# A morphologists perspective on Creole complexity

Olivier Bonami[1] Ana R. Luís[2]

[1] U. Paris-Sorbonne & IUF & Laboratoire de Linguistique Formelle (U. Paris Diderot & CNRS)
olivier.bonami@paris-sorbonne.fr

[2] Universidade de Coimbra & CELGA
aluis@fl.uc.pt

Rethinking Creole Morphology
CIL 19, Genève, July 2013

# Introduction

- Morphology has figured prominently in discussions about the simplicity/complexity of creole grammars.
- General claim: creoles are 'simpler' than non-creoles because they have 'simpler' morphology.
- In the domain of inflection, simplicity is formulated as follows:
  1. very little/simple inflectional morphology (McWhorter, 2001)
  2. less inflectional morphology (making fewer distinctions) than the lexifier (Siegel, 2004; Plag, 2006; Holm, 2007).
  3. loss of most of the lexifier inflections, preservation of some and/or development of other (Becker and Veenstra, 2003; Kihm, 2003; Siegel, 2004).
- From a general linguist's perspective, are two problems with this line of argumentation.

# Problem 1: size and structure

- The complexity of an inflectional system does not reduce to:
  - The number of inflected word forms
  - number of inflectional morphs per word
  - number of inflectional affixes per language

☞ Of course, without the presence of these there is no inflection at all!

- Research on inflectional complexity in non-creole languages focuses on the **structure** of the system rather than its size.
  - ☞ Finkel and Stump (2007); Corbett (2007); Ackerman et al. (2009); Baerman (2012); Brown and Evans (2012); Walther (2013); Stump and Finkel (in press); Ackerman and Malouf (in press); Bonami and Luís (to appear), etc.
- What is new in these studies is the potential for a quantitative take on the structure of inflectional systems.

# Problem 2: unknown correlations

- Underlying the debate (e.g. bebtween McWhorter, 2001; DeGraff, 2001) is an asumption that morphological complexity is a sign of overall complexity.
- Yet it is far from evident that such a correlation holds.
- Two directly relevant recent studies:
  - Nichols (2009):
    - Studies hand-picked quantitative properties in a carefully sampled set of languages
    - Observes a positive correlation between degree of synthesis on verbs and syntactic complexity (multiplicity of basic word orders / alignment types)
    - Problem: arbitrary choice of quantitative measures
  - Moscoso del Prado Martín (2011):
    - Evaluates the inflectional and syntactic complexity of large parallel corpora in different languages
    - Applies information-theoretic measures of complexity to modified versions of the corpora that remove all inflection or all syntax
    - Conclusion: variation in morphological and syntactic complexity, but these balance out and the overall complexity is nearly identical in all 5 languages.
    - Problem: very skewed sample of languages (only Romance and Germanic)

# Taking stock

- While there are interesting attempts to address linguistic complexity in language as a whole, we are still far from having consensual objective measures that could then be applied to the particular case of creoles.
- Even if we did, two major methodological difficulties:
  - Such measures require access to linguistic resources (large lexica, large tagged corpora) that are unlikely to ever be available for many creoles
  - The sample of existing creole languages is hopelessly small and skewed, so that any quantitative claim on what is true of 'possible creole languages' is suspicious.
- Thus it is not clear that we will learn something on the complexity of creoles by looking at the complexity of their morphology.
- This does not entail that studying various aspects of the inflection of creoles in quantitative terms is pointless.
- Arguably this is a more interesting endeavour: learning something about the structure of the languages.

# Goals of this talk

- Examine quantifiable aspects of creole morphology by applying the same concepts and methodology that have been applied to non-creole languages, without making claims about the complexity of the creoles as a whole
- Apply the methodology to the verbal inflection of
  - Mauritian French
  - Korlai Indo-Portuguese
- Provide evidence which clearly shows that
  - The same complex inflectional phenomena found in non-creoles are also found in creole languages
  - That complexity can be quantified
  - There is no compelling evidence that creoles are more or less complex overall from that point of view.

