Towards a robust assessment of implicative relations in inflectional systems

Olivier Bonami

U. Paris-Sorbonne
Institut Universitaire de France
Laboratoire de Linguistique Formelle

Computational approaches to morphological complexity
Surrey Morphology Group @ U. Paris-Sorbonne, February 2013

Defining implicative structure

- Inflectional paradigms have what Wurzel (1984) calls implicative structure.

> *The inflectional paradigms are, as it were, kept together by implications. There are no paradigms (except highly extreme cases of suppletion) that are not based on implications valid beyond the individual word, so that we are quite justified in saying that inflectional paradigms generally have an implicative structure, regardless of deviations in the individual cases.*
>
> Wurzel (1989, 114)

- Discussions of implicative structure usually focus on hard cases, but as Wurzel emphasizes, implicative structure is present even in trivial paradigms.
- A trivial example: if an English verb has Xing as its present participle, then its bare infinitive is X.

☞ Implicative structure is an empirical property of paradigms, not a theoretical hypothesis on the nature of morphology.

Illustrations: simple implications

lexeme	INF	PRS.1PL	PRS.2PL	IPFV.1PL	IPFV.2PL
LAVER wash	lave	lavɔ̃	lave	lavjɔ̃	lavje
DIRE say	diʁ	dizɔ̃	dit	dizjɔ̃	dizje
PEINDRE paint	pɛ̃dʁ	pɛɲɔ̃	pɛɲe	pɛɲɔ̃	pɛɲe
POUVOIR can	puvwaʁ	puvɔ̃	puve	puvjɔ̃	puvje

- The IPFV.1PL is Xɔ̃ if and only if the IPFV.2PL is Xe
  ⇒ general, bidirectional, categorical
- If the PRS.2PL is Xe, then the PRS.1PL is Xɔ̃.
  ⇒ general, monodirectional, categorical
- If the PRS.1PL is Xɔ̃, then the PRS.2PL is Xe.
  ⇒ general, monodirectional, almost categorical
- If the PRS.1PL is Xɔ̃, then the INF is Xe.
  ⇒ general, monodirectional, noncategorical
- If the INF is Xɛ̃dʁ, then the IPFV.1PL is Xɛɲɔ̃.
  ⇒ local, monodirectional, categorical
- If the INF is Xwar, then the IPFV.1PL is Xɔ̃.
  ⇒ local, monodirectional, noncategorical

Implications with a disjunctive consequent

- In many cases, noncategorical implications come in families, which can be grouped using disjunction in the consequent.
- Typical example: dropped theme vowels in Latin

conj.	1SG	2SG	3SG	1PL	2PL	3PL
I	amō	amās	amat	amāmus	amātis	amant
II	deleō	delēs delet	delēmus	delētis	delent	
III	legō	legis	legit	legimus	legitis	legunt
IIIm	capiō	capis	capit	capimus	capitis	capiunt
IV	audiō	audīs	audit	audīmus	audītis	audiunt

☞ If the PRS.1SG is in XCō, then the PRS.1PL is either in XCāmus or in XCīmus

- Knowing the likelihood of each possible outcome is relevant.

Implications with a complex antecedent

- Many interesting implications mention 2 paradigm cells in the antecedent

lexeme	INF	PRS.2PL	PST.PTCP
LAVER wash	lave	lave	lave
LISSER smooth	lise	lise	lise
FINIR finish	finiʁ	finise	fini
TONDRE mow	tɔ̃dʁ	tɔ̃de	tɔ̃dy
MORDRE bite	mɔʁdʁ	mɔʁde	mɔʁdy
SORTIR go out	sɔʁtiʁ	sɔʁte	sɔʁti
MOURIR die	muʁiʁ	muʁe	mɔʁ

☞ If the INF is Xiʁ and the PRS.2PL is Xise, the PST.PTCP is always Xi.

☞ If the INF is XCiʁ and the PRS.2PL is XCe, the PST.PTCP is most often XCy.

- We call such things binary implicative relations
- *n*-ary implicative relations underly the idea of principal parts: sets of *n* cells from which a categorical implication exists to all other cells.

The information-theoretic perspective

- Research program laid out in (Ackerman et al., 2009):
  - Use of information-theoretic tools to study implicative structure directly
  - Further applied and developed in (Sims, 2010; Malouf and Ackerman, 2010; Bonami et al., 2011)
  - Closely related to but distinct from the research program laid out in (Stump and Finkel, in press)
- This research program was elaborated in the context of a debate on the primacy of morphotactic or implicative structure
  - Blevins (2006), Ackerman et al. (2009), Baerman and Corbett (2012), (Stump and Finkel, in press, chap. 9)
- The focus here is different: use the same tools as a way of exploring what implicative structure there is.
  - How it is most perspicuously modeled is an important but separate issue.

The general approach

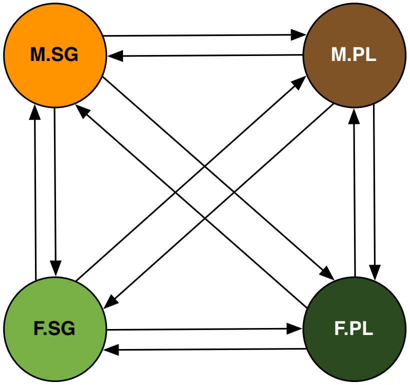
- Instrumented descriptive morphology:
  - Fully implemented analyses (with help from Gilles Boyd Delphine Tribout)
  - Applied to real-size datasets (thousands of lexemes)
- For practical reasons, focus on French for now
- Based on `flexique`, a new inflectional lexicon of French (Bonami et al., in preparation) derived from `Lexique 3` (New et al., 2007)

POS	lexemes	words
nouns	33,716	67,353
adjectives	11,420	45,680
verbs	5,325	271,575
total	50,461	384,608

Structure

French adjectives

- Looking at French adjectival paradigms and disregarding M.SG liaison forms, there are 12 relationships from one cell to another to explore:



Zoom in: [M.SG ⇒ M.PL]

- There are exactly two patterns of alternation relating M.SG to M.PL

#	description	examples		
		lexeme	M.SG	M.PL
p_1	Xal $\sim$ Xo	LOYAL	lwajal	lwajo
p_2	$X \sim X$	CALME	kalm	kalm
		BANAL	banal	banal

- There are exactly two relevant classes of M.SG which exhibit different behavior:
  - Words ending in -al
  - Words not ending in -al
- These are the relevant classes because they determine what patterns are eligible: words that do not end in -al can't follow p_1, but words that do can follow p_2.

