Assessing empirically the inflectional complexity of Mauritian Creole

Olivier Bonami[1] Fabiola Henri[2]

[1]U. Paris-Sorbonne &
Institut Universitaire de France
UMR 7023 "Laboratoire de Linguistique Formelle"

[2]UMR 7023 "Laboratoire de Linguistique Formelle"

FACS II
Berlin, November 9, 2010

Outline

Introduction

- Our goal: assess empirically the claim that creole languages have a simpler inflectional system than their lexifier (e.g. Plag, 2006)
- To this end, we compare the complexity of Mauritian Creole with that of French
- We take for granted that Mauritian makes a morphological distinction between long and short verb forms (Veenstra, 2004; Henri, 2010).

LF	brize	brije	vɑ̃de	amɑ̃de	kɔ̃siste	ɛgziste	fini	vini
SF	briz	brije	van	amɑ̃d	kɔ̃siste	ɛgzis	fini	vin
TRANS.	'break'	'glow'	'sell'	'amend'	'consist'	'exist'	'finish'	'come'

- We look at three aspects of complexity:
  - Structure of the paradigm
  - Interface between morphology and syntax/semantics
  - Predictibility relations between cells in the paradigm

Dimension 1: paradigm size

☞ French: 51 cells

Finite forms

TAM	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	lav	lav	lav	lav-ɔ̃	lav-e	lav
PST.IND.IPFV	lav-ɛ	lav-ɛ	lav-ɛ	lav-jɔ̃	lav-je	lav-ɛ
PST.PFV	lavɛ	lava	lava	lava-m	lava-t	lavɛ-r
FUT.IND	lav-ʁɛ	lav-ʁa	lav-ʁa	lav-ʁɔ̃	lav-ʁe	lav-ʁɔ̃
PRS.SBJV	lav	lav	lav	lav-jɔ̃	lav-je	lav
PST.SBJV	lava-s	lava-s	lava	lava-sjɔ̃	lava-sje	lava-s
COND	lav-ʁɛ	lav-ʁɛ	lav-ʁɛ	lav-ʁjɔ̃	lav-ʁje	lav-ʁɛ
IMP	---	lav	---	lav-ɔ̃	lav-e	---

Nonfinite forms

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

☞ Mauritian: 2 cells

LF	SF
lave	lav

Dimension 2: number of processes

- French: allomorphic stem selection + at most 3 suffixes

(1)	a.	*all-ons*	
		go[PRS]-1PL	
	b.	*i-r-i-ons*	
		go-FUT-ANA-1PL	

- Mauritian: allomorphic stem selection, no true affixation

(2)	a.	tõbe	bʁije
		shiver-LF	mix-LF
	b.	tom	bʁije
		shiver-SF	mix-SF

Dimension 3: number of features

- French: disputed. According to Bonami and Boyé (2007), 6 features:
  - Tense
  - Mood
  - Temporal reference type (Verkuyl et al., 2004) and/or aspect
  - Person
  - Number
  - Gender
- Mauritian: undecidable.
  - At least one feature
  - No stable morphosyntactic import

Two further dimensions

- These 3 dimensions are probably what people usually have in mind
- Much recent work in morphology focuses on other aspects of morphological complexity
  - Prevalence of irregularity
  - Number and nature of inflection classes
  - Prevalence of syncretism
  - etc.
- We propose looking at two important dimensions
  - Morphosyntactic transparency: to what extent do the distinctions encoded by the paradigm correspond to 'natural' syntactic and/or semantic classes? (Aronoff, 1994)
  - Interpredictibility: how difficult is it to predict the content of some cell in the paradigm from the content of other cells? (Ackerman et al., 2009)
- Why these dimensions?
  - They definitely matter to speakers
  - Contribution to currently central issues of morphological theory

Outline

The issue

- Starting with Aronoff (1994), growing interest in morphological phenomena that do not correlate with syntactic and/or semantic features in a straightforward way.
- Morphomic pattern: the distribution of some morphological distinction is featurally incoherent
- May concern either affixal exponents or stem allomorphy
- Most celebrated case: distribution of stems in Romance conjugation (Maiden, 1992, 2005; Pirelli and Battista, 2000; Bonami and Boyé, 2002)