# Outline

# Two main dimensions of complexity

1. Size — Malouf & Ackerman's 'enumerative complexity'
   - How many morphs in a word
   - How many cells in a paradigm
   - How many inflection classes in the system
2. Predictibility
   - Difficulty of predicting one form from another form

     Malouf & Ackerman's 'integrative complexity'
   - Difficulty of predicting the feature content of a word from its form
   - Difficulty of predicting the form of a word from its feature content

- In principle,
  - Size is a challenge for memory: many forms to learn
  - Lack of predictibility is a challenge for processing, given imperfect knowledge, ambiguity, and noise.

# A typical mildly complex system: Czech

- Sample of Czech declensions:

| | | I (M.ANIM) | II (M.ANIM) | III (M.INAN) | IV (F) | V (F) | VI (N) |
|---|---|---|---|---|---|---|---|
| SG | NOM | host | lingvista | most | věta | kost | město |
| | GEN | hosta | lingvisty | mostu | věty | kosti | města |
| | DAT | hostovi, hostu | lingvistovi | mostu | větě | kosti | městu |
| | ACC | hosta | lingvistu | most | větu | kost | město |
| | VOC | hoste | lingvisto | moste | věto | kosti | město |
| | LOC | hostovi, hostu | lingvistovi | mostu, mostě | větě | kosti | městě, městu |
| | INS | hostem | lingvistou | mostem | větou | kostí | městem |
| PL | NOM | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | GEN | hostů, hostí | lingvistů | mostů | vět | kostí | měst |
| | DAT | hostům | lingvistům | mostům | větám | kostem | městům |
| | ACC | hosty | lingvisty | mosty | věty | kosti | města |
| | VOC | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | LOC | hostech | lingvistech | mostech | větách | kostech | městech |
| | INS | hosty | lingvisty | mosty | větami | kostmi | městy |
| | | `guest' | `linguist' | `bridge' | `sentence' | `bone' | `town' |

# A typical mildly complex system: Czech

- Syncretism: in some declension, cells *A* and *B* use the same exponent.

$\Rightarrow$ The exponent carried by a form does not help disambiguate the syntactic function.

| | | I (M.ANIM) | II (M.ANIM) | III (M.INAN) | IV (F) | V (F) | VI (N) |
|---|---|---|---|---|---|---|---|
| SG | NOM | host | lingvista | most | věta | kost | město |
| | GEN | hosta | lingvisty | mostu | věty | kosti | města |
| | DAT | hostovi, hostu | lingvistovi | mostu | větě | kosti | městu |
| | ACC | hosta | lingvistu | most | větu | kost | město |
| | VOC | hoste | lingvisto | moste | věto | kosti | město |
| | LOC | hostovi, hostu | lingvistovi | mostu, mostě | větě | kosti | městě, městu |
| | INS | hostem | lingvistou | mostem | větou | kostí | městem |
| PL | NOM | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | GEN | hostů, hostí | lingvistů | mostů | vět | kostí | měst |
| | DAT | hostům | lingvistům | mostům | větám | kostem | městům |
| | ACC | hosty | lingvisty | mosty | věty | kosti | města |
| | VOC | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | LOC | hostech | lingvistech | mostech | větách | kostech | městech |
| | INS | hosty | lingvisty | mosty | větami | kostmi | městy |
| | | `guest' | `linguist' | `bridge' | `sentence' | `bone' | `town' |

# A typical mildly complex system: Czech

- Paradigmatic opacity: two declensions have the same exponent in cell *A* but different exponents in cell *B*.

⇒ When facing the *A* form of a novel lexeme, hard to predict the *B* form L.

| | | I (M.ANIM) | II (M.ANIM) | III (M.INAN) | IV (F) | V (F) | VI (N) |
|---|---|---|---|---|---|---|---|
| | NOM | host | lingvista | most | věta | kost | město |
| | GEN | hosta | lingvisty | mostu | věty | kosti | města |
| | DAT | hostovi, hostu | lingvistovi | mostu | větě | kosti | městu |
| SG | ACC | hosta | lingvistu | most | větu | kost | město |
| | VOC | hoste | lingvisto | moste | věto | kosti | město |
| | LOC | hostovi, hostu | lingvistovi | mostu, mostě | větě | kosti | městě, městu |
| | INS | hostem | lingvistou | mostem | větou | kostí | městem |
| | NOM | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | GEN | hostů, hostí | lingvistů | mostů | vět | kostí | měst |
| | DAT | hostům | lingvistům | mostům | větám | kostem | městům |
| PL | ACC | hosty | lingvisty | mosty | věty | kosti | města |
| | VOC | hosté, hosti | lingvisté, lingvisti | mosty | věty | kosti | města |
| | LOC | hostech | lingvistech | mostech | větách | kostech | městech |
| | INS | hosty | lingvisty | mosty | větami | kostmi | městy |
| | | `guest' | `linguist' | `bridge' | `sentence' | `bone' | `town' |