Unary implicative relations

- A unary implicative relation expresses the likelihood of different forms filling cell B for a coherent class of forms filling cell A
- A unary implication array is a set of unary implicative relations whose antecedents constitute a partition of the set of A forms.

class	description	patterns	examples		
			lexeme	M.SG	M.PL
C_1	ending in al	$p_1 : X\mathrm{al} \sim X\mathrm{o}$	LOYAL	lwajal	lwajo
		$p_2 : X \sim X$	BANAL	banal	banal
C_2	not ending in al	$p_2 : X \sim X$	CALME	kalm	kalm

The unary implication array [M.SG ⇒ M.PL]

- Important decisions:
  - How do we infer the patterns?
  - How do we estimate the likelihood of a particular outcome?

Inferring the patterns

- We borrow the strategy of the Minimal Generalization Learner (Albright, 2002).
  - Assume a decomposition of segments into distinctive features.
  - Assume that each pair of forms is related by a single SPE-style rule (Chomsky and Halle, 1968).
  - For each $\langle \text{INPUT}, \text{OUTPUT} \rangle$ pair:
    Determine the most specific rule $A \rightarrow B/\#C\_\_D\#$ such that

    $$\text{INPUT} = CAD \text{ and } \text{OUTPUT} = CBD,$$

    maximizing C and minimizing A.
  - For each set of rules R sharing the same structural change $A \rightarrow B$:
    Determine the least general rule of the form

    $$r = A \rightarrow B/(\#|X)[\text{feat}^{+}]^{*}\text{seg}^{*}\_\_\_\text{seg}^{*}[\text{feat}^{+}]^{*}(Y|\#)$$

    such that all rules in R are specializations of r.

Inferring the patterns: example

- As the program explores the lexicon, it computes incrementally more general rules.

input	output	rule	
final	fino	al → o /	#fin___#
penal	peno	al → o /	#C[−voice]V[+high,−back]n___#
vɛʁbal	vɛʁbo	al → o /	X[+voice]C[+voice]___#
djalɛktal	djalɛkto	al → o /	C___#
aʁeal	aʁeo	al → o /	___#

- Order of presentation does not matter
- Tractable computation: for n structural changes, $n-1$ rule comparisons in the worst case.
- This is a rather crude method (e.g. won't do well on discontinuous inflection) but sufficient for present purposes

Estimating the likelihood of the choice of a pattern

- Using type frequency information from `flexique`, we can estimate the conditional probability of a pattern given a class

class	size	patterns	freq.	examples		
				lexeme	M.SG	M.PL
C_1	428	Xal $\sim$ Xo	399	LOYAL	lwajal	lwajo
		$X \sim X$	29	BANAL	banal	banal
C_2	8797	$X \sim X$	8797	CALME	kalm	kalm

$$p(C_1) = \frac{428}{9225} \approx 0.046 \qquad p(X\text{al} \sim X\text{o} | C_1) = \frac{399}{428} \approx 0.932$$

$$p(X \sim X | C_1) = \frac{29}{428} \approx 0.068$$

$$p(C_2) = \frac{8797}{9225} \approx 0.954 \qquad p(X \sim X | C_2) = 1$$

- The distribution of these conditional probabilities is our model of the implication array.

Using conditional entropy as a summary of the distribution

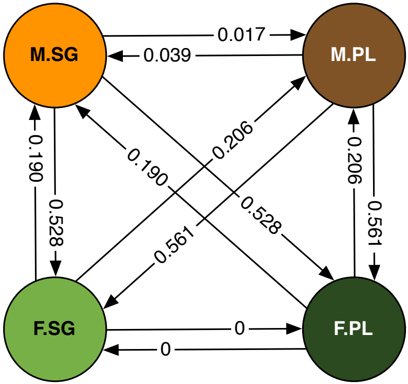
$$H(Y \mid X) = -\sum_{x \in X} P(x) \left(\sum_{y \in Y} P(y \mid x) \log_2 P(y \mid x) \right)$$

- Positive number that grows as uncertainty rises
  - Rises with the number of possible outcomes
  - Rises when the probabilities are distributed more uniformly
- Calibrated so that for 2^n equiprobable possibilities, entropy is n.
- Here:

$$\begin{aligned} H(\text{M.SG} \sim \text{M.PL} \mid \text{M.SG}) &= -\left(\frac{428}{9225} \left(\frac{399}{428} \log_2 \frac{399}{428} + \frac{29}{428} \log_2 \frac{29}{428} \right) + \frac{8797}{9225} (1 \times \log_2 1) \right) \\ &\approx -\left(\frac{428}{9225} \times 0.357 + \frac{8797}{9225} \times 0 \right) \\ &\approx 0.017 \end{aligned}$$

French adjectives: unary implication arrays

- Entropy values for French adjectives:



- $H([\text{F.SG} \Rightarrow \text{F.PL}]) = H([\text{F.PL} \Rightarrow \text{F.SG}]) = 0$: full interpredictibility.

