. 203  
Caspar, F. 198  
Cerrón-Palomino, R. 203–204  
Çetinoglu, Ö. 253  
Charry, E. 22  
Chatelain, H. 120  
Chaudenson, R. 368  
Chibaka, E. F. 67  
Chien, Y. 262–268, 270, 272, 275, 277–279  
Christaller, J. G. 120, 121  
Christensen, A.-D. 342  
Clahsen, H. 21  
Clark, J. M. 128–129  
Clyne, M. G. 13, 249  
Coehorst, E. 299  
Coler, M. 202–203  
Collard, I. F. 169  
Comrie, B. 180, 184  
Cornips, L. 17, 31, 341–350, 352, 354, 359–360  
Coşeriu, E. 272  
Costa-Faidella, J. 241  
Costa, A. 241  
Creemers, C. 322  
Crevels, E. I. 13, 22, 27, 61, 179–180, 182–183  
Croft, W. 271  
Cruson, C. 20

- D
- Da Cruz, A. 92
- Dahl, Ö. 162–164, 168, 171–172
- Dako, K. 71–72
- Danielsen, S. 20, 26, 138, 154, 188, 199–200
- Daval-Markussen, A. 90, 98, 102, 108, 110–111, 115, 124, 145, 151–152
- de Araujo, G. A. 181
- de Drie, A. 39, 53
- De Josselin de Jong, J. P. B. 367, 385
- de Kleine, C. M. 55
- de Rooij, V. 17, 31, 341–345, 348–350, 352, 354, 359
- de Vries, J. W. 20
- Decker, T. 69
- DeGraff, M. 123, 144
- Delgado, R. 241
- Demirçay, D. 248–249
- den Besten, H. 9–11, 366
- Deuchar, M. 18, 239–241, 243
- Devonish, H. 65
- Devos, M. 233
- Dewulf, M. 284
- Dietrich, W. 87, 93–94
- Diller, A. 169
- Dimmendaal, G. 232
- DiScullo, A.-M. 17, 238
- Dixon, R. M. W. 133, 169, 175, 184, 187
- Doğruöz, A. S. 243, 246
- Donicie, A. 39
- Driessen, G. 344
- Dryer, M. S. 98, 110, 142
- Duden Online 324
- Dunham, M. 218, 233
- Dunn, M. J. 26, 128, 136–138, 154
- E
- Eaton, H. 228–229, 232
- Ebach, M. C. 128, 130
- Eberhard, D. M. 196–197
- Eckert, P. 350
- Edelsten, P. 233
- Edema 225
- Eersel, H. 44
- Ellis, P. D. 117
- Emeneau, M. B. 164
- Endepols, H. 284, 289–292, 298, 300, 302–303, 305, 307, 308, 324
- Enfield, N. J. 164, 175
- Epps, P. 180, 183
- Eriksen, L. 102, 170, 179
- Escure, G. 8
- Evans-Pritchard, E. E. 350
- Extra, G. 16, 347
- F
- Faraclas, N. G. 63, 65–67, 76
- Farquharson, J. T. 96
- Faust, N. W. 101
- Fennig, C. D. 64, 67
- Fernández Garay, A. V. 204
- Finck, F. N. 328–330
- Finney, M. A. 64, 69–70
- Fischer, O. 133–134
- Fleck, D. W. 189
- Flores Farfán, J. A. 13
- Focke, H. C. 39, 50
- Foley, R. A. 169
- Fon Sing, G. 117–118, 123
- Fonka, H. M. 63
- Furman, R. A. 243
- Fyle, C. N. 69
- G
- Gabas Jr., N. 197
- Gaines, R. 42
- Gal, S. 350
- Galboran, L. 232
- Galucio, A. V. 102, 213
- Gani-Ikilama, T. O. 67
- Gibson, H. 224
- Gibson, M. L. 207
- Giesbers, H. 347, 357, 359
- Gift, A. 66
- Gilbert, G. G. 3
- Gilligan, G. 185, 190
- Giribet, G. 126
- Gobardhan-Rambocus, L. 55
- Gomes, D. M. 200
- Gómez-Imbert, E. 180
- Good, J. 109
- Goossens, H. M. W. 290–291, 300
- Goossens, J. 290
- Gorbunova, I. M. 269
- Görlach, M. 63, 69
- Goschler, J. 242
- Green, D. W. 237, 249
- Green, M. J. 68
- Greenhill, S. J. 156, 169, 172
- Grinevald, C. 183
- Gronдона, V. 180
- Grosjean, F. 256
- Guillaume, A. 198–199
- Güldemann, T. 169
- Gullberg, M. 18, 237, 241–242, 256
- Gumperz, J. J. 170
- Günther, L. 285
- Guzzardo Tamargo, R. E. 242
- H
- Haalboom, A. 286
- Hagen, A. 347, 357, 359
- Hall, J. 181
- Hall, K. 344
- Hammarström, H. 20, 169
- Hancock, I. F. 63, 69, 110, 140
- Hanks, W. 342
- Hannß, K. 20
- Harjula, L. 225
- Harries, L. P. 225
- Hart, H. 185–186
- Haspelmath, M. 26, 98, 110, 133, 143, 152, 154, 175
- Haude, K. 22
- Haynie, H. 181–182
- Heine, B. 62, 175, 218–219
- Hekking, E. 22
- Helsinki Swahili Corpus (HSC) 220–222, 224
- Hendle, J. 233
- Hennig, W. 127, 144
- Hermant, P. 284
- Hernández-Campoy, J. M. 365–366
- Hernández, M. 241
- Hinskens, F. IX, 16, 21, 30, 283–284, 290, 293, 299, 304–305, 344, 373, 385
- Hobsbawm, E. 294
- Hofweber, J. 237, 241–242, 244, 253
- Hoiting, N. 42
- Holm, J. A. 99, 110–113, 115–116, 118, 120, 122–124
- Holman, E. W. 166