# Large but predictable: Tundra Nenets

- Nominal declension in Tundra Nenets (Uralic, Samoyedic; see Salminen (1997))
- Every noun can appear in one of 4 declensions:
  - The absolute declension: 21 forms, 17 synthetic

| | NOM | ACC | GEN | DAT | LOC | ABL | PROS |
|---|---|---|---|---|---|---|---|
| SG | ∅ | -m | -h | -n°h | -x°na | -xød | -w°na |
| DU | -x°h | -x°h | -x°h | | periphrastic | | |
| PL | -q | ∅ | -q | -x°q | -x°qna | -xøt° | -qm°na |

  - The predestinative declension: 27 forms, all synthetic

| | SINGULAR | | | DUAL | | | PLURAL | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| NOM | -døm° | -dør° | -d°da | -d°m'ih | -d°r'ih | -d°d'ih | -d°maq | -d°raq | -d°doh |
| ACC | -døm° | -dømt | -d°mta | -d°m'ih | -d°mt'ih | -d°mt'ih | -d°maq | -d°mtaq | -d°mtoh |
| GEN | -døn° | -dønt° | -d°nta | -d°n'ih | -d°nt'ih | -d°nt'ih | -d°naq | -d°ntaq | -d°ntoh |

Paradigm of predestinative nouns

(Predestinative codes person and number of a beneficiary, but not number of the noun)

# The paradigm of possessed nouns

- The possessive declension: 189 forms, 153 synthetic

| | | 1sg | 2sg | 3sg | 1du | 2du | 3du | 1pl | 2pl | 3pl |
|---|---|---|---|---|---|---|---|---|---|---|
| | NOM | -m° | -r° | -da | -m'ih | -r'ih | -d'ih | -maq | -raq | -doh |
| | ACC | -m° | -mt | -mta | -m'ih | -mt'ih | -mt'ih | -maq | -mtaq | -mtoh |
| | GEN | -n° | -nt° | -nta | -n'ih | -nt'ih | -nt'ih | -naq | -ntaq | -ntoh |
| SG | DAT | -xøn° | -xønt° | -x°nta | -x°n'ih | -x°nt'ih | -x°nt'ih | -x°naq | -x°ntaq | -x°ntoh |
| | LOC | -x°nan° | -x°nant° | -x°nanta | -x°nan'ih | -x°nant'ih | -x°nant'ih | -x°nanaq | -x°nantaq | -x°nantoh |
| | ABL | -x°døn° | -x°dønt° | -x°dønta | -x°døn'ih | -x°dønt'ih | -x°dønt'ih | -x°dønaq | -x°døntaq | -x°døntoh |
| | PROS | -m°nan° | -m°nant° | -m°nanta | -m°nan'ih | -m°nant'ih | -m°nant'ih | -m°nanaq | -m°nantaq | -m°nantoh |
| | NOM | -x°yun° | -x°yud° | -x°yuda | -x°yun'ih | -x°yud'ih | -x°yud'ih | -x°yunaq | -x°yudaq | -x°yudoh |
| DU | ACC | -x°yun° | -x°yud° | -x°yuda | -x°yun'ih | -x°yud'ih | -x°yud'ih | -x°yunaq | -x°yudaq | -x°yudoh |
| | GEN | -x°yun° | -x°yut° | -x°yuta | -x°yun'ih | -x°yut'ih | -x°yut'ih | -x°yunaq | -x°yutaq | -x°yutoh |
| | | local cases | | | | periphrastic | | | | |
| | NOM | -n° | -d° | -da | -n'ih | -d'ih | -d'ih | -naq | -daq | -doh |
| | ACC | -n° | -d° | -da | -n'ih | -d'ih | -d'ih | -naq | -daq | -doh |
| | GEN | -qn° | -t° | -ta | -qn'ih | -t'ih | -t'ih | -qnaq | -taq | -toh |
| PL | DAT | -xøqn° | -xøt° | -x°ta | -x°qn'ih | -x°t'ih | -x°t'ih | -x°qnaq | -x°taq | -x°toh |
| | LOC | -x°qnan° | -x°qnant° | -x°qnata | -x°qnan'ih | -x°qnat'ih | -x°qnat'ih | -x°qnanaq | -x°qnataq | -x°qnatoh |
| | ABL | -x°tøn° | -x°tøt° | -x°tøta | -x°tøn'ih | -x°tøt'ih | -x°tøt'ih | -x°tønaq | -x°tøtaq | -x°tøtoh |
| | PROS | -qm°nan° | -qm°nat° | -qm°nata | -qm°nan'ih | -qm°nat'ih | -qm°nat'ih | -qm°nanaq | -qm°nataq | -qm°natoh |