An important caveat

- Entropy is a summary of a probability distribution.
  - Thus there can be structure in the distribution that it masks.
  - In the case of [M.SG ⇒ M.PL]: all the uncertainty is located in a definite corner of the search space, forms ending in -al.
  - The same entropy could have been obtained with scattered irregularities.

[M.SG ⇒ F.SG]: patterns

- For [M.SG ⇒ F.SG] the distribution is very different:
  - 26 patterns:

Pattern			freq.
ϵ→ϵ	/	__#	6153
ϵ→ʁ	/	{ʃ,ʒ,j}{e,ɛ}__#	110
ϵ→t	/	[+son,−lat]__#	1178
ϵ→z	/	[+voc,−cons,−nas]__#	506
ϵ→d	/	[−cons,−high]__#	133
ϵ→s	/	__#	22
ϵ→ʃ	/	#{p,b,f,v},{l,r},{ɛ,a,ɛ̃,ɑ̃}__#	3
f→v	/	[+voc,−cons,−nas,−low]__#	271
ɑ̃→ɑn	/	__#	29
ɛ̃→ɛn	/	__#	339
ɛ̃→in	/	[+cons]__#	94
ɔ̃→on	/	[+cons],[−voc]__#	38
~→yn	/	[+voice][+cons,−high]__#	7

Pattern			freq.
ʁ→ʁis	/	t__#	164
ʁ→z	/	[+cons]__#	153
ϵ→ɛs	/	[+son][+cons][−back]ʁ__#	6
o→ɛl	/	[+cons,+ant]__#	4
ϵ→kt	/	[−cons,+voc,−low]{ɛ,ɛ̃}__#	4
u→ol	/	#{p,b,f,v,m}__#	2
ϵ→g	/	lɔ̃__#	2
ϵ→l	/	#su__#	2
ϵ→j	/	#ʒɑ̃ti__#	1
→ɛj	/	#vj__#	1
ɛ̃→iɲ	/	#ben__#	1
ϵ→v	/	#sɛʁ__#	1
ʁ→ʃ	/	#sɛ__#	1

[M.SG ⇒ F.SG]: classes

class	size	patterns	frees	examples		
				lexeme	M.SG	M.PL
C_1	3439	$\epsilon \rightarrow \epsilon$	3439	LAVABLE washable	lavabl	lavabl
C_2	1591	$\epsilon \rightarrow \epsilon$	1113	GAI joyful	gɛ	gɛ
		$\epsilon \rightarrow$ z	381	NIAIS stupid	njɛ	njɛz
		$\epsilon \rightarrow$ t	79	PR ready	pʁɛ	pʁɛt
		$\epsilon \rightarrow$ d	11	LAID ugly	lɛ	lɛd
		$\epsilon \rightarrow$ s	7	AIS thick	epɛ	epɛs
C_3	913	$\epsilon \rightarrow$ t	876	CONTENT happy	kɔ̃tɑ̃	kɔ̃tɑ̃t
		ɑ̃ $\rightarrow$ an	24	PERSAN persian	pɛʁsɑ̃	pɛʁsan
		$\epsilon \rightarrow \epsilon$	9	ARGENT silver	aʁʒɑ̃	aʁʒɑ̃
		$\epsilon \rightarrow$ d	4	GRAND large	gʁɑ̃	gʁɑ̃d
		$\epsilon \rightarrow$ s	0	—	—	—
⋮	⋮	⋮	⋮	⋮ ⋮	⋮	
C_{41}	1	k $\rightarrow$ ʃ	1	SEC dry	sɛk	sɛʃ
		$\epsilon \rightarrow \epsilon$	0	—	—	—

Comparison with an artificial dataset

	patterns	classes	entropy
[M.SG $\Rightarrow$ M.PL]	2	2	0.017
[M.SG $\Rightarrow$ F.SG]	26	41	0.528

- Now imagine a language K where [M.SG $\Rightarrow$ F.SG] for adjectives is as follows:

class	size	patterns	freqs	examples		
				lexeme	M.SG	M.PL
C	9225	a $\rightarrow$ u	8494	KALABA	kalaba	kalabu
		a $\rightarrow$ i	731	KOLOBA	koloba	kolobi

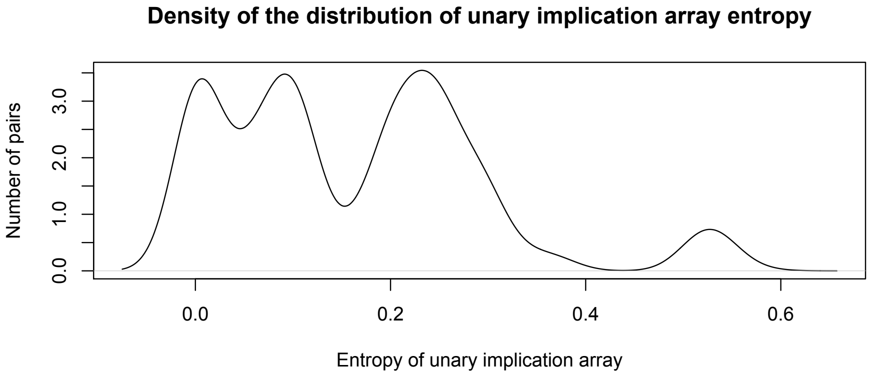
- Clearly K is very different from French. Yet:

language	array	patterns	classes	entropy
French	[M.SG $\Rightarrow$ F.SG]	26	41	0.528
K	[M.SG $\Rightarrow$ F.SG]	2	1	0.528