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	bwa	bwa	bwa	buv-ɔ̃	buv-e	bwav
PST.IND.IPFV	buv-ɛ	buv-ɛ	buv-ɛ	buv-jɔ̃	buv-je	buv-ɛ

Partial paradigm of *boire* 'drink'

- The presence of morphomic patterns is an element of morphological complexity
- Their prevalence varies widely from language to language

Morphomes in French conjugation

- Cf. (Bonami and Boyé, 2002, 2003, 2007):
  - Affixes have a very simple distribution:
    - no inflection class distinction
    - no morphomic distribution
  - Intricate system of stem allomorphy relying on morphomic patterns

Finite forms

TAM	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	$stem_3$	$stem_3$	$stem_3$	$stem_1$-ɔ̃	$stem_1$-e	$stem_2$
PST.IND.IPFV	$stem_1$-ɛ	$stem_1$-ɛ	$stem_1$-ɛ	$stem_1$-jɔ̃	$stem_1$-je	$stem_1$-ɛ
PST.PFV	$stem_{11}$	$stem_{11}$	$stem_{11}$	$stem_{11}$-m	$stem_{11}$-t	$stem_{11}$-r
FUT.IND	$stem_{10}$-ʁɛ	$stem_{10}$-ʁa	$stem_{10}$-ʁa	$stem_{10}$-ʁɔ̃	$stem_{10}$-ʁe	$stem_{10}$-ʁɔ̃
PRS.SBJV	$stem_7$	$stem_7$	$stem_7$	$stem_8$-jɔ̃	$stem_8$-je	$stem_7$
PST.SBJV	$stem_{11}$-s	$stem_{11}$-s	$stem_{11}$	$stem_{11}$-sjɔ̃	$stem_{11}$-sje	$stem_{11}$-s
COND	$stem_{10}$-ʁɛ	$stem_{10}$-ʁɛ	$stem_{10}$-ʁɛ	$stem_{10}$-ʁjɔ̃	$stem_{10}$-ʁje	$stem_{10}$-ʁɛ
IMP	---	$stem_5$	---	$stem_6$-ɔ̃	$stem_6$-e	---

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
$stem_9$	$stem_4$-ɑ̃	$stem_{12}$	$stem_{12}$	$stem_{12}$	$stem_{12}$

Morphomes in Mauritian conjugation?

- The syntactic contexts in which the two forms appears do not form natural classes (Henri and Abeillé, 2008; Henri, 2010)
- In lexeme formation processes, both forms are used in a way that does not reflect any morphosyntactic property (Henri, 2010)

Syntactic distribution of the SF, 1/2

- The SF is triggered by nonclausal complements

(3) a. *Mo ti manz/*manze kari.*
1SG PST eat.SF/LF curry
'I ate curry.'

b. *Sa stati la dat/*date depi lepok lager.*
DEM statue DEF date.SF/LF from period war
'This statue dates back from the war period.'

- Note that the postverbal argument of unaccusative verbs counts as a complement

(4) a. *Inn ariv/*arive enn aksidan.*
PRF arrive.SF/LF INDF accident
'There has been an accident.'

Syntactic distribution of the SF, 2/2

- The SF also appears with predicative APs and locative goals
- Verbs with a clausal complement take a SF only if another nonclausal complement precedes it

(5) a. *Nou res/*reste malad.*
1PL stay.SF/LF sick
'We are still sick.'

b. *Li pe mars lor disab.*
3SG PROG walk.SF on sand
'She is walking towards the sand.'

c. *Mari inn demann/*demande [ar tou dimounn] [kiler la].*
Mary PERF ask.SF/LF with all people what_time DEF
'Mari asked everyone what time it was.'