# The paradigm of predicative nouns

- The predicative declension (a.k.a. nominal conjugation): 72 forms

| | SINGULAR | | | DUAL | | | PLURAL | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| AOR | -d°m | -n° | ∅ | -n'ih | -d'ih | -x°h | -waq | -daq | -q |
| PRET | -dømc'° | -nc'° | -s'° | -n'inc'° | -d'inc'° | -xønc'° | -wac'° | -dac'° | -c'° |

Paradigm of absolute predicative nouns

| | | POSSESSOR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | SINGULAR | | | DUAL | | | PLURAL | | |
| | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| AOR | SG | -m° | -r° | -da | -m'ih | -r'ih | -d'ih | -maq | -raq | -doh |
| | DU | -x°yun° | -x°yud° | -x°yuda | -x°yun'ih | -x°yud'ih | -x°yud'ih | -x°yunaq | -x°yudaq | -x°yudoh |
| | PL | -n° | -d° | -da | -n'ih | -d'ih | -d'ih | -naq | -daq | -doh |
| PRET | SG | -møs'° | -røs'° | -das'° | -m'inc'° | -r'inc'° | -d'inc'° | -mac'° | -rac'° | -donc'° |
| | DU | -x°yunøs'° | -x°yudøs'° | -x°yudas'° | -x°yun'inc'° | -x°yud'inc'° | -x°yud'inc'° | -x°yunac'° | -x°yudac'° | -x°yudonc'° |
| | PL | -nøs'° | -døs'° | -das'° | -n'inc'° | -d'inc'° | -d'inc'° | -nac'° | -dac'° | -donc'° |

Paradigm of possessed predicative nouns

# Tundra nenets: conclusions

- Large paradigms for nouns, with 269 synthetically expressed cells
- Requires about 25 affixes located in 8 position classes according to Ackerman et al. (2012)'s analysis
- Interesting cases of fusion, odd morphophonemics, etc.
- Sizable quantity of syncretism contributing to the complexity of the system.
- However, it is trivial to infer the whole paradigm for any single form: all nouns inflect alike.

# Gratuitous inflection can be harmless

- In itself, having inflection classes does not produce unpredictivity
- If any form is a predictor of class, then inflection classes are gratuitous but cheap.

| | class I | | class II | |
|---|---|---|---|---|
| gender | SG | PL | SG | PL |
| I | j- | s- | b- | t- |
| II | g- | s- | n- | t- |
| III | g- | j- | n- | b- |
| IV | j- | j- | b- | b- |
| V | j- | g- | b- | n- |
| VI | g- | g- | n- | n- |

Burmeso verbs (from Corbett, 2009, quoting Donohue 2001)

# Extreme unpredictibility: Nuer

Baerman (2012), using data from Frank 1999:

- 6 cells and 3 suffixes, but 25 different patterns
- This arguably makes the inflection system of little functional value.