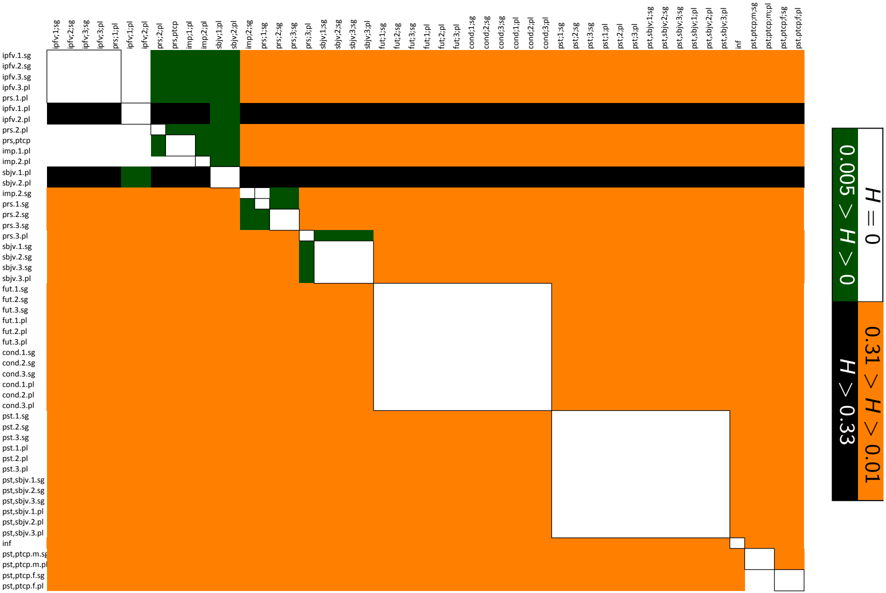
Structure

The big picture

- $51 \times 50 = 2550$ unary arrays
- Average entropy 0.1618
- Distribution of entropy values:



Basic classification



Alliances of forms

- We uncover 16 zones of perfect interpredictibility:

Finite forms

<table>
<tr><th>TEMPS</th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2">6</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td>8</td><td>9</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">10</td><td colspan="2">11</td><td>10</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">13</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">13</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

Nonfinite forms

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>F.SG</th><th>M.PL</th><th>F.PL</th></tr>
<tr><td>14</td><td>8</td><td colspan="2">15</td><td colspan="2">16</td></tr>
</table>

The effects of phonological neutralization

- The worst predictors of other cells are, by far: IPFV.1PL,IPFV.2PL,SBJV.1PL,SBJV.2PL
  - The entropy from one of those cells to any other cells is always above 0.33
  - The entropy from any other cell to any cell is always below 0.31
- This is entirely due to regular phonological processes
  - Homorganic vowel insertion between a branching onset and a glide
  - Simplification of geminate glides

IPFV.1PL		IPFV.1SG	lexeme	trans.
surface ϕ	underlying ϕ			
kadʁijɔ̃	kadʁjɔ̃	kadʁɛ	CADRER	‘frame’
kadʁijɔ̃	kadʁijjɔ̃	kadʁijɛ	QUADRILLER	‘cover’

- Important lesson: phonology has a strong impact on predictibility.

Another look

- If we focus on a set of distillations:

	PRS.1.SG	PRS.2.SG	PRS.1.PL	PRS.2.PL	PRS.3.PL	IPFV.1.PL	IMP.2.SG	IMP.2.PL	SBJV.1.SG	SBJV.1.PL	FUT.1.SG	PST.1.SG	INF	PRS.PTCP	PST.PTCP.M.SG	PST.PTCP.F.SG
PRS.1.SG	---	0,0011	0,2582	0,2558	0,234	0,2401	0,0008	0,2573	0,2447	0,2395	0,0839	0,2434	0,2786	0,2599	0,2166	0,2365
PRS.2.SG	0,0004	---	0,2681	0,2743	0,238	0,2764	0,0004	0,256	0,2462	0,2403	0,0849	0,2437	0,2896	0,2764	0,2164	0,2362
PRS.1.PL	0,2556	0,26	---	0,0012	0,055	0	0,2556	0,0016	0,0577	0,0026	0,2946	0,2495	0,3017	0,0004	0,2633	0,2585
PRS.2.PL	0,2545	0,2589	0	---	0,055	0	0,2545	0,0004	0,0577	0,0026	0,2902	0,2491	0,2974	0,0004	0,2598	0,2552
PRS.3.PL	0,207	0,207	0,0722	0,0734	---	0,0517	0,201	0,0734	0,0022	0,0529	0,2349	0,2998	0,3038	0,0722	0,2873	0,2851
IPFV.1.PL	0,5111	0,5181	0,3663	0,3672	0,3314	---	0,5111	0,3675	0,335	0,0042	0,544	0,5225	0,5825	0,3666	0,5374	0,5336
IMP.2.SG	0	0,0004	0,259	0,256	0,2443	0,2409	---	0,2519	0,2444	0,2404	0,0849	0,2437	0,2789	0,2607	0,2161	0,2359
IMP.2.PL	0,2549	0,2544	0	0	0,0546	0	0,2566	---	0,0597	0,0022	0,2839	0,2478	0,2955	0	0,2593	0,2546
SBJV.1.SG	0,2017	0,2017	0,0772	0,0785	0,0039	0,0568	0,2017	0,1216	---	0,0562	0,2364	0,3011	0,303	0,0773	0,2883	0,286
SBJV.1.PL	0,5095	0,5093	0,3652	0,3662	0,3316	0,0051	0,51	0,3677	0,3341	---	0,5357	0,5172	0,5697	0,3659	0,5235	0,5191
FUT.1.SG	0,0177	0,0177	0,2346	0,2254	0,1931	0,2142	0,0177	0,2299	0,1887	0,2059	---	0,2012	0,2056	0,2349	0,2039	0,2109
PST.1.SG	0,1067	0,1067	0,1066	0,0936	0,162	0,0968	0,106	0,0932	0,163	0,0909	0,1067	---	0,0612	0,1064	0,0476	0,0854
INF	0,0673	0,0684	0,0725	0,0732	0,1199	0,0847	0,0673	0,0713	0,1199	0,0805	0,0544	0,0152	---	0,072	0,0424	0,0711
PRS.PTCP	0,2553	0,2606	0	0,0012	0,0546	0	0,2553	0,0012	0,0578	0,0022	0,2938	0,2485	0,3021	---	0,2634	0,2586
PST.PTCP.M.SG	0,0913	0,0913	0,0801	0,078	0,1231	0,076	0,0902	0,0781	0,1249	0,0716	0,074	0,0228	0,0458	0,0799	---	0,1004
PST.PTCP.F.SG	0,0726	0,0726	0,047	0,042	0,0958	0,0449	0,0716	0,042	0,0964	0,0419	0,0637	0,0147	0,025	0,047	0	---

