Information-theoretic measures of inflectional complexity: Empirical challenges and analytic rewards

Olivier Bonami

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

Partially based on joint work with Gilles Boyé & Fabiola Henri

Introduction

- Domain of inquiry: implicative structure of inflection systems.
  - Speakers have knowledge of implicative relations between paradigm cells that allow them to make (more or less reliable) inferences on the inflection of novel lexemes (e.g. Morin, 1987; Wurzel, 1989).
  - The organization and reliability of these implicative relations are one of the determinants of the complexity of an inflection system (Ackerman et al., 2009; Malouf and Ackerman, 2010)
  - Many previous studies of implicative relations (e.g. Bonami and