- Holmes, E. C. 127–128, 131, 348, 350  
Honnens, P. 300  
Huber, M. 63–64, 69, 71  
Huls, E. 248  
Huttar, G. L. 39
- I
- Ichikawa, Y. 264  
Igboanusi, H. 67  
Ihemere, K. U. 64  
Indefrey, P. 18, 237  
Irvine, J. 350  
Isles, S. 368, 370
- J
- Jake, J. 238  
Jansen, B. 20  
Jensen, C. 91, 99  
Ješina, J. 332–333  
Johnson, F. 220  
Johnstone, B. 341, 344, 349, 359, 361–362  
Jones, E. D. 69  
Joseph, B. 165  
Juffermans, K. 64
- K
- Kagaya, R. 225, 233  
Kaiser, S. 264, 275  
Kaji, S. 233  
Kanté, M. 120  
Kapinga, M. C. 222  
Kaplan, R. B. 76  
Karabağ, İ. 239–240, 243, 248  
Kats, J. 352  
Kaufman, T. 86, 91, 166, 170, 270, 345  
Kearney, M. 128–129  
Kelly-Holmes, H. 348, 350  
Khan, A. 18  
Kinda, J. 219  
Kisbey, W. H. 233  
Kluge, F. 285, 290  
Knippenberg, W. H. Th. 291  
Knotter, A. 345–346  
Kobayashi, N. 264  
Koefoed, G. 22  
Kohlberger, M. 190  
Kolipakam, V. 20  
Kook, H. 15, 17
- Koopman, H. 8  
Kootstra, G. J. 15–16  
Kornfilt, J. 239–240, 250–252  
Kortmann, B. 90  
Köster Henke, W. 285  
Kouega, J.-P. 68  
Kouwenberg, S. 10–11, 25–26, 107–110, 123, 144, 149, 151, 154  
Krasnoukhova, O. V. 20, 180, 183  
Krifka, M. 218  
Kuiken, F. 16  
Kusters, W. 110  
Kuteva, T. 219
- L
- Labbé, J. 299  
Labov, W. 366  
Lacoste, V. 366  
Ladegaard, H. J. 372  
Laine, M. 242  
Lalleman, J. 20  
Lambert-Brétière, R. 42  
LangHeinrich, F. 228  
Lathrap, D. W. 189  
Law, P. 8  
Leach, M. B. 225  
Lee, M. K. 92  
Lefebvre, C. 6, 38, 48–49, 56  
Leimgruber, J. 365  
Lemle, M. 93–94  
Leoue, J. 118, 123  
Leppänen, S. 342  
Lewis, M. P. 64, 67  
Leyew, Z. 175  
Lieberman, B. S. 126–128, 131  
Lieberkühn, S. 376, 378, 380–384  
Lijongwa, C. 233  
Lim, L. 110  
Lindstedt, J. 170  
Lipiski, J.M. 73  
Los, B. L. J. 15  
Lumsden, J. S. 38, 56  
Lynn, M. 73
- M
- Mace, R. 169  
MacSwan, J. 238, 254, 256  
Maganga, C. 232  
Magens, J. M. 368, 375–377, 383–384
- Maia, A. da S. 120  
Marinis, T. 237, 241  
Marlo, M. R. 225  
Martins, A. 201  
Masica, C. 165  
Massey, D. 348  
Mathijssen, M. 349  
Matras, Y. 38, 56, 166, 208, 289, 329, 331, 335  
Matthews, S. J. 109–110  
Mayer, M. 39  
Mbangwana, P. 68  
McAlister, K. 256  
McGlynn, C. 64  
McWhorter, J. H. 6, 90, 96–98, 102, 108–110, 366  
Meader, R. E. 181  
Meakins, F. ix, 88, 272–273, 280  
Meijer, G. 8  
Meira, S. 15, 182  
Mello, H. 116  
Messineo, C. 180  
Michael, L. 86, 90, 93–95  
Michaelis, S. M. 26, 110, 143, 152, 154  
Migge, B. 39  
Mihas, E. 187–188, 200  
Milroy, J. 347  
Milroy, L. 17  
Mitchell, S. 154  
Mithun, M. 170  
Moen, D. S. 129  
Mol, F. 227  
Molindo, E. A. 68  
Momoh, A. 66  
Moormann, J. G. M. 284–285, 289, 291, 297–301, 306–308, 310, 315, 322, 324–325, 327, 330–333  
Morrison, M. 233  
Mous, M. 18, 28, 217, 224, 233  
Moyer, M. 18  
Mufwene, S. S. 91, 267  
Mulira, E. M. K. 232  
Müller, A. 93  
Müller, N. J. 20, 93  
Münstermann, H. 347  
Muntendam, A. G. 14, 16, 333  
Murayari, R.M. 86  
Muriungi, P. 225  
Murphy, J. D. 232