| Pattern # | | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NOM | — | — | — | — | — | — | — | — | — | — | — | — | — |
| SG | GEN | — | -kä | -kä | — | — | — | -kä | — | -kä | — | -kä | -kä | — |
| | LOC | — | -kä | -kä | — | — | -kä | — | — | — | -kä | -kä | -kä | — |
| | NOM | — | — | -ni | -ni | — | — | -ni | — | — | -ni | — | — | — |
| PL | GEN | -ni | -ni | -ni | -ni | — | -ni | -ni | -ni | -ni | -ni | -ni | — | — |
| | LOC | -ni | -ni | -ni | -ni | — | -ni | -ni | — | -ni | -ni | — | -ni | -ni |

| Pattern # | | XIV | XV | XVI | XVII | XVIII | XIX | XX | XXI | XXII | XXIII | XXIV | XXV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NOM | — | — | — | — | — | — | — | — | — | — | — | — |
| SG | GEN | -kä | -kä | — | -ä | -ä | -kä | -kä | — | -ä | — | -kä | -kä |
| | LOC | -kä | — | -kä | -ä | -ä | -ä | -ä | -ä | -kä | -ä | -kä | -kä |
| | NOM | — | — | — | -ni | — | -ni | — | — | — | -ni | -ni | — |
| PL | GEN | — | -ni | — | -ni | -ni | -ni | -ni | -ni | -ni | -ni | — | -kä |
| | LOC | — | — | -ni | -ni | -ni | -ni | -ni | -ni | -ni | -ni | -ni | -ni |

# Outline

# The issue

- Creoles are definitely on the simple side in terms of size
- How do they fare in terms of predictibility?
- For creoles with no inflection at all, the issue is moot.
- Here we look at evidence from a few creole languages which do have inflection:
  - Korlai Indo-Portuguese Creole (data from Clements, 1996)
  - Mauritian Creole (data from Bonami et al., 2011)
- Our conclusion: where Creoles have inflection, it shows no sign of being more transparent than inflection in noncreoles.

# Predictibility in Korlai Creole

- 4 paradigm cells, 4 inflection classes
- Downsizing of the Portuguese system + innovation of a 4th class for loans of substratic origin

| | kata `sing' | bebe `drink' | irgi `get up' | lotu `push' |
|---|---|---|---|---|
| UNMARKED | katá | bebé | irgí | lotú |
| PAST | kató | bebéw | irgíw | lotú |
| GERUND | katán | bebén | irgín | lotún |
| COMPLETIVE | katád | bebíd | irgíd | lotúd |

Korlai Creole Portuguese
(adapted from Clements, 1996)

- Presence of syncretism in the **u** conjugation: UNMARKED vs. PAST
- Presence of opacity in the COMPLETIVE: **e** vs. **i** conjugation.

# Predictibility in Mauritian Creole

- Mauritian Creole verbs have two forms
- The alternation between these two forms is definitely morphology, as shown in detail in Henri (2010)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| LF | briye | briye | insiste | existe | fini | vini | paste | pas | bande | ban |
| SF | briy | briye | insiste | exis | fini | vinn | pas | pas | ban | ban |
| TRANS. | `glow' | `mix' | `insist' | `exist' | `finish' | `come' | `filter' | `pass' | `bandage' | `ban' |

- The alternation codes a complex array of syntactic, morphological and information-structure oppositions (Henri, 2010)
  - ☞ In the simpler cases: presence/absence of a nonclausal following complement

(1) a. *Nou res toultan malad.*
1PL stay.SF always sick
Lit. `We always remain sick.'

b. *Nou reste toultan.*
1PL stay.LF always
`We always stay.'

# Predictibility in Mauritian Creole

- The existence of a large dictionary Carpooran (2011) allows one to make more precise observations.
- Large number of inflection classes for a system of this side

| class | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| size | 1417 | 305 | 13 | 3 | 2 | 1 | 1 | 1 | 1 | 113 | 3 | 218 |
| LF | lave | adapte | bande | reste | tombe | ronfle | tramble | segonde | rantre | fini | vini | abat |
| SF | lav | adapte | bann | res | tom | ronf | tram | segon | rant | fini | vinn | abat |

- High prevalence of syncretism: 31% of verbs have identical long and short forms
- High prevalence of opacity:
  - 90% of verbs have a LF which does not identify inflection class
  - 80% of verbs have a SF which does not identify inflection class

# Outline

# The method

- Building on a strategy pioneered by Ackerman et al. (2009), we can use standard tools from information theory to quantify different aspects of predictivity.
  - ☞ Ackerman et al. (2009), Sims (2010), Bonami et al. (2011), Bonami and Luís (to appear), Ackerman and Malouf (in press), Blevins (to appear)
- Here we use simplified measures that are
  - Simpler to illustrate
  - Applicable given limited data
- However, scaling up is easy given appropriate data.