Syntactic distribution of the LF, 1/2

- Conversely, the LF appears when the verb has no complement, the complement is extracted , or it is clausal

(6) a. *Mo ti manze/*manz.*
1SG PST eat.LF/SF
'I ate.'

b. *Tibaba ki mo mama ti veye/*vey toule zour.*
little_baby COMP POSS mother PST look_after.LF/SF every day
'It's little babies that my mother looked after every day.'

c. *Mari inn demande/*demann [kiler la] [ar tou dimounn] .*
Mary PERF ask.LF/SF what_time DEF with all people
'Mari asked everyone what time it was.'

Syntactic distribution of the LF, 2/2

- Adjuncts also trigger the LF.

(7) *Li pe marse lor disab.*
3SG PROG walk.LF on sand
'She is walking on the sand.'

- The alternation is **not** phonologically conditioned: a complement that is not adjacent to the verb still triggers the SF.

(8) a. *Nou res/*reste toultan malad.*
1PL stay.SF/LF always sick
'Lit. We remain always sick.'

b. *Nou manze/*manz toultan.*
1PL eat.SF/LF always
'We keep eating.'

Discursive import of the LF

- Interestingly, the LF may appear with a nonclausal complement under certain discursive conditions, precisely in counter-oriented moves (deferments, counter-implicative and counter-propositional moves).
  - ☞ In such contexts, the LF is analyzed as an exponent of Verum Focus (Henri et al., 2008; Henri, 2010).

(9) *Mo ti krwar Mari pa* MANZE/*MANZ *kari poul!*
1SG PST think Mary NEG eat.LF/SF curry chicken
'I thought Mary DIDN'T eat chicken curry!'

SF and LF in reduplication, 1/2

- The two forms are used in "attenuative" reduplication which is a derivational process creating new verbal lexemes (Henri, 2010).
  - ☞ The short reduplicated form is the concatenation of two copies of the base's SF
  - ☞ The long reduplicated form is the concatenation of the base's SF with the base's LF

LF	SF	gloss	red. LF	red. SF	trans.
sãte	sãt	'sing'	sãtsãte	sãtsãt	'hum'
reste	res	'stay'	ʁesʁeste	ʁesʁes	'stay occasionally'
soɚti	soɚt	'get out'	soɚtsoɚti	soɚtsoɚt	'get out occasionally'
balje	balje	'sweep'	baljebalje	baljebalje	'sweep carelessly'

Examples of attenuative reduplication

SF and LF in reduplication, 2/2

- Attenuative reduplication contrasts with intensive reduplication
  - It is a syntactic rather than a lexical process
  - Both the base and the reduplicant are always exact copies

(10) a. *Mo ti manze, manze, manze.*
1SG PST eat.LF eat.LF eat.LF
'I ate, ate, ate.'

b. *Zan nek sant sega, sant sega mem enn lazourne.*
John only sing.SF sega sing.SF sega still day
'John keeps singing the sega, singing the sega all day long.'

Summary

		Distribution	SF	LF
		Syntax		
No	Verum Focus	V with canonical phrasal complements (NPs,APs,ADVPs,VPs,PPs)	yes	no
		V with no complements	no	yes
		V with adjuncts	no	yes
		V with clausal complements	no	yes
		Extracted complements	no	yes
Verum	Focus	In Counter-Oriented moves: deferment, denials (counter-propositional and counter implicative)	dispreferred	yes
		Morphology		
		reduplicant	yes	no
		base	yes	yes

Table: Constraints on verb form alternation

Outline

The issue

- Interpredictibility: how difficult is it to predict the content of some cell in the paradigm from the content of other cells?
- We want a way to assess this in a way that makes sense when comparing languages with different paradigm size and different numbers of inflectional processes.
- Proposed strategy:
  - For each pair of cells $\langle \sigma, \tau \rangle$, measure the amount of information that knowledge of the content of σ gives you on the content of τ
  - Average over all pairs of cells
- One way of doing this: use standard techniques from information theory (as suggested by Ackerman *et al.* 2009)

Entropy

- The entropy of a random variable measures the uncertainty as to what the value of that variable is.