(darker is more unpredictable)

- Some unidirectional categorical implications
- Some cells are better predictors than others
- Variability in what is easy to predict.

Structure

The issues

- We now have a quantitative assessment of unary implication arrays for French verbs and adjectives, over a large lexicon.
- Question: would we get the same results if we change some parameters in the procedure?
- I will address the following variations:
  1. Taking into account more or less phonotactic similarity
  2. Focusing on written forms or phonetized forms
  3. Looking at a full lexicon or a set of exemplary lexemes (with type frequency information)
  4. Varying the size of the lexicon
- In each case we will conclude that variations result in sizable differences in the final analysis.

More or less phonotactic sensitivity?

- The previous analysis was based on a full decomposition of segments in distinctive features.
- This sometimes leads to the inference of quite subtle patterns, e.g. for [INF ⇒ PST.PTCP]:

  $$\text{ʁ} \rightarrow \text{y} \;/ [+\text{ant}, -\text{lat}][+\text{son}, +\text{cont}][+\text{cons}, -\text{cont}, -\text{nas}]\_\_\#$$

- What happens if we use a less subtle decomposition, or no decomposition at all?

C/V as the only phonotactic category

Full decomposition of segments

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2">6</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td>8</td><td>9</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">10</td><td colspan="2">11</td><td>10</td></tr>
<tr><td>FUT
COND</td><td colspan="6">12</td></tr>
<tr><td>PST
PST.SBJV</td><td colspan="6">13</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.
PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>14</td><td>8</td><td colspan="2">15</td><td colspan="2">16</td></tr>
</table>

Average conditional entropy: 0.1618

C/V distinction

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2">6</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td>8</td><td>9</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">10</td><td colspan="2">11</td><td>10</td></tr>
<tr><td>FUT</td><td colspan="6">12</td></tr>
<tr><td>COND</td><td colspan="3">12</td><td colspan="2">17</td><td>12</td></tr>
<tr><td>PST
PST.SBJV</td><td colspan="6">13</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.
PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>14</td><td>8</td><td colspan="2">15</td><td colspan="2">16</td></tr>
</table>

Average conditional entropy: 0.2268

- Sizeable rise of average entropy
- Loss of predictibility of the COND.12PL

lexeme	trans.	COND.2SG	COND.2PL
MOURIR	'die'	muʁʁɛ	muʁʁje
MOUDRE	'grind'	mudʁɛ	mudʁije

No phonotactic sensitivity

- We can also completely drop phonotactic sensitivity and assume that all changes can happen to all forms

Full decomposition of segments

	1SG	2SG	3SG	1PL	2PL	3PL
PRS	1	2		3	4	5
IPFV	3			6		3
IMP	—	7	—	8	9	—
PRS.SBJV	10			11		10
FUT	12					
COND						
PST	13					
PST.SBJV						

INF	PRS. PTCP	PST.PTCP M.SG	M.PL	F.SG	F.PL
14	8	15		16	

Average conditional entropy: 0.1618

No phonotactic sensitivity

	1SG	2SG	3SG	1PL	2PL	3PL
PRS	1	2		3	4	5
IPFV	18			6		18
IMP	—	7	—	8	9	—
PRS.SBJV	10			11		10
FUT	12					
COND				17		
PST	13					
PST.SBJV						

INF	PRS. PTCP	PST.PTCP M.SG	M.PL	F.SG	F.PL
14	19	15		16	

Average conditional entropy: 0.3463

- Steep rise of average entropy
- Now we are loosing track of the similarity between the PRS.1PL and the IPFV

Written vs. spoken medium

- We compare the phonetized data from `flexique` to the orthographic forms from the Le*fff* (Sagot et al., 2006)

☞ 4846 verbs in the intersection of the two lexica

Surface phonological forms (C/V)

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2">6</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td>8</td><td>9</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">10</td><td colspan="2">11</td><td>10</td></tr>
<tr><td>FUT</td><td colspan="6">12</td></tr>
<tr><td>COND</td><td colspan="3">12</td><td colspan="2">17</td><td>12</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">13</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.
PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>14</td><td>8</td><td colspan="2">15</td><td colspan="2">16</td></tr>
</table>

Average conditional entropy: 0.2268

Written forms (C/V)

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>IPFV</td><td colspan="6">4</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td colspan="2">4x</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="2">8</td><td>8x</td><td colspan="2">4y</td><td>8y</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">9</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">10</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.
PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>11</td><td>4</td><td colspan="4">12</td></tr>
</table>

Average conditional entropy: 0.1401

- Overall, entropy is higher with spoken forms
- However not always, e.g. between PRS.2SG and PRS.3SG

A paradoxical situation

- The written data give rise to an unusual situation.

lexeme	trans.	PRS.2PL	SBJV.2PL	IMP.2PL
voir	'see'	voyez	voyiez	voyez
payer	'pay'	payez	payiez	payez
avoir	'have'	avez	ayez	ayez
e	'be'	sommes	soyez	soyez

- In this table, any cell is fully interpredictable from the other except for [IMP.2PL $\Rightarrow$ SBJV.2PL]
  - The same situation arises with SBJV.1SG, SBJV.3SG and SBJV.3PL.
- Because of this, there is no single smallest partition of the paradigm in classes of fully mutually interpredictable cells.