# Quantifying predictivity

- 
- The impact of syncretism on ambiguity can be evaluated by looking at the conditional probability of a paradigm cell given knowledge of a form and the identity of the lexeme filling that form.
- ☞ For the Czech noun HOST:

| SG | PL |
|---|---|
| $P(\text{case} = \text{NOM} \mid \text{form} = \textit{host}) = 1$ | $P(\text{case} = \text{NOM} \mid \text{form} = \textit{hosté}) = \frac{1}{2}$ |
| $P(\text{case} = \text{GEN} \mid \text{form} = \textit{hosta}) = \frac{1}{2}$ | $P(\text{case} = \text{GEN} \mid \text{form} = \textit{hostů}) = 1$ |
| $P(\text{case} = \text{DAT} \mid \text{form} = \textit{hostovi}) = \frac{1}{2}$ | $P(\text{case} = \text{DAT} \mid \text{form} = \textit{hostům}) = 1$ |
| $P(\text{case} = \text{ACC} \mid \text{form} = \textit{hosta}) = \frac{1}{2}$ | $P(\text{case} = \text{ACC} \mid \text{form} = \textit{hosty}) = \frac{1}{2}$ |
| $P(\text{case} = \text{VOC} \mid \text{form} = \textit{hoste}) = 1$ | $P(\text{case} = \text{VOC} \mid \text{form} = \textit{hosté}) = \frac{1}{2}$ |
| $P(\text{case} = \text{LOC} \mid \text{form} = \textit{hostovi}) = \frac{1}{2}$ | $P(\text{case} = \text{DAT} \mid \text{form} = \textit{hostech}) = 1$ |
| $P(\text{case} = \text{INS} \mid \text{form} = \textit{hostem}) = 1$ | $P(\text{case} = \text{ACC} \mid \text{form} = \textit{hosty}) = \frac{1}{2}$ |

- From this we can compute the conditional entropy of a cell given a form, which quantifies the unpredictibility.
- For HOST: $H(\text{cell} \mid \text{form}) = \frac{4}{7} \approx 0.57$

# Quantifying predictivity

- Averaging over all classes, we get

| Language fragment | Tundra Nenets | Czech | Nuer | Korlai | Mauritian |
|---|---|---|---|---|---|
| $H(\text{cell} \mid \text{form})$ | 0.91 | 0.93 | 1.26 | 0.125 | 0.25 |

- This is a much idealized calculation...
  - The paradigm cells do not have the same frequency, and hence the same probability
  - The paradigm cells are not equally likely to be left ambiguous by context
  - We should average over lexemes rather than inflection classes
  - etc.
- ...but
  - it captures the intuitions on relative complexity
  - it can readily be made more precise when relevant data is available

# Quantifying predictivity

- In the same fashion, we can quantify the impact of paradigmatic opacity on inflection class determination
- In our Czech sample, for a noun in the NOM.SG:

$$P(\text{class} = \text{I} \mid \text{exp} = \emptyset) = \tfrac{1}{3} \quad P(\text{class} = \text{II} \mid \text{exp} = \text{-}a) = \tfrac{1}{2}$$
$$P(\text{class} = \text{III} \mid \text{exp} = \emptyset) = \tfrac{1}{3} \quad P(\text{class} = \text{IV} \mid \text{exp} = \text{-}a) = \tfrac{1}{2}$$
$$P(\text{class} = \text{V} \mid \text{exp} = \emptyset) = \tfrac{1}{3} \quad P(\text{class} = \text{VI} \mid \text{exp} = \text{-}o) = 1$$

- From this we can compute the conditional entropy of class given exponent, for each paradigm cell.
- For the NOM.SG: $H(\text{class} \mid \text{exp}) = \frac{\log 3}{2} + \frac{1}{3} \approx 1.13$
- Averaging over all cells:

| Language fragment | Tundra Nenets | Czech | Nuer | Korlai | Mauritian |
|---|---|---|---|---|---|
| $H(\text{class} \mid \text{cell})$ | 0 | 1.03 | 3.68 | 0.125 | 2.66 |

- The same caveats as before apply

# Getting more accurate results

- Where the relevant data is available, we can get more accurate results by taking into account type frequency.
  - ☞ Drop the simplifying assumption that there is an equal chance for a lexeme to fall in any inflection class

Syncretism (with type frequency information):

- | Language fragment | Tundra Nenets | Czech | Nuer | Korlai | Mauritian |
  |---|---|---|---|---|---|
  | $H(\text{cell} \mid \text{form})$ | 0.91 | 0.89 | 1.36 | — | 0.31 |

- Opacity (with type frequency information):

| Language fragment | Tundra Nenets | Czech | Nuer | Korlai | Mauritian |
|---|---|---|---|---|---|
| $H(\text{cell} \mid \text{form})$ | 0 | 0.68 | 2.60 | — | 0.62 |

- For really well documented languages, we can use frequency data from a large tagged corpus to get even more accurate results.

☞ We won't get very far for creoles however.

# A better assesment of predictibility

- The present assesment of paradigm opacity has limitations.

| class | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| size | 1417 | 305 | 13 | 3 | 2 | 1 | 1 | 1 | 1 | 113 | 3 | 218 |
| LF | lave | adapte | bande | reste | tombe | ronfle | tramble | segonde | rantre | fini | vini | abat |
| SF | lav | adapte | bann | res | tom | ronf | tram | segon | rant | fini | vinn | abat |

  - Depends on questionable decisions on inflectional classification (segmentation, grain of the classification, number of classes, etc.)
  - Misses much phonotactic information
  - Does not correspond to a realistic practical problem for speakers
  - Attributes less predictivity to systems with more paradigm cells
- A better strategy (Ackerman et al., 2009):
  - Focus on predicting **one** cell rather than the whole paradigm
    - This is the Paradigm Cell Filling Problem: given exposure to one form of an unknown lexeme, what inferences can a competent speaker draw on other forms of that lexeme?
  - Average predictivity for all pairs of cells.

# A better assesment of predictibility

- Bonami et al. (2011) use a variant of that methodology to compare the level of opacity in Mauritian Creole and in French
  - 2078 verbs from Carpooran (2011) matched again the 2078 most frequent nondefective French verbs
  - Use of two fully transcribed inflected lexica
  - No precoded morphological analysis
  - Patterns of relatedness between pairs of forms inferred using a variant of the Minimal Generalization strategy (Albright and Hayes, 2003)
  - We now look at the value, for all pairs of cells (u,v), of $H([u \Rightarrow v]|u)$
- Results:

| language | Mauritian | French |
|---|---|---|
| average cond. entropy over all pairs | 0.744 | 0.446 |
| minimal cond. entropy | 0.563 | 0 |
| maximal cond. entropy | 0.925 | 0.916 |

- In terms of paradigm opacity, Mauritian is decidedly more complex than its lexifier.

# Conclusions

- We have argued that:
  - The issue of creole complexity is interesting to the extent that it teaches us something on the structure of the languages
  - Local complexity (of a particular factor in a particular dimension of grammar) is easier to assess and more informative than global complexity (of the whole linguistic system)
  - Creoles should be compared to the full typology of noncreole languages
- We have shown that:
  - There are at least two important dimensions to the complexity of inflection systems: size and predictibility
  - While creole inflection systems are definitely small on average, they exhibit the very same complex phenomena found in noncreole languages.
  - Where we have relevant data, the evidence does not show creoles to be more predictable than those of noncreoles

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.

Ackerman, F., Bonami, O., and Nikolaeva, I. (2012). `Systemic polyfunctionality and morphology-syntax interdependencies'. Presented at the Conference on Defaults in Morphological Theory, Lexington, Kentucky.

Ackerman, F. and Malouf, R. (in press). `Morphological organization: the low conditional entropy conjecture'. *Language*, 89.

Albright, A. C. and Hayes, B. P. (2003). `Rules vs. analogy in english past tenses: A computational/experimental study'. *Cognition*, 90:119--161.

Baerman, M. (2012). `Paradigmatic chaos in Nuer'. *Language*, 88:467--494.

Becker, A. and Veenstra, T. (2003). `The survival of inflectional morphology in French-related Creoles.' *SSLA*, 25:285--306.

Blevins, J. P. (to appear). *Word and Paradigm Morphology*. Oxford: Oxford University Press.

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.

Bonami, O. and Luís, A. R. (to appear). `Sur la morphologie implicative dans la conjugaison du portugais : une étude quantitative'. *Mémoires de la Société de Linguistique de Paris*, 22.

Brown, D. and Evans, R. (2012). `Morphological complexity and unsupervised learning: validating Russian inflectional classes using high frequency data'. In F. Kiefer, M. Ladányi, and P. Siptár (eds.), *Current Issues in Morphological Theory: (Ir)regularity, analogy and frequency*. Amsterdam: John Benjamins, 135--162.

Carpooran, A. (2011). *Diksioner Morisien*. Sainte Croix (Mauritius): Koleksion Text Kreol, 2nd Edition.

Clements, J. C. (1996). *The Genesis of a Language: The Formation and Development of Korlai Portuguese*. Amsterdam: John Benjamins.

Corbett, G. G. (2007). `Canonical typology, suppletion and possible words'. *Language*, 83:8--42.

--------- (2009). `Canonical inflection classes'. In F. Montermini, G. Boyé, and J. Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*. Somerville: Cascadilla Press, 1--11.

DeGraff, M. (2001). `Morphology in creole genesis: Linguistics and ideology'. In *Ken Hale: A Life in Language*. Cambridge: MIT Press, 53--121.

Finkel, R. and Stump, G. T. (2007). `Principal parts and morphological typology'. *Morphology*, 17:39--75.

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

Holm, J. (2007). `Creolization and the fate of inflections'. In T. Stolz, D. Bakker, and R. Palomo (eds.), *Aspects of language contact: new theoretical, methodological and empirical findings with special focus on Romanisation processes*. Mouton de Gruyter.

Kihm, A. (2003). `Inflectional categories in creole languages'. In I. Plag (ed.), *Phonology and Morphology in Creole Languages*. Tübingen: Niemeyer, 333--363.

McWhorter, J. (2001). `The world's simplest grammars are creole grammars'. *Linguistic Typology*, 5:125--166.

Moscoso del Prado Martín, F. (2011). `The mirage of morphological complexity'. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*. 3524--3529.

Nichols, J. (2009). `Linguistic complexity: a comprehensive definition and survey'. In G. Sampson, D. Gil, and P. Trudgill (eds.), *Language complexity as an evolving variable*. Oxford: Oxford University Press, 110--125.

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

Salminen, T. (1997). *Tundra Nenets Inflection*, vol. 227 of *Mémoires de la Société Finno-Ougrienne*. Helsinki: Suomalais-Ugrilainen Seura.

Siegel, J. (2004). `Morphological simplicity in pidgins and creoles'. *Journal of Pidgin and Creole Languages*, 19:139--162.

Sims, A. (2010). `Probabilistic paradigmatics: Principal parts, predictability and (other) possible particular pieces of the puzzle'. Paper presentend at the Fourteenth International Morphology Meeting, Budapest.

Stump, G. T. and Finkel, R. (in press). *Morphological Typology: From Word to Paradigm*. Cambridge: Cambridge University Press.

Walther, G. (2013). *De la canonicité en morphologie: perspective empirique, théorique et computationnelle*. Ph.D. thesis, Université Paris Diderot.

# Another dimension: periphrasis

- Opacity, syncretism and overabundance introduce complexity insofar as they undermine the functional value of having inflection rather than not.
- Another property with the same character: presence of inflectional periphrasis
- In Tundra Nenets, in the dual, local cases are expressed by combining an inflected postposition with a genitive form of the noun.

| | NOM | ACC | GEN | DAT | LOC | ABL | PROS |
|---|---|---|---|---|---|---|---|
| SG | ∅ | -m | -h | -n°h | -x°na | -xød | -w°na |
| DU | -x°h | -x°h | -x°h | ti-x°h n'ah | ti-x°h n'ana | ti-x°h n'ad° | ti-x°h n'amna |
| PL | -q | ∅ | -q | -x°q | -x°qna | -xøt° | -qm°na |

- The potential complexity here is the blur between morphology and syntax introduced by periphrastic inflection: a single category is sometimes expressed by morphological means, sometimes by syntactic means.