☞ If X is a random variable and p gives the probability of each event x in X,

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

- Intuitively:
  - 0 corresponds to a situation where there is no uncertainty
  - 1 corresponds to a situation where there are two equiprobable possibilities
  - Entropy grows when there are more possible outcomes
  - Entropy decreases when there is a larger difference between the likelihood of the outcomes

Conditional entropy

- The conditional entropy of Y knowing X is the amount of uncertainty there is as to the value of Y once you know X.
- Suggested measure of morphological complexity:
  - Suppose that we want to predict the content of cell τ in the paradigm from the content of cell σ.
  - The complexity of that task is measured by the conditional entropy of the patterns of relatedness between τ and σ knowing what pattern *could* be applicable to σ.
  - In other words, we evaluate how much knowledge of the overall morphological system helps in predicting τ from σ.

NB: This strategy is derived from Ackerman et al. (2009), but slightly different.

  - Both approaches rely on conditional entropy, but the random variables are different
  - The current approach has the advantage of not relying on a previously established classification of lexemes in inflection classes

A concrete example

- We exemplify the approach with a toy lexicon of 4 Mauritian verbs; we try to predict the short form from the long form.
- First step: identify the alternation patterns present in the language fragment

☞ In the present context, simple pattern matching: maximize the common stem

- Second step: identify which patterns each LF *could* exemplify, given its ending

LF	SF	stem	pattern	set of possible patterns
lave	lav	lav	$X\text{e} \to X$	$\{X\text{e} \to X, X \to X\}$
bʁije	bʁije	bʁije	$X \to X$	$\{X\text{e} \to X, X \to X\}$
fini	fini	fini	$X \to X$	$\{X\text{i} \to X, X \to X\}$
vini	vin	vin	$X\text{i} \to X$	$\{X\text{i} \to X, X \to X\}$

A concrete example

- Third step: take each set of possible patterns to characterize morphologically a class of input forms.

LF	SF	stem	pattern	set of possible patterns	class of LF
lave	lav	lav	$X\text{e} \to X$	$\{X\text{e} \to X, X \to X\}$	class A
bʁije	bʁije	bʁije	$X \to X$	$\{X\text{e} \to X, X \to X\}$	class A
fini	fini	fini	$X \to X$	$\{X\text{i} \to X, X \to X\}$	class B
vini	vin	vin	$X\text{i} \to X$	$\{X\text{i} \to X, X \to X\}$	class B

- Fourth step: compute the entropy on this basis

$$H(\text{LF} \sim \text{SF} | \text{LF}) = 1$$

More precisely

- If an item has a class A LF, it is equally likely that it exhibits the $X\text{e} \rightarrow X$ or the $X \rightarrow X$ pattern (1 verb out of 2 in each subclass).

$$H(\text{LF} \sim \text{SF}|\text{LF}: A) = 1$$

- If an item has a class B LF, it is equally likely that it exhibits the $X\text{i} \rightarrow X$ or the $X \rightarrow X$ pattern (1 verb out of 2 in each subclass).

$$H(\text{LF} \sim \text{SF}|\text{LF}: B) = 1$$

- The global conditional entropy is the weighted mean of the local conditional entropies:

$$H(\text{LF} \sim \text{SF}|\text{LF}) = \frac{2}{4}H(\text{LF} \sim \text{SF}|\text{LF}: A) + \frac{2}{4}H(\text{LF} \sim \text{SF}|\text{LF}: B) = \frac{1}{2} + \frac{1}{2} = 1$$

Variations

- If there are more patterns, the entropy may grow

LF	SF	stem	pattern	set of possible patterns	class of LF
lave	lav	lav	$X\text{e} \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
bʁije	bʁije	bʁije	$X \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
fini	fini	fini	$X \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B
vini	vin	vin	$X\text{i} \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B
egziste	egzis	egzis	$X\text{te} \rightarrow X$	$\{X\text{te} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	class C
aʁete	aʁet	aʁet	$X\text{e} \rightarrow X$	$\{X\text{te} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	class C
kõsiste	kõsiste	kõsiste	$X\text{te} \rightarrow X$	$\{X\text{te} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	class C

$$H(\text{LF} \sim \text{SF}|\text{LF}) = 1.250$$

More precisely

- As before, $H(\text{LF} \sim \text{SF}|\text{LF} : A) = 1$
- As before, $H(\text{LF} \sim \text{SF}|\text{LF} : B) = 1$
- $H(\text{LF} \sim \text{SF}|\text{LF} : C) = 1.585$
- $H(\text{LF} \sim \text{SF}|\text{LF}) = \frac{2}{7} \times 1 + \frac{2}{7} \times 1 + \frac{3}{7} \times 1.585 = 1.250$

Variations

- If we extend the lexicon, type frequency kicks in: not all patterns are equally likely, so entropy drops.