☞ Full mutual interpredictibility is not a transitive relation.

? Is this an artifact of an odd orthographic systems, or do languages sometimes naturally have that property?

Full lexicon or exemplars+frequency

- The methods used here rely heavily on the phonotactic properties of the lexicon
  - ☞ Phonotactic conditions on the context of use of a pattern of alternation are inferred automatically through an exhaustive examination of the lexicon.
- Easier strategy: work from exemplars of all inflection classes and knowledge of the type frequency of each of these patterns
  - ☞ See e.g. Stump and Finkel (in press) on French conjugation

☞ What are the benefits of using a large scale lexicon?

- To investigate this, we emulated Stump and Finkel's dataset by creating a lexicon containing *n* copies of each exemplar lexeme, where *n* is the type frequency of the relevant inflection class.

Results with normal phonotactic sensitivity

Full lexicon, with phonotactic sensitivity

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2">6</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>7</td><td>—</td><td>8</td><td>9</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">10</td><td colspan="2">11</td><td>10</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">12</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">13</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS. PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>14</td><td>8</td><td colspan="2">15</td><td colspan="2">16</td></tr>
</table>

Average conditional entropy: 0.1618

Exemplars+frequency, with phonotactic sensitivity

<table>
<tr><th></th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>1</td><td colspan="2">2</td><td colspan="3">3</td></tr>
<tr><td>IPFV</td><td colspan="3">3</td><td colspan="2" rowspan="3">4</td><td>3</td></tr>
<tr><td>IMP</td><td>—</td><td>5</td><td>—</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td colspan="3">6</td><td>6</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">7</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">8</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS. PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>M.PL</th><th>F.SG</th><th>F.PL</th></tr>
<tr><td>7</td><td>3</td><td colspan="2">9</td><td colspan="2">10</td></tr>
</table>

Average conditional entropy: 0.018

- The results are too good to be true.
- Clearly this is due to a misuse of the Minimal Generalization strategy:
  - With only 72 actually distinct lexemes, the algorithm has very little to generalize over.
  - Thus patterns are unnaturally specific, so that situations of uncertainty barely ever arise.

Results with no phonotactic sensitivity

- If we discard phonotactic sensitivity, we get the same partition, but unexpectedly low average entropy:

Full lexicon, no phonotactic sensitivity

	1SG	2SG	3SG	1PL	2PL	3PL
PRS	1	2		3	4	5
IPFV	18			6		18
IMP	—	7	—	8	9	—
PRS.SBJV	10			11		10
FUT	12					
COND				17		
PST	13					
PST.SBJV						

INF	PRS. PTCP	PST.PTCP M.SG	PST.PTCP M.PL	PST.PTCP F.SG	PST.PTCP F.PL
14	19	15		16	

Average conditional entropy: 0.3463

Exemplars+frequency, no phonotactic sensitivity

	1SG	2SG	3SG	1PL	2PL	3PL
PRS	1	2		3	4	5
IPFV	18			6		18
IMP	—	7	—	8	9	—
PRS.SBJV	10			11		10
FUT	12					
COND				17		
PST	13					
PST.SBJV						

INF	PRS. PTCP	PST.PTCP M.SG	PST.PTCP M.PL	PST.PTCP F.SG	PST.PTCP F.PL
14	19	15		16	

Average conditional entropy: 0.1733

Results with no phonotactic sensitivity

- The skewed sampling happens to favor predictibility:

imperfective stem end	true lexicon	fake lexicon
complex onset	235	10
/j/	585	447
other	4504	5969

- Conclusion:
  - Using exemplars seems to be harmless if we are only interested in categorical implications (Stump and Finkel, in press).
  - However there is implicative structure based on the phonotactic shape of the lexicon that will unavoidably be missed.
  - Even putting this factor aside, sampling on the basis of overall inflection classes skews the distribution of patterns of alternation between individual pairs of cells.

What size of lexicon should we use?

- All the results discussed so far are dependent on the size of the lexicon we work from.
  - As the lexicon grows, the number of patterns it exemplifies grows.
  - Even once we reach the full set of patterns in the lexicon, adding in more lexemes will change the frequency distribution of the patterns
  - The token frequency of inflection classes may relate in complex ways to their type frequency: some highly populated classes may contain only rare lexemes, and vice-versa.
- Here we have worked with an unnaturally large lexicon
  - Probably larger than what any speaker masters
  - Definitely more than what a field linguist can hope to collect
- Thus we can conduct sampling experiments and try to assess the added value of collecting a large lexicon

The sampling strategy

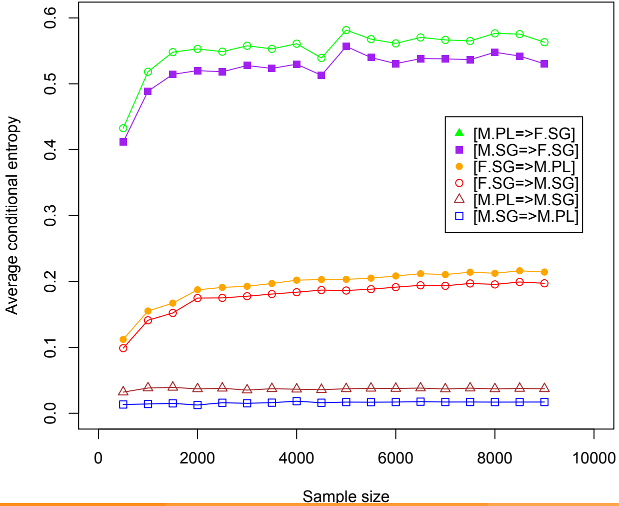
- We emulate the strategy of a learner (or a field linguist):
  - We pick a lexeme of the relevant category in a large corpus at random
  - We continue picking random lexemes until we have a list of n distinct lexemes
- Thus:
  - High frequency lexemes have a higher probability of being picked
  - There is still room for quite an amount of variation in what gets picked

☞ In practice, rather than actually sampling from a corpus, we rely on the token frequencies of lexemes collected in `Lexique 3`

- We then compare 50 distinct samples for each lexicon size
- We will focus on adjectives rather than verbs, because it is easier to work with small paradigms.