- Murphy, K. 117  
 Mutonyi, N. 225  
 Muysken, P. C. 8–22, 26–29, 61–62, 85–86, 98, 107–108, 153–154, 168, 170–171, 173–175, 180, 183, 203, 237–239, 243–248, 253–256, 261–262, 267, 270, 272–274, 279, 307, 366  
 Muzale, H. 225  
 Myers-Scotton, C. 238, 255–256, 271
- N**  
 Nakhleh, L. 136  
 Ndawula, E. G. M. 232  
 Neba, A. N. 67  
 Nelson, F. G. 385  
 Nettle, D. 169  
 Ngefac, A. 67–68  
 Ngonyani, D. 225  
 Ngunga, A. 225  
 Nichols, J. 135, 137–138, 171, 173–175  
 Nicolle, S. 225  
 Nikiema, N. 219  
 Nimuendajú, C. 182  
 Nkengasong, N. 68  
 Nonato, R. 201  
 Norde, M. 133  
 Nordén, A. H. 71  
 Nortier, J. W. R. 16, 239  
 Nurse, D. 177, 225–226, 233
- O**  
 O'Connor, L. M. 261  
 O'Hagan, Z. 87, 90, 94, 99, 101–102  
 Ó hAodha, M. 301  
 Ogden, T. H. 129  
 Ojarikre, A. 67  
 Olbertz, H. 16  
 Oldendorp, Chr. G. A. 182, 370–371  
 Ole-Kotikash, L. 227  
 Olson, H. S. 225  
 Onar Valk, P. 244  
 Orkaydo, O. O. 224–225, 232  
 Osei-Tutu, K. O. A. 71  
 Overall, S. 61, 76, 189–190
- P**  
 Padua, A. J. 181  
 Page, R. D. M. 127–128, 131, 310  
 Pagel, M. 169  
 Pakendorf, B. 170, 172  
 Parafita Couto, M. C. 244  
 Pardeshi, P. 184, 205  
 Parkvall, M. 102, 108–111, 118, 144, 148–149, 152, 155, 166  
 Patrick, P. L. 16, 99, 104, 110–113, 115–116, 118, 120, 122–124, 155  
 Patterson, C. 130  
 Payne, David 183, 185, 187–188, 208  
 Payne, Doris L. 227  
 Payne, T. E. 183, 185, 187–188, 208, 214, 216, 227, 232  
 Peeke, C. 185  
 Penny, D. 131  
 Peperkamp, M. G. J. 187  
 Perridon, H. 133  
 Perrois, L. 73  
 Peterson, D. A. 184, 205  
 Petzell, M. 223, 233  
 Pfaff, C. W. 238, 246  
 Philippson, G. 233  
 Pietikäinen, S. 348, 350  
 Pillinger, S. 232  
 Plag, I. 48, 110  
 Plaza Martínez, P. 203  
 Polinsky, M. 184, 191  
 Pontoppidan, E. 367–368  
 Poplack, S. 238, 245  
 Preston, D. R. 347
- Q**  
 Qiu, P. 264–265  
 Queixalós, F. 190
- R**  
 Rau, D.-H. V. 264, 269  
 Rich, R. G. 181, 185  
 Riester, J. 181  
 Rodrigues, A. D. 86–87, 91–92, 94, 198  
 Rodriguez-Fornells, A. 242  
 Roehl, K. 228  
 Rojas-Berscia, L. M. 261  
 Rose, F. 200
- Rosenberg, M. 129  
 Ross, M. 133  
 Rothman, J. 243  
 Rott, J. A. 21  
 Rowan, O. 181  
 Rugemalira, J. 225
- S**  
 Sabino, R. 366–367, 385  
 Sacleux, C. 220  
 Sakel, J. J. 38, 56, 202  
 Sanada, S. 262–263, 265–268, 270, 277–278  
 Sande, A. 241  
 Sandfeld, K. 164  
 Schadeberg, Th. C. 232  
 Schaefer, R. P. 42  
 Schilling, N. 365, 377  
 Schleicher, C. O. 93  
 Schneider, E. W. 141  
 Schrock, T. 232  
 Schröder, A. 67–68  
 Schroeder, Ch. 242  
 Schuchardt, H. 385  
 Schumann, C. L. 39, 44, 51–53  
 Sebastián-Gallés, N. 241  
 Sebba, M. 44–45, 54  
 Seifart, F. 183, 188  
 Seki, L. 180  
 Seuren, P. A. M. 98, 261–262, 274–275, 277–279  
 Sewel, W. 368  
 Shibatani, M. 184, 205  
 Siewert, K. 286  
 Silva-Corvalán, C. 246  
 Silva, L. de J. 201, 246  
 Silverstein, M. 271, 342, 347  
 Simons, G. F. 64, 67  
 Singerman, A. R. 198  
 Singh, R. 10, 17, 238  
 Singler, J. V. 25–26, 107–108, 123, 154  
 Sippola, E. 335  
 Slobin, D. I. 41–43, 55  
 Smakman, D. 343  
 Smeets, I. 204  
 Smith, N. S. H. 1, 8–12, 30, 33, 38–39, 52, 61–63, 154, 266–267, 274, 280  
 Snow, C. E. 20