LF	SF	stem	pattern	set of possible patterns	class of LF
lave	lav	lav	$X\text{e} \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
aʁete	aʁet	aʁet	$X\text{e} \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
bʁije	bʁij	bʁij	$X\text{e} \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
bʁije	bʁije	bʁije	$X \rightarrow X$	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	class A
fini	fini	fini	$X \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B
sãti	sãti	sãti	$X \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B
paɚti	paɚti	paɚti	$X \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B
vini	vin	vin	$X\text{i} \rightarrow X$	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	class B

$$H(\text{LF} \sim \text{SF}|\text{LF}) = 0.811$$

More precisely

- If an item has a class A LF, it is thrice more likely likely that it exhibit the $X\text{e} \rightarrow X$ than the $X \rightarrow X$ pattern (3 out of 4 verbs).

$$H(\text{LF} \sim \text{SF}|\text{LF} : A) = 0.811$$

- If an item has a class B LF, it is thrice more likely likely that it exhibit the $X \rightarrow X$ than the $X\text{i} \rightarrow X$ pattern (3 out of 4 verbs).

$$H(\text{LF} \sim \text{SF}|\text{LF} : B) = 0.811$$

- Conclusion:

$$H(\text{LF} \sim \text{SF}|\text{LF}) = \frac{4}{8} H(\text{LF} \sim \text{SF}|\text{LF} : A) + \frac{4}{8} H(\text{LF} \sim \text{SF}|\text{LF} : B) = 0.811$$

Application to Mauritian

- We collected the 2079 distinct Mauritian verbs listed in Carpooran (2009), and coded their LF and SF in phonemic transcription.
- In parallel, we extracted from the lexique database (New et al., 2001) the phonemic transcription of all 51 forms from the 2079 most frequent nondefective verbs of French.
- We implemented a python script systematizing exactly the algorithm presented above.
- Overall results:

Mauritian	0.744
French	0.416

- Conclusion: predicting one cell of the paradigm from another on the basis of morphological information is noticeably more complex in Mauritian than in French.

Confirmation

- This result seems quite robust:
  - If we now just compare the LF ~ SF relation just to the INF ~ PRS.3SG relation (to compare what is most directly comparable):

(Mauritian) LF ↦ SF	(French) INF ↦ PRS	(Mauritian) SF ↦ LF	(French) PRS ↦ INF
0.563	0.338	0.925	0.355

  - One might argue that type frequency information is information about the structure of the lexicon, not morphology. If we leave out this information (take all classes to be equiprobable):

Mauritian	1.316
French	0.681

  - Combining the two restrictions:

(Mauritian) LF ↦ SF	(French) INF ↦ PRS	(Mauritian) SF ↦ LF	(French) PRS ↦ INF
1.076	0.537	1.557	1.303

Limitations

- It would be interesting to see whether the same difference between French and Mauritian also holds when token frequency is taken into account.
    - This is currently not feasible: no accessible data for Mauritian
- The pattern matching method that is used to identify classes is extremely crude
    - Devising more subtle methods would definitely be worthwhile
    - However there is no reason to believe that the crudeness of the method has the effect of missing more generalizations in one language than in the other.

Why this result?