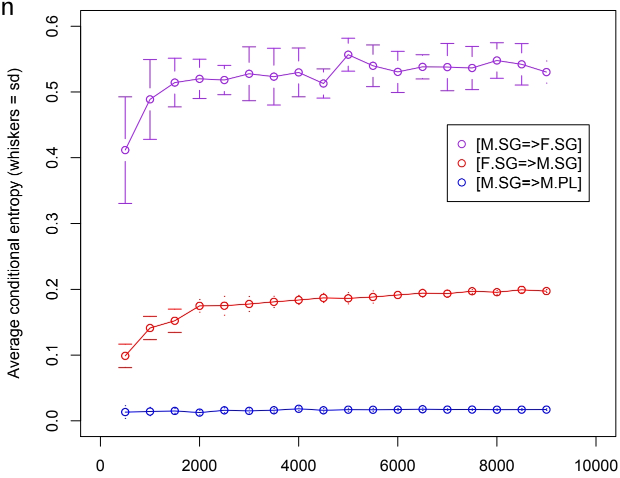
Rising average entropy

- Entropy almost always rises as the lexicon grows
- For some pairs of cells the rise is dramatic



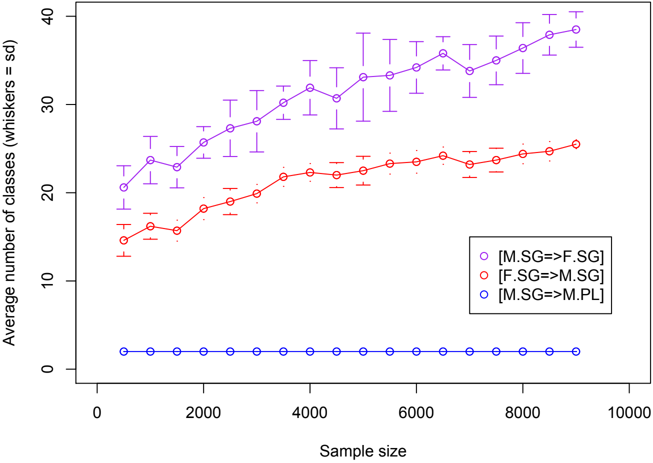
Diminishing variance

- At low sample sizes (< 1000) quite a lot of variance
- ☞ We should be careful when interpreting results drawn from a small lexicon



Rising number of classes

- The rise in entropy is due to a rise in the number of classes



Structure

Conclusions

- I have motivated a particular way of investigating the implicative structure of paradigms:
  - Study directly the frequency distribution of patterns of alternations (Ackerman et al., 2009)
  - Use of Albright's minimal generalization strategy for fast and easy inference of patterns
  - Application of fully automated analytic tools to semi-exhaustive, unanalyzed datasets
  - ☞ Systematicity of analytic choices
- Here we focused on unary implicative arrays, but the approach extends readily to the study of *n*-ary implicative arrays, and thus to the inference of principal parts.
- I have shown how details of execution influence the final results:
  - ☞ Constitution of the lexicon, pattern inference method, sampling method, sample
- In the evaluation of any study of implicative relations, each of these aspects should be examined critically.

Thanks

For useful input, references, discussion, collaboration, thanks to:

- Farrell Ackerman
- Jim Blevins
- Gilles Boyauthier Caron
- Fabiola Henri
- Ana R. LuClnt Plancq
- Greg Stump
- Delphine Tribout

References I

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.

Albright, A. C. (2002). *The Identification of Bases in Morphological Paradigms*. Ph.D. thesis, University of California, Los Angeles.

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

Baerman, M. and Corbett, G. G. (2012). 'Stem alternations and multiple exponence'. *Word Structure*, 5:52–68.

Blevins, J. P. (2006). 'Word-based morphology'. *Journal of Linguistics*, 42:531–573.

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.

——— (2007). 'Remarques sur les bases de la conjugaison'. In E. Delais-Roussarie and L. Labrune (eds.), *Des sons et des sens*. Paris: Hermès, 77–90. Ms, Université Paris 4 & Université Bordeaux 3.

Bonami, O., Boyé, G., Giraudo, H., and Voga, M. (2008). 'Quels verbes sont réguliers en français?' In *Actes du premier Congrès Mondial de Linguistique Française*. 1511–1523.

Bonami, O., Boyé, G., and Henri, F. (2011). 'Measuring inflectional complexity: French and Mauritian'. In *Workshop on Quantitative Measures in Morphology and Morphological Development*. San Diego.

Bonami, O., Caron, G., and Plancq, C. (in preparation). '`Flexique`: a large scale phonetized inflectional lexicon for French'. Laboratoire de Linguistique Formelle.

Bonami, O., Henri, F., and Luís, A. R. (2012). 'Tracing the origins of inflection in Creoles: a quantitative analysis'. Paper presented at the 9th Creolistics Workshop, Aarhus, Denmark.