- Soares, C. 256  
 Song, J. J. 184  
 Sordam, M. 44  
 Sorensen, A. P. 180  
 Soveri, A. 242  
 Stefanich, S. 241  
 Stegen, O. 228  
 Stein, P. 11, 370, 385  
 Stengs, I. 341, 350, 359  
 Stenzel, K. 180  
 Stirnimann, H. 225  
 Stolz, T. 162, 165–168, 172–173, 175  
 Ström, E.-M. 223–224, 233  
 Stroomer, H. 232  
 Stüve, L. 290, 299, 306  
 Stuyck, K. 351  
 Suzuki, M. 116  
 Swift, K. E. 184  
 Szmrecsanyi, B. 90, 98
- T**  
 TUKI (Taasisi ya Uchunguzi wa Kiswahili) 220  
 Taeldeman, J. 344  
 Talmy, L. 23, 37–43, 47, 49  
 Taylor, C. 353, 361  
 Telles, S. 181, 196  
 The World Bank 76  
 Thissen, L. 343  
 Thomason, S. G. 26, 55, 86, 91, 133, 162–164, 166, 170, 270–271, 345  
 Todd, L. 271  
 Toribio, A. J. 256  
 Treffers-Daller, J. 29, 237–239, 241–243, 245–248, 250–255  
 Tripp, R. 201  
 Tsurumi, E. P. 266  
 Tuggy, D. 184
- U**  
 UNESCO 76  
 United Nations 62
- V**  
 Vallejos, R. 86–87, 90, 94–97, 99–102, 190  
 Valdés Kroff, J. 242, 255  
 Valenzuela, P. 186–187, 189  
 Vallejos, R. 86–87, 90, 94–97, 99–102, 190  
 Valls, L. 385  
 van Baarle, P. 182, 215  
 van Bolhuis, E. G. 285  
 van de Kamp, J. 299  
 van de Kerke, S. VII, 201–205  
 van den Berg, M. C. 8, 11–12, 32, 52–53  
 van der Sijs, N. 308, 311, 335  
 van der Voort, H. 9–11, 27, 32, 179, 365  
 van der Wal, J. 222, 335  
 van der Wouden, T. 19  
 van Dyk, P. 39, 51  
 van Eeden, R. 20  
 van Gijn, R. 20, 22, 26–27, 161  
 van Ginneken, J. 289  
 Van Hauwermeiren, P. 283–286, 288, 306, 335  
 van Hout, R. W. N. M. 17, 21–22  
 van Lieshout, L. 288  
 van Linden, A. 200  
 van Nerum, H. 286, 288  
 van Putten, S. 42  
 van Rossem, C. 11–12, 32, 365–369, 372–376, 378, 383  
 van Winkel, F. 291  
 Vandamme, A.-M. 126–127  
 Vedder, P. 15, 17  
 Veenstra, T. IX, 1, 10–12, 33, 171  
 Vek, J. 295, 303  
 Villerius, S. 15  
 von Sowa, R. 337  
 Voorhoeve, J. 371, 377  
 Vos, P. 288, 335  
 Vydrin, V. 64
- W**  
 Walker, R. S. 93  
 Wang, S.-L. 18, 239–240
- Wang, X. 347  
 Wapishana Language Project 182  
 Warnow, T. 135, 137–138  
 Watson, J. 344  
 Weber, G. 368, 370  
 Wei, L. 237, 249  
 Weijnen, A. 352  
 Weinreich, U. 166  
 Wekker, H. 261–262, 275, 277–279  
 Wheeler, W. C. 126  
 Wichmann, S. 93, 166  
 Wiens, J. J. 129  
 Wietz, I.L. 381, 383–384  
 Wijers, C. 364  
 Wiley, E. O. 126–128, 131  
 Wilkinson, M. 128  
 Willems, J. F. 302, 322  
 Williams, D. 128, 130  
 Wilner, J. 39, 44  
 Wilson, R. 170  
 Winford, D. 114, 133, 208  
 Winkler, J. 289–291  
 Wise, M. R. 183, 185, 187–191, 208  
 Woerfel, T. 242  
 Wolf, S. A. 306  
 Woolard, K. 344  
 Wyse, A. J. G. 70
- Y**  
 Yakpo, K. 14–15, 24, 52–53, 56, 61–65, 73–74, 76, 78  
 Yalçın, K. 239, 243  
 Yamamoto, H. 264  
 Yangklang, P. 42, 50  
 Yankson, S. 72  
 Yoneda, N. 233  
 Yukawa, Y. 233  
 Yuval-Davis, N. 342
- Z**  
 Zarco, M. de 74  
 Zlatev, J. 42, 50  
 Zúñiga, F. 204, 208



# Language index

## A

AAVE 140  
Afrikaans 245  
Aguaruna 189–190, 192, 206, 208  
Aikanã 181, 193, 195, 206–207  
Ainu 119, 120  
Akan 71, 72, 118–121  
Aku 64–65  
Algonquian languages 42, 143  
Aluku 141  
Amele 143  
Andoa 185, 206  
Angolar 112–113, 115, 152–153  
Apiaka 181  
Apinajê 201  
Arabela 181, 185, 187, 206  
Arabic 98, 111–112, 115–116, 143, 219–220, 222, 225, 239  
Arabic languages 98  
Araona 199, 207  
Arawakan languages 138, 188, 200  
Arawakan Campa languages 187  
Arikapú 181  
Ashéninka 187–188, 200, 206  
Ashéninka Alto Perené 187–188, 200  
Ashéninka Ucayali-Pajonal 188, 200  
Assamese 111–112, 115–116  
Atayal 30, 261–274, 276–279  
Athabaskan languages 42  
Austronesian languages 2, 86  
Avatime 42  
Aymara 202–203, 207

## B

Bahamian 140  
Bambara 118–121  
Bantu languages 28, 217–218, 224–226

Bargoens 30–31, 283–294, 297–301, 304, 307–322, 324–333  
Basque 143, 147  
Baure 188, 199–200, 206  
Bena 233  
Berbice Dutch 112–113, 115–116, 123  
Bohemian Romani 332–333  
Borana 225, 232  
Bororo 201, 207  
Brahui 118–120, 123, 142  
Bubi 64, 75  
Bukusu 225  
Burmese 143–146  
Burushaski 143, 147

## C

Cahuapanan 188  
Cameroon(ian) Pidgin 24, 61–65, 67–68, 74–75, 77  
Canela 201, 207  
Cape Verdean 102, 112–113, 115–116  
Caquinte 184  
Cariban 180, 182  
Cavineña 198–199, 207  
Chaga 226, 233  
Chasu 226  
Chinese 42, 262–263, 270, 272  
ChiNgoni 225  
Chinuk Wawa 99, 102  
Chipaya 203, 207  
Church Sranan 371, 377  
Cocamilla 91  
Cocama 85, 91, 94  
Cree 143, 272  
D

## D

Daiso 225  
Digo 225  
Djeomitxi 201  
Dominican 112–115

Dutch 5, 7, 24, 30–32, 37–40, 50, 52–56, 58, 79–80, 82, 112, 115, 117, 238–241, 243–244, 246, 248–249, 253, 255–256, 283–288, 290, 293–299, 301–302, 304–307, 310–313, 316–317, 319–322, 324, 327, 329–333, 341, 345–347, 349, 351–361, 366–369, 371–372, 376–377, 381–382, 384  
Dyirbal 143, 146–147

## E

East African Bantu languages 225  
Eastern Tukanooan languages 180  
Ekoti 226, 232  
Emerillon 200  
English 23–25, 30, 32, 37–40, 43–45, 48–50, 52, 54–56, 61–62, 64–68, 71–72, 76–77, 92, 101, 112, 115, 117–120, 123–125, 140–144, 153, 184, 223, 242–245, 251, 293–294, 296–297, 304–306, 310, 313, 317, 319, 321–322, 332–333, 341–342, 354, 357–358, 360–361, 369, 372, 380–385  
Ewe 37–40, 42–43, 45–50, 53, 55–56, 72  
Ewegbe 123

## F

Fa d'Ambu 152  
Fanakalo 150  
Fijian 143  
Fon 39, 42, 123  
Français Tirailleur 150  
French 30, 38, 64, 67–68, 73, 91, 98, 112, 115, 117, 143, 163, 238, 255–256, 272, 288–289, 293, 301, 304, 307



- G**
- Gbe languages 23–24, 37–39, 46, 50, 52, 55–56
- German 29–30, 143, 237, 239–256, 283, 289, 293, 296, 298–300, 316, 324, 327, 354–355, 360, 371, 376–382
- Ghanaian Pidgin (English) 24, 61–65, 70–72, 75
- Gikūyū 233
- Gogo 225
- Gooniyandi 146–147
- Groenstraat Bargoens 30–31, 283–285, 289–290, 293–295, 298, 309, 311–314, 322, 324–326, 330
- Guahiboan languages 190, 206
- Guajajara 200
- Guaraní 93, 143, 200
- Guarayo/Guarayu 93
- Guaycuruan languages 180
- Guinea Bissau 112–113, 115
- Gullah 140
- Gurindji Kriol 152, 272, 279
- Guyanese 141
- Gweno 226, 233
- H**
- Ha 225
- Haitian 38, 91, 102, 112–115, 118, 123
- Harakmbut 183, 200–201, 206
- Hausa 64–65, 72
- Haya 226, 233
- Hebrew 143, 289, 297–299, 301–302, 327
- Heerlen Dutch 31, 341, 345, 352
- Henese Fleck 283, 290, 299–301, 306, 308, 324, 326, 330
- Hindi 143, 147
- Hiri Motu 99
- Hmong 42, 143, 150–151
- Hunzib 143, 147
- I**
- Igbo 51, 65
- Ik 225, 228, 232
- Ilwana 226, 233
- Indonesian 119–120, 122–123
- Iquito 187
- Irantxe 181
- Iraqw 225, 227, 232
- Irish Gaelic 301
- Italian 32, 341, 354, 356, 360–361
- Itonama 81, 183, 190, 211
- J**
- Jamaican 112–113, 115, 118, 141
- Jaminjungan 42
- Japanese 9–30, 261–279
- Jita 225
- Jivaroan languages 189–190, 192, 206
- Judeo-German 289, 298–299, 327
- K**
- Kagulu 223, 233
- Kamtok 67
- Kannada 133
- Kanoé 181, 195, 206
- Karo 197, 207
- Kayardild 146
- Kempen traders' cant 283, 299–300, 302, 306, 322, 324
- Kiitharaka 225
- Kikongo 23, 39
- Kilivila 143
- Kimbundu 118–121, 142
- Kisii 233
- Kitemoka 181
- Kivunjo Chaga 226, 233
- Klamath-Takelman languages 42
- Kobon 143
- Kolyma Yukaghir 118–120
- Konso 224–225, 232
- Korlai 112, 115–116
- Koyra Chiini 119–122
- Krahô-Canela 201
- Krio 24, 61–65, 69–70, 74–76, 112, 114–115, 141
- Kukama 25, 85–103
- Kokama-Kokamilla 190, 198, 206
- Kupwar Marathi 133
- Kuria 225
- Kwaza 181, 190–191, 198, 206–208
- Kwinti 141
- L**
- Lahu 143–146
- Latin 163, 262, 289, 304, 311, 367, 375, 383
- Latundê 181, 196–197, 206
- Leko 190, 201
- Lekoudesch 289
- Lepper Talp 301
- Light Warlpiri 272, 279
- Limburgish 301, 344
- Lingala 225
- Língua Geral Amazônica (see Nheengatu) 91–92
- Língua Geral Paulista 91–92
- Lokono 182
- Lottegorisch 289
- Luganda 226, 232
- Luluyia 225
- Luvale 143
- M**
- Maasai 225, 227, 232
- Macro-Jê languages 180–181, 201, 207
- Madang languages 143
- Maipuran languages 188
- Makhuwa 222
- Makua 228, 232–233
- Makonde 225–226, 233
- Makuna 181
- Makwe 233
- Mamaindê 196–197, 206
- Mandarin 118–120, 123–124, 262, 265, 270, 272
- Mangarrayi 146–147
- Maninkakan-Bamanankan-Jula (Manding) cluster 64
- Maori 143
- Mapudungun 28, 204–205, 207–208
- Martuthunira 143, 146
- Matacoan languages 180
- Matawai 141
- Matengo 226, 233
- Maung 146–147
- Mbugu 152, 224, 226–227, 229, 233
- Mbugwe 26, 233
- Media Lengua 152, 261, 272–273
- Mina 119–120

- Miraña 183  
 Mon-Khmer languages 42  
 Moré 181, 210  
 Mosetén 190, 201–202, 207  
 Muinane 183  
 Mundurukú 188, 200, 207  
 Munich 206  
 Muylaq' Aymara 202–203  
 Mwera 225
- N**
- Nadëb 181  
 Naduhup languages 180–181  
 Nagamese 102, 112, 115–116  
 Napeka 181  
 Ndamba 226, 233  
 Ndengeleko 226, 233  
 Ndyuka 112, 114–115, 141, 150  
 Negerhollands (*see* Virgin Island Dutch Creole) 112, 114–115, 367  
 Ngiyambaa 143, 146–147  
 Nheengatu 91–93, 99, 182  
 Niger-Congo languages 23, 42, 86, 90, 115, 118–120, 123, 142, 154  
 Nigerian Pidgin English 24, 61–67, 70–71, 74–77  
 Nilamba 233  
 Nomatsiguenga 187  
 Nubi 102, 112, 115–116  
 Nunggbuyu 146–147  
 Nyamwezi 222, 226, 232  
 Nyiha 225
- O**
- Oceanic languages 143  
 Omagua 86–87, 90–95, 99, 101, 103
- P**
- Palenquero 112, 114–116  
 Palikur 182  
 Pama-Nyungan languages 136–137, 143, 146–147  
 Pangwa 225  
 Panoan languages 183, 187–189  
 Papiamentu 109–110, 112, 114–115, 123, 144, 149, 151  
 Paramacca 141  
 Paraguayan Guarani 200
- Pare 226, 233  
 Paresi 181, 200, 206  
 Passamaquoddy-Maliseet 143  
 Peba-Yaguan languages 185, 188, 206  
 Pichi 24, 61–62, 64–65, 70, 73–76  
 Pirahã 119–120, 123  
 Plains Indian Sign Language 99  
 Pogoro 233  
 Portuguese 30, 51, 64, 73, 92, 112, 115, 117, 180  
 Pottefers Cant 30, 283, 285–286, 288, 296–297, 305, 307–318, 320, 322, 325, 335  
 Potteferstaal 283  
 Providence Island Creole 140  
 Puinave 181  
 Puquina 204, 207
- Q**
- Quechua 6, 203, 205, 207, 261–262, 272, 274
- R**
- Rangî 224, 226, 228, 233  
 Rapanui 99, 143  
 Rendille 225, 232  
 Resígaro 188, 206  
 Rikbaktsa 201, 207  
 Rimi 225  
 Roeselare Bargoens 318–320  
 Romani 31, 283, 285, 289, 297, 299, 301, 304, 311, 327–334  
 Rotwelsch 285–286, 291, 299, 306  
 Rutara 225, 230
- S**
- St. Kitts Creole 140  
 Sabanê 181  
 Sandawe 225, 228–229, 232  
 Sango 150  
 Santomense 152  
 Saramaccan 109, 141, 375, 380–384  
 Šatrovački 301  
 Scottish Gaelic 327  
 Sea Islands Creole 140  
 Seediq 266–267, 271–272
- Semitic languages 143  
 Seychellois 112, 115  
 Shambala 226, 228, 233  
 Shawi 28, 185–186, 205–206  
 Shelta 301  
 Shiwiar 190, 192, 206  
 Shiwilu 186–187, 206  
 Sikuanî 190, 206  
 Sino-Tibetan languages 42, 119  
 Sinti-Romani 327–328, 330–332, 334  
 Siriono 93  
 Sorbian 96  
 Southern Min 265, 272  
 Spanish 6, 25, 30, 40, 62, 64–65, 73, 76, 101, 112, 115, 117, 143, 163, 180, 199, 210, 244, 272, 346  
 Sranan 23–24, 32, 37–40, 42–57, 102, 141, 366, 371, 375, 377, 381, 383  
 Sri Lankan Malay 152  
 Surinamese creoles 98  
 Swahili 28, 143, 217–222, 224–229  
 Swedish 143
- T**
- Tai-Kadai languages 42  
 Takanan languages 189, 198–199, 207  
 Tariana 181  
 Tarumã 182  
 Taiwanese Mandarin 265, 272  
 Tayo 153  
 Tiötensprache 299–300, 302, 306  
 Tiriyo 182  
 Tiwi 146–147  
 Tobago Creole 140  
 Tok Pisin 112, 115  
 Trinidad Creole 140  
 Tsimane' 202  
 Tupari 181, 198, 207  
 Tupi languages 86, 91–93, 181, 188, 200  
 Tupi-Guarani languages 25, 85–87, 89, 91–94, 103, 143, 181–182, 188, 190, 200  
 Tupinamba 85–87, 90, 92–96, 99–101, 182  
 Turkana 225, 232  
 Turkish 29, 237, 239–256

## U

Urdu 133  
Urubu-Kapoor 143  
Usan 143  
Uto-Aztecan languages 138

## V

Verlan 301  
Virgin Island Dutch Creole  
32, 365

## W

Wardaman 146-147  
Wapishana 182  
Welsh 243  
Witoto 188, 206

## X

Xavante 181  
Xikrin 201

## Y

Yagua 185, 206  
Yaminahua 187  
Yanesha' 187  
Yao 225  
Yiddish 289, 327  
Yidiny 143, 146-147  
Yilan Japanese 29-30, 261-263,  
270-277, 279  
Yilan Creole 30, 99, 261-269,  
274, 276  
Yurakaré 190, 201-202, 207  
Yoruba 65, 124, 143-146, 153

## Z

Zamboangueno 112, 115  
Zaparoan languages 181, 185,  
187-188, 207  
Zele Bargoens 312, 314-315, 317,  
320, 325  
Zigula 226, 233  
Zulu 124, 143

This book contains twelve articles written in honour of the massive contribution to Contact Linguistics of Pieter C. Muysken in the period 1974-2019. His work on Creole Languages, Mixed Languages, Code-Switching, Bilingualism, Areal Linguistics and Borrowing has been seminal. It has been an inspiration to numerous other scholars working on the various facets of this expanding field. These articles cover a wide range of topics that fall under the scope of what is now termed Contact Linguistics, with four sections entitled: *Creole languages and creole studies*, *Linguistic areas*, *Mixed languages and language mixing*, and *Sociolinguistic aspects of language contact*. The volume also contains a comprehensive bibliography of Pieter Muysken's Language Contact publications.

ISBN 978 90 272 0756 2



9 789027 207562

JOHN BENJAMINS PUBLISHING COMPANY