- Let us examine the classes the algorithm arrives at, comparing the LF ↦ SF relation with the INF ↦ PRS.3SG relation.
- In Mauritian, we find 11 patterns giving rise to 10 classes.

class	patterns	example		# of lex.	entropy
1	$\{X\text{e} \rightarrow X, X \rightarrow X\}$	kwafe	kwaf	1138	0.565
2	$\{X\text{te} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	gʁijote	gʁijot	268	0.845
3	$\{X \rightarrow X\}$	sufeɚ	sufeɚ	225	0.0
4	$\{X\text{ʁe} \rightarrow X\text{ɚ}, X\text{ʁe} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	kofʁe	kofʁe	159	0.835
5	$\{X\text{le} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	dekole	dekol	138	0.927
6	$\{X\text{i} \rightarrow X, X \rightarrow X\}$	fini	fini	116	0.173
7	$\{X\text{ãde} \rightarrow X\text{an}, X\text{e} \rightarrow X, X \rightarrow X\}$	ʁãde	ʁan	15	0.567
8	$\{X\text{ble} \rightarrow X\text{m}, X\text{le} \rightarrow X, X\text{e} \rightarrow X, X \rightarrow X\}$	ʁeduble	ʁeduble	13	0.391
9	$\{X\text{ɔ̃be} \rightarrow X\text{ɔm}, X\text{e} \rightarrow X, X \rightarrow X\}$	plɔ̃be	plɔ̃b	3	0.918
10	$\{X\text{õde} \rightarrow X\text{on}, X\text{e} \rightarrow X, X \rightarrow X\}$	fekõde	fekõd	4	0.811

Classification of Mauritian LFs on the basis of their possible relatedness with the SF

- Three well populated classes with a high entropy (# 2, 4, 5)

☞ For verbs whose LF ends in -te, -ʁe or -le, the SF is quite unpredictable

- Even for the remaining verbs in -e the predictibility is far from being total

Why this result?

- Compare the French situation:

class	patterns	example		# of lex.	entropy
1	$\{X\text{e} \rightarrow X\}$	asyme	asym	1279	0.0
2	$\{X\text{je} \rightarrow X\text{i}, X\text{je} \rightarrow X, X\text{e} \rightarrow X\}$	pije	pij	171	1.515
3	$\{X\text{le} \rightarrow \text{v}X, X\text{e} \rightarrow X\}$	ale	va	153	0.057
4	$\{X\text{iʁ} \rightarrow X, X\text{ʁ} \rightarrow X\}$	finiʁ	fini	142	0.313
5	$\{X\text{dʁ} \rightarrow X, X\text{ʁ} \rightarrow X\}$	kudʁ	ku	55	0.0
6	$\{X\text{tiʁ} \rightarrow X, X\text{iʁ} \rightarrow X, X\text{ʁ} \rightarrow X\}$	paʁtiʁ	paʁ	33	0.994
7	$\{X\text{tʁ} \rightarrow X, X\text{ʁ} \rightarrow X\}$	konɛtʁ	konɛ	32	0.0
8	$\{X\text{ɥe} \rightarrow X\text{y}, X\text{e} \rightarrow X\}$	tɥe	ty	31	0.0
9	$\{X\text{əniʁ} \rightarrow X, X\text{jɛ̃} \rightarrow X, X\text{ʁ} \rightarrow X\}$	vəniʁ	vj ɛ	22	0.0
10	$\{X\text{ʁ} \rightarrow X\}$	fɛʁ	fE	21	0.0
...	...	...	...	...	...
(22 other classes with less than 20 members)					

Classification of French INFs on the basis of their possible relatedness with the PRS.3SG

- The infinitive is an excellent predictor of the present, except for verbs ending in -je or in -tir
- For the vast majority of verbs (73% of the 2079 most frequent) there is no uncertainty at all

Morphology vs. phonological information

- These results were initially puzzling to us because of a previous study.
- We trained Albright's (2002) *Minimal Generalization Learner* on French and Mauritian, to see how good it was at inferring the form of particular verbs
- The MGL is known for capturing efficiently morphophonological generalizations on the lexicon, in a way that correlates with experimental studies (Albright, 2003; Albright & Hayes, 2003).
- Results:

(Mauritian) LF ↦ SF	(French) INF ↦ PRS	(Mauritian) SF ↦ LF	(French) PRS ↦ INF
96.82%	96.27%	93.18%	90.70%

- We submit that the results of the MGL exhibit the effects of statistical lexical phonological knowledge rather than morphological knowledge

☞ Lexemes whose stems sound alike tend to follow the same morphological patterns; the MGL captures this.