Boyé, G. (2011). 'Régularité et classes flexionnelles dans la conjugaison du français'. In M. Roché, G. Boyé, N. Hathout, S. Lignon, and M. Plénat (eds.), *Des unités morphologiques au lexique*. Hermes Science, 41–68.

Boyé, G. and Cabredo Hofherr, P. (2006). 'The structure of allomorphy in spanish verbal inflection'. In *Cuadernos de Lingüística*, vol. 13. Instituto Universitario Ortega y Gasset, 9–24.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper and Row.

References II

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

——— (2009). 'Principal parts and degrees of paradigmatic transparency'. In J. P. Blevins and J. Blevins (eds.), *Analogy in Grammar*. Cambridge: Cambridge University Press, 13–54.

Malouf, R. and Ackerman, F. (2010). 'Paradigms: The low entropy conjecture'. Paper presented at the Workshop on Morphology and Formal Grammar, Paris.

Montermini, F. and Bonami, O. (to appear). 'Stem spaces and predictibility in verbal inflection'. *Lingue e Linguaggio*.

Montermini, F. and Boyé, G. (2012). 'Stem relations and inflection class assignment in Italian'. *Word Structure*, 5:69–87.

Morin, Y.-C. (1987). 'Remarques sur l'organisation de la flexion en français'. *ITL Review of Applied Linguistics*, 77–78:13–91.

New, B., Brysbaert, M., Veronis, J., and Pallier, C. (2007). 'The use of film subtitles to estimate word frequencies'. *Applied Psycholinguistics*, 28:661–677.

Sagot, B., Clément, L., La Clergerie, E. d., and Boullier, P. (2006). 'The le*fff* 2 syntactic lexicon for French: architecture, acquisition, use'. In *Proceedings of LREC 2006*.

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.

Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie-Verlag. Translated as (Wurzel, 1989).

——— (1989). *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.

Stem spaces

- Family of analyses of Romance conjugation by Boyd colleagues
  - (Bonami and Boyé, 2002; Boyé and Cabredo Hofherr, 2006; Bonami and Boyé, 2007; Bonami et al., 2008; Boyé, 2011; Montermini and Boyé, 2012; Montermini and Bonami, to appear)
- Ultimately grounded in (Aronoff, 1994)'s view of stem allomorphs and (Morin, 1987)'s view of implicative relations
- Uniform methodology:
  - Abstract away lexeme-specific suppletive forms
  - Abstract away constant inflection
  - Identify alliances of forms
    - The resulting distillation is a stem space
  - Identify reliable implicative relations within the stem space, under the following assumptions:
    - The number of links between stems should be minimized
    - Implicative relations between two cells rely on a single default strategy

Comparing the partitions

Stem space based partition

Finite forms

<table>
<tr><th>TEMPS</th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td colspan="3">3</td><td colspan="2"></td><td>2</td></tr>
<tr><td>IPFV</td><td colspan="6">1</td></tr>
<tr><td>IMP</td><td>—</td><td>5</td><td>—</td><td colspan="2">6</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td>7</td><td>7</td><td>7</td><td colspan="2">8</td><td>7</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">10</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">11</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

Nonfinite forms

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>F.SG</th><th>M.PL</th><th>F.PL</th></tr>
<tr><td>9</td><td>4</td><td colspan="4">12</td></tr>
</table>

Entropy-based partition

Finite forms

<table>
<tr><th>TEMPS</th><th>1SG</th><th>2SG</th><th>3SG</th><th>1PL</th><th>2PL</th><th>3PL</th></tr>
<tr><td>PRS</td><td>3A</td><td colspan="2">3B</td><td>1B</td><td>1C</td><td>2</td></tr>
<tr><td>IPFV</td><td colspan="3">1B</td><td colspan="2">1A</td><td>1B</td></tr>
<tr><td>IMP</td><td>—</td><td>5</td><td>—</td><td>6A</td><td>6B</td><td>—</td></tr>
<tr><td>PRS.SBJV</td><td>7</td><td>7</td><td>7</td><td colspan="2">8</td><td>7</td></tr>
<tr><td>FUT</td><td colspan="6" rowspan="2">10</td></tr>
<tr><td>COND</td></tr>
<tr><td>PST</td><td colspan="6" rowspan="2">11</td></tr>
<tr><td>PST.SBJV</td></tr>
</table>

Nonfinite forms

<table>
<tr><th rowspan="2">INF</th><th rowspan="2">PRS.PTCP</th><th colspan="4">PST.PTCP</th></tr>
<tr><th>M.SG</th><th>F.SG</th><th>M.PL</th><th>F.PL</th></tr>
<tr><td>9</td><td>4</td><td colspan="2">12A</td><td colspan="2">12B</td></tr>
</table>

Discussion

- The simpler partition of (Bonami and Boyé, 2002) is entirely due to:
  - Leaving out data (so-called suppletive inflected forms)
  - Abstracting away regular phonological processes
- Both moves are valid (though disputable) within the construction of a constructive formal analysis
- Neither is justified by direct empirical evidence
- Ultimately, the drive towards segmentation (i.e. reducing implicative structure to morphotactics) was responsible for these analytic choices. In retrospect it is not clear that they are motivated.

Principal part analyses

- (Finkel and Stump, 2007, 2009; Stump and Finkel, in press) explore a research program that shares much of our goals.
- Important differences:
  - Focus on categorical implications, hence a subset of what we studied.
  - Focus on principal parts
    - Principal part systems are very sensitive to the exact lexicon they are built on, whereas speakers are exposed to varied lexica.
    - There are often multiple optimal principal part systems.
    - This is not a problem for pedagogy, but calls into question the usefulness of principal parts as descriptive devices.
  - Uses segmented inputs
    - ☞ Often improves the predictive power of a cell
  - Uses exemplars rather than full paradigms
    - No sensitivity to the phonological structure of stems
    - ☞ Often reduces the predictive power of a cell