

The Emergence of morphomic structure in Romance-based Creoles

Olivier Bonami¹ Fabiola Henri² Ana R. Luís³

¹U. Paris-Sorbonne & IUF & LLF
olivier.bonami@paris-sorbonne.fr

²U. Sorbonne Nouvelle & LLF
henrifabiola@gmail.com

³Universidade de Coimbra & CELGA
aluis@fl.uc.pt

Workshop on the Role of Autonomous Morphology in Language Change
ICHL 2011, Osaka, July 2011

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

Introduction

- ▶ Part of a larger study of inflection in Romance-based Creoles.
- ▶ Issues we addressed in previous work:
 - ▶ **Existence** of inflectional morphology in Indo-Portuguese (Luís, 2008) and Mauritian (Becker and Veenstra, 2003; Henri, 2010)
 - ▶ Existence of **morphomic phenomena** (Aronoff, 1994), that is, purely morphological phenomena that are not the direct expression of syntax, semantics, or phonology (Luís, 2008; Henri, 2010)
 - ▶ Quantitative evaluation of the complexity (Bonami et al., 2011)
- ▶ Here we address a slightly different set of issues:
 - ▶ Given that these creoles have inflectional morphology, what is the origin of that morphology? Lexifier, substrate, or innovation?
 - ▶ Can the study of the lexifier shed light on the creole system?
- ▶ Guiding idea: creolization as untutored Second Language Acquisition (e.g. Andersen, 1983; Plag, 2008; Siegel, 2008; Luís, 2008)

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

Small paradigms, inflection classes

- ▶ Verbal paradigms in Indo-Portuguese creoles contain a small number of verb forms
 - ☞ Despite the reduced size of paradigms, verb forms are organised into inflection classes.

	kanta 'sing'	kume 'eat'	subi 'go up'	beblu 'mutter'
BASE	kanta a	kume e	subi i	beblu u
PAST	kant-o	kume-u	subi-u	beblu
PROGRESSIVE	kanta-n	kume-n	subi-n	bebli-n
COMPLETIVE	kanta-d	kumi-d	subi-d	beblu-d

Daman Creole Portuguese

(adapted from (Clements and Koontz-Garboden, 2002))

- ▶ Inflection classes are signalled by theme vowels which constitute genuine morphemes and are therefore visible to inflectional processes

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

The Portuguese conjugation system

- ▶ Portuguese verbal paradigm: 66 cells
- ▶ 3 conjugation classes, each with its own perceptible theme vowel
- ▶ Number of features expressed by inflected verbs: person, number, tense, mood and aspect (and gender, for Participles)

The Portuguese conjugation system

► lavar 'wash' (class1)

TAM	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	lav- o	lav a -s	lav a	lav a -mos	lav a -is	lav a -m
IND.PST.IPFV	lav a -va	lav a -vas	lav a -va	lav á -vamos	lav a -veis	lav a -vam
IND.PST.PFV	lav- ei	lav a -ste	lav ou	lav á -mos	lav a -stes	lav a -ram
IND.PST.PRF	lav á -ra	lav á -ras	lav á -ra	lav á -ramos	lav á -reis	lav á -ram
IND.FUT	lav a -rei	lav a -rás	lav a -rá	lav a -remos	lav a -reis	lav a -rão
SBJV.PRS	lav- e	lav e -s	lav e	lav e -mos	lav e -is	lav e -m
SBJV.PST	lav a -sse	lav a -ssem	lav a -sse	lav á -ssemos	lav a -ssemis	lav a -ssem
SBJV.FUT	lav a -r	lav a -res	lav a -r	lav a -rmos	lav a -rdes	lav a -rem
COND	lav a -ria	lav a -rias	lav a -ria	lav a -ríamos	lav a -ríeis	lav a -riam
IMP	---	lav a	lav e	lav e -mos	lav a -i	lav e -m
INF.PERS	lava- r	lav a -res	lav a -r	lav a -rmos	lav a -rdes	lav a -rem
	INF.IMPERS lav a -r		PTCP lav a -do/a		GER lav a -ndo	

The Portuguese conjugation system

▶ lavar 'wash' (class1)

TAM	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	lav- o	lav a -s	lav a	lav a -mos	lav a -is	lav a -m
IND.FUT	lav a -rá	lav a -rás	lav a -rá	lav a -remos	lav a -reis	lav a -rão

▶ beber 'drink' (class2)

TAM	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	beb- o	bebe e -s	bebe e	bebe e -mos	bebe e -is	bebe e -m
IND.FUT	bebe e -rá	bebe e -rás	bebe e -rá	bebe e -remos	bebe e -reis	bebe e -rão

▶ subir 'go up' (class3)

TAM	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	sub- o	sobe e -s	sobe e	subi i -mos	subi i -s	sobe e -m
IND.FUT	subi i -rá	subi i -rás	subi i -rá	subi i -remos	subi i -reis	subi i -rão

The Portuguese conjugation system

- ▶ Some inflectional processes & morphomic patterns
 - ▶ regular affixation:

lavá- va-mos	‘wash.IND.PST.IPFV.1PL’
lava- r-ía-mos	‘wash.COND.1PL’

 - ▶ stem selection:

fazer	‘do.INF’
faço	‘do.IND.FUT.1SG’
farei	‘do.IND.FUT.1SG’

 - ▶ suffix allomorphy:

lav- ei ,	beb- i ,	dorm- i	IND.PST.PFV.1SG
lav- o ,	bebe- u ,	dormi- u	IND.PST.PFV.3SG

 - ▶ syncretisms:

lavava	IND.PST.IPFV.1 3.SG
--------	---------------------

 - ▶ hetroclisis:

beb i do,	sub i do	IND.PST.PFV.2 3.SG
------------------	-----------------	--------------------

Paradigm downsizing

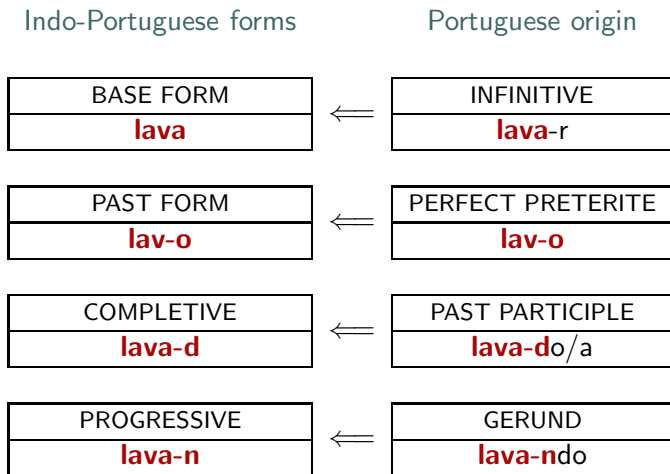
The verbal paradigms of Korlai Portuguese, Daman Portuguese and Diu Portuguese

- ▶ loss of P/N agreement (as is typical of creole languages)
- ▶ **Only** four tense values: unmarked, past, gerund and completive.
- ▶ each lexeme has 4 inflected verb forms

BASE	lav a	'go up'
PAST	lav-o	'went up'
COMPLETIVE	lav a -d	'gone up'
PROGRESSIVE	lav a -n	'going up'

Daman Creole Portuguese

The survival of the inflection class system



Inflectional class extension in Indo-Portuguese

- ▶ In Korlai Portuguese and Daman Portuguese, verbs of substrate origin (Marathi/Gujrati), take an **u** theme vowel

	beblu (DCP) 'mutter'	tepu(KCP) 'heat up'
BASE	bebl u	tepu u
PAST	bebl u	tepu u
PROGRESSIVE	bebl u -n	tepu u -n
COMPLETIVE	bebl u -d	tepu u -d

- ▶ Indo-Portuguese creoles of Korlai and Daman have **adopted** and **extended** the Portuguese conjugation class system

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

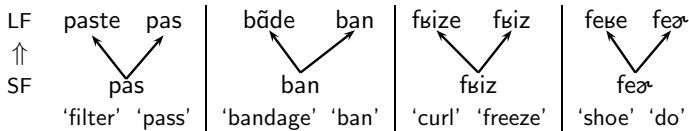
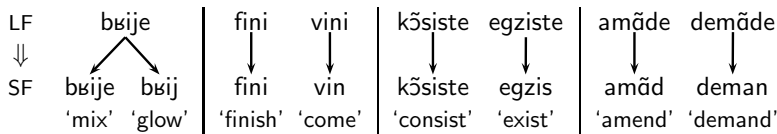
References

Tiny paradigms, but true morphology

- ▶ The Mauritian verbal paradigm has a two-cells paradigm

LF	bɔize	bɔije	vāde	amāde	kɔsiste	egziste	fini	vini
SF	bɔiz	bɔije	van	amād	kɔsiste	egzis	fini	vin
TRANS. 'break' 'mix' 'sell' 'amend' 'consist' 'exist' 'finish' 'come'								

- ▶ Morphological alternation, (contra Corne, 1982): the alternation is not phonologically predictable



The morphosyntactic import of the distinction

- ▶ The alternation encodes two types of features: syntactic and discursive (Henri, 2010)
 - ☞ The division of labor between LF and SF is morphomic (Aronoff, 1994)

		Distribution	SF	LF
Syntax				
No Verum Focus	V with nonclausal complements (NPs,APs,ADVPs,VPs,PPs)		yes	no
	V with no complements		no	yes
	V with clausal complements		no	yes
	V with extracted complements		no	yes
	Verum Focus		no	yes
Morphology				
		reduplicant	yes	no
		base	yes	yes

Constraints on verb form alternation

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

The French conjugation system

► 51 cells

☞ *laver* 'wash':

Finite forms

TAM	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	lav	lav	lav	lav- ɔ̃	lav- e	lav
PST.IND.IPFV	lav- ε	lav- ε	lav- ε	lav-j- ɔ̃	lav-j- e	lav- ε
PST.PFV	lave	lava	lava	lava- m	lava- t	lavε- ɒ
FUT.IND	lavə- ɒ-ε	lavə- ɒ-a	lavə- ɒ-a	lavə- ɒ-ɔ̃	lavə- ɒ-e	lavə- ɒ-ɔ̃
PRS.SBJV	lav	lav	lav	lav-j- ɔ̃	lav-j- e	lav
PST.SBJV	lava- s	lava- s	lava	lava- s-j-ɔ̃	lava- s-j-e	lava- s
COND	lavə- ɒ-ε	lavə- ɒ-ε	lavə- ɒ-ε	lavə- ɒ-j-ɔ̃	lavə- ɒ-j-e	lavə- ɒ-ε
IMP	---	lav	---	lav- ɔ̃	lav- e	---

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
lave	lav- ũ	lave	lave	lave	lave

The French conjugation system

- ▶ Only 1 productive conjugation (Kilani-Schoch and Dressler, 2005)
 - ▶ Dozens of irregular classes of various sizes, with cross-cutting subregularities
- 👉 Bonami and Boyé (2002): full description with 12 principal parts

MSPS	LAVER	FINIR	SORTIR	BOUILLIR	PERDRE	BOIRE	ÊTRE
IND.PRS.1.PL	lav ^õ	finis ^õ	sort ^õ	buj ^õ	pærd ^õ	byv ^õ	sɔm
IND.PRS.3.PL	lav	finis	sort	bu	pærd	bwav	et
IND.PRS.3.SG	lav	fini	sor	bu	pæ	bwa	ɛ
PTCP.PRS	lav ^ã	finis ^ã	sort ^ã	buj ^ã	pærd ^ã	byv ^ã	et ^ã
IMP.2.SG	lav	fini	sor	bu	pæ	bwa	swa
IMP.1.PL	lav ^õ	finis ^õ	sort ^õ	buj ^õ	pærd ^õ	byv ^õ	swaj ^õ
SBJV.PRS.3.SG	lav	finis	sort	bu	pærd	bwav	swa
SBJV.PRS.1.PL	lav ^{jõ}	finis ^{jõ}	sort ^{jõ}	buj ^{jõ}	pærd ^{jõ}	byv ^{jõ}	swaj ^{jõ}
INF	lave	fini	sorti	buji	pærd	bwa	ɛt
IND.FUT.3.SG	lav	fini	sorti	bu	pærd	bwa	sə
IND.PST.PFV.3.SG	lava	fini	sorti	buji	pærdi	by	fy
PTCP.PST	lave	fini	sorti	buji	pærdy	by	ete

Origin of the Mauritian forms

- ▶ All French first conjugation verbs have a striking contrast between a bare stem form and a longer form ending in **-e**
 - ▶ Syncretism between the PRS.SG, the PRS.3PL, the IMP.2SG, the SBJV.SG, and the SBJV.3PL (bare stem)
 - ▶ Syncretism between the infinitive, the past participle, the PRS.2PL and the IMP.2PL
- ▶ (Becker and Veenstra, 2003; Veenstra, 2004) argue that these two syncretic forms are at the origin of the Mauritian LF and SF.
- ☞ Substrate languages may have had an influence on the **function** of the alternation (Wal and Veenstra, 2011), but its **form** definitely stems from statistically prevalent patterns in the input data from French.
- ▶ Supporting evidence:
 - ▶ In 18th century French, infinitive final **-r** was consistently dropped for verbs of all conjugations, except those with a final schwa (Rosset, 1911, Y.-C. Morin, p.c.).
 - ▶ Thus in the French to which future creole speakers were exposed, few verbs marked a difference between INF and PST.PTCP.

Elaborating on Becker & Veenstra's hypothesis

- ▶ For those verbs which did mark a difference between INF and PST.PTCP in 18th century French, the infinitive was overwhelmingly chosen as the long form:

17th-18th c. French	INF=PST.PTCP	INF	PST.PTCP	PRS.SG
# of Mauritian verbs	1769	118	12	11

Origin of the Mauritian long forms; type frequencies compiled from (Carpooran, 2009)

- ▶ Note that 9 of the 12 LF with a French PST.PTCP as origin end in **-r** (e.g. *ouver* 'open'), and could thus have been mistakenly identified as an INF.

Elaborating on Becker & Veenstra's hypothesis

- ▶ Crucial point:
 - ▶ In no instance did a Mauritian verb take up a French INF as its long form and the corresponding irregular PRS.SG as its short form.
 - ▶ Rather, when an irregular PRS.SG was inherited, it has always been inherited as a syncretic LF|SF
- ▶ In addition, short forms are absent from the early Mauritian texts compiled by Baker et al. (2007)



Conclusions:

- ▶ All the evidence points to a single form (almost always the infinitive) as the etymological origin of the whole Mauritian paradigm
- ▶ Thus the alternation between LF and SF is a morphological innovation of Mauritian, not the adaptation of a French alternation.
- ▶ This does not mean that the existence of an $X \sim X_e$ alternation in French played no part in shaping the Mauritian system
 - ▶ The alternation may have been interpreted at first as a rule of sandhi.

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

The structure of the argument

- ▶ Previous work (e.g. Luís 2008) has focused on **why** there is morphology in some creoles, and not in others.
- ▶ Here we address another issue: in Creoles that have inflectional morphology, why do they have **that particular type** of inflectional morphology.
- ▶ The guiding idea: imperfect untutored SLA (see above)
- ☞ Learners should be sensitive to statistically salient features of the word tokens they hear
- ▶ We compare statistical properties of French and European Portuguese inflection to see whether they correlate with the different features of the creoles under investigation.
- ☞ In the absence of large corpora of spoken 18th century Romance languages, we extrapolate on the basis of reliable data from the contemporary languages (both written and spoken)

Data sources

Written EP **CETEMPúblico** (Santos and Rocha, 2001): tagged corpus of Portuguese (180M words), taken from issues of the newspaper *Público* from 1991 to 1998.

Written French 2 years of the newspaper *Le Monde* (2003–2004; 38.5M words), tagged and lemmatized using MElt (Denis and Sagot, 2009)

Spoken EP and French **C-ORAL-ROM** (Cresti et al., 2004), collection of balanced corpora of spoken French, Italian, Spanish and Portuguese (~ 300000 words for each language), transcribed, tagged and lemmatized

Informal French **Lexique 3** (New et al., 2007): database of French inflected words with frequency data compiled from film subtitles

Comparing inflection class size

- ▶ Both languages have a very prominent first conjugation

	Type frequency (written corpus)	Token frequency (written corpus)	Token frequency (spoken corpus)
Portuguese	75.9%	50.0%	30.11%
French	88.7%	45.0%	29.56%

Proportion of first conjugation verbs in both languages
(data from **CETEMPúblico**, **Le Monde** and **C-ORAL-ROM**)

- ▶ Not much can be concluded from this, because class membership may have varied considerably in the last 300 years

Comparing inflection class visibility

- ▶ Inflection class visibility is the extent to which the shape of the forms filling paradigm cells is informative on the inflection class that form belongs to.
- ▶ Theme vowels are clear contributors to inflection class visibility.

	PST.IPFV.3.SG		INF
Portuguese	plissava [plisave]	→	plissar
French	plissait [plisɛ]	→	plisser 'pleat'

Comparing inflection class visibility

- ▶ In EP, almost all cells in the paradigm contain a theme vowel providing diagnostic information on the conjugation class
 - ☞ Ignoring the strange case of the present subjunctive, this is true for all but **1** (PRS.IND.1SG) of the **60** paradigm cells
- ▶ By contrast, in French, only the infinitive, the past participle, the simple past and the (barely used) past subjunctive contain a theme vowel giving unambiguous information on conjugation class.
 - ☞ That is, only **14** out of **51** cells are diagnostic.

	Type frequency	Token frequency (written corpus)	Token frequency (spoken corpus)
Portuguese	98%	99.96%	92.57%
French	27%	33.77%	28.53%

Proportion of paradigm cells with a diagnostic vowel alternation
(data from **CETEMPúblico**, **Le Monde** and **C-ORAL-ROM**)

Discussion

- ▶ Inflection class is much more prominent statistically in Portuguese speech than in French speech
- ▶ In untutored SLA, acquiring inflection classes is easy and useful for Portuguese, hard and inefficient for French
- ▶ If creolization has some relationship to untutored SLA, the inheritance of a conjugation class system in Indo-Portuguese and its disappearance in Mauritian is to be expected.

Recasting Becker & Veenstra's argument

- ▶ Learners of French are strongly attuned to an $X \sim X\mathbf{e}$ alternation. In terms of token frequency:

	among 1st conj. tokens		among all verb tokens	
	C-ORAL-ROM	lexique 3	C-ORAL-ROM	lexique 3
'long form'	49.4%	49.1%	14.6%	19.3%
'short form'	40%	40.1%	11.8%	15.8%
contrasting forms	89.4%	89.2%	26.4%	35.2%

Visibility of the long/short alternation in French

- ▶ There is a strong incentive to interpret this as significant, although because of syncretism, there is little evidence in the input on the morphosyntactic use of the alternation in the lexifier.
- ▶ If creolization has some relationship to untutored SLA, it is not surprising that this alternation was integrated as a morphological innovation in Mauritian.

Outline

Introduction

Inflectional morphology in Indo-Portuguese

What morphology?

Origin in the lexifier

Inflectional morphology in Mauritian

What morphology?

Origin in the lexifier

Towards explaining the differences

Conclusions

References

Conclusions

- ▶ We have shown that:
 - ▶ Indo-Portuguese kept and extended the inflection class system of its lexifier
 - ▶ Mauritian innovated a morphological distinction absent from its lexifier
 - ▶ In both cases, statistical characteristics of the input data from the lexifier helps explain what kind of morphology is found in the creole
- ☞ New type of evidence showing that untutored SLA played an important role in shaping of creole verbal morphology.
- ▶ Limits of this study:
 - ▶ Based on corpora from the contemporary languages
 - ▶ Lack of knowledge of the dialectal characteristics of European immigrants' speech
 - ▶ Lack of knowledge of the extent of substrate influence
- ▶ If anything, this highlights the need for better descriptions of the early stages of creole formation.

- Andersen, R. W. (1983). *Pidginization and creolization as language acquisition*. New-York: Cambridge University Press.
- Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.
- Baker, P., Fon-Sing, G., and Hookoomsing, V. Y. (2007). 'The corpus of Mauritian Creole texts'. In P. Baker and G. Fon-Sing (eds.), *The making of Mauritian Creole*. London: Battlebridge Publications, 1–61.
- Becker, A. and Veenstra, T. (2003). 'The survival of inflectional morphology in French-related Creoles.' *SSLA*, 25:285–306.
- Bonami, O. and Boyé, G. (2002). 'Suppletion and stem dependency in inflectional morphology'. In F. Van Eynde, L. Hellan, and D. Beerman (eds.), *The Proceedings of the HPSG '01 Conference*. Stanford: CSLI Publications.
- Bonami, O., Boyé, G., and Henri, F. (2011). 'Measuring inflectional complexity: French and mauritian'. In *Quantitative Measures in Morphology and Morphological Development*. San Diego: University of California.
- Carpooran, A. (2009). *Diksioner Morisien*. Sainte Croix (Mauritius): Koleksion Text Kreol.
- Clements, J. C. and Koontz-Garboden, A. J. (2002). 'Two indo-portuguese creoles in contrast'. *Journal of Pidgin and Creole Languages*, 2:191–236.
- Corne, C. (1982). 'The predicate in Isle de France Creole.' In P. Baker and C. Corne (eds.), *Isle de France Creole. Affinities and Origins*. Ann Arbor: Karoma, 31–48.
- Cresti, E., do Nascimento, F. B., Sandoval, A. M., Veronis, J., Martin, P., and Choukri, K. (2004). 'The C-ORAL-ROM CORPUS. a multilingual resource of spontaneous speech for Romance languages.' In *Proceedings of LREC 2004*.
- Denis, P. and Sagot, B. (2009). 'Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort'. In *Proceedings of PACLIC 2009*.
- Henri, F. (2010). *A Constraint-Based Approach to verbal constructions in Mauritian*. Ph.D. thesis, University of Mauritius and Université Paris Diderot.
- Kilani-Schoch, M. and Dressler, W. (2005). *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.
- Luís, A. R. (2008). 'Tense marking and inflectional morphology in indo-portuguese'. In S. Michaelis (ed.), *Roots of creole structures: Weighing the contribution of substrates and superstrates*, Creole Language Library. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- New, B., Brysbaert, M., Veronis, J., and Pallier, C. (2007). 'The use of film subtitles to estimate word frequencies'. *Applied Psycholinguistics*, 28:661–677.
- Plag, I. (2008). 'Creoles as interlanguages: inflectional morphology'. *Journal of Pidgin and Creole Languages*, 1:109–130.
- Rosset, T. (1911). *Les origines de la prononciation moderne étudiées au XVII^e siècle*. Paris: Armand Colin.
- Santos, D. and Rocha, P. (2001). 'Evaluating cetempúblico, a free resource for Portuguese'. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. 442–449.
- Siegel, J. (2008). *The emergence of Pidgin and Creole languages*. Oxford University Press.
- Veenstra, T. (2004). 'What verbal morphology can tell us about Creole genesis: the case of French-related Creoles'. In I. Plag (ed.), *Phonology and Morphology of Creole Languages*, no. 478 in *Linguistische Arbeiten*. Max Niemeyer Verlag GmbH.
- Wal, J. v. d. and Veenstra, T. (2011). 'Verb allomorphy in mauritian creole and the cj/dj distinction in bantu'. In *Berlin Bantu Conference (B4ntu)*. Berlin.