A partial confirmation

- To confirm this intuition, we made a new series of entropy calculations.
- We want to evaluate how useful knowledge of the phonological shape of the input cell is when predicting the output cell.
- To do this we look at the conditional entropy of the patterns of relatedness between two cells given knowledge of the *n* last segments of the input cell.
- Results:

	(Mauritian) LF ↦ SF	(French) INF ↦ PRS	(Mauritian) SF ↦ LF	(French) PRS ↦ INF
last segment	0.950	1.154	0.430	0.566
last 2 segments	0.471	0.462	0.290	0.271
last 3 segments	0.043	0.149	0.094	0.086
last 4 segments	0.022	0.052	0.035	0.016

- Once we look at more than one segment (which typically corresponds to an inflectional affix) we see find very little difference between the two languages

☞ In both cases, phonological family resemblance between words helps in the same way

Outline

Conclusion

- Mauritian conjugation is undisputably simpler than French conjugation in some respects
  - ☞ Paradigm size, number of features, number of processes
- Mauritian seems more opaque morphosyntactically than French.
  - ☞ But currently no way of measuring this precisely
- Interpredictibility:
  - We propose an information-theoretic metric for specifically morphological aspects of interpredictibility
  - In this respect Mauritian is more complex than French
- Lessons:
  - There are many dimensions to morphological complexity.
  - Thus it is not self-evident that creoles are less complex than their lexifiers in all dimensions.
  - Any quantitative measure can (and should) be evaluated critically, but they are the only way of making meaningful claims in this area.

References

Ackerman, F., Blevins, J. P., and Malouf, R. (2009). 'Parts and wholes: implicative patterns in inflectional paradigms'. In J. P. Blevins and J. Blevins (eds.), *Analogy in Grammar*. Oxford: Oxford University Press, 54–82.

Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.

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.

——— (2003). 'Supplétion et classes flexionnelles dans la conjugaison du français'. *Langages*, 152:102–126.

——— (2007). 'French pronominal clitics and the design of Paradigm Function Morphology'. In *Proceedings of the fifth Mediterranean Morphology Meeting*. 291–322.

Carpooran, A. (2009). *Diksioner Morisien*. Sainte Croix (Mauritius): Koleksion Text Kreol.

Henri, F. (2010). *A Constraint-Based Approach to verbal constructions in Mauritian*. Ph.D. thesis, University of Mauritius and Université Paris Diderot.

Henri, F. and Abeillé, A. (2008). 'Verb form alternations in Mauritian'. In S. Müller (ed.), *Proceedings of the 15th Conference on HPSG*. Stanford: CSLI Publications, 378–398.

Henri, F., Marandin, J.-M., and Abeillé, A. (2008). 'Information structure coding in Mauritian: Verum Focus expressed by long forms of verbs'. Paper presented at the Workshop on Predicate Focus, Verum Focus, Verb Focus.

Maiden, M. (1992). 'Irregularity as a determinant of morphological change'. *Journal of Linguistics*, 28:285–312.

——— (2005). 'Morphological autonomy and diachrony'. In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 2004*. Dordrecht: Kluwer, 137–175.

New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). 'Une base de données lexicales du français contemporain sur internet: Lexique'. *L'Année Psychologique*, 101:447–462.

Pirelli, V. and Battista, M. (2000). 'The paradigmatic dimension of stem allomorphy in italian verb inflection'. *Rivista di Linguistica*, 12.

Plag, I. (2006). 'Morphology in Pidgins and Creoles'. In K. Brown (ed.), *Encyclopedia of Language and Linguistics, 2nd Edition*, vol. 8. 304–308.

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.

Verkuyl, H., Vet, C., Borillo, A., Bras, M., Le Draoulec, A., Molendijk, A., Swart, H. d., Vetters, C., and Vieu, L. (2004). 'Tense and aspect in sentences'. In F. Corblin and H. d. Swart (eds.), *Handbook of French Semantics*. Stanford: CSLI Publications, 233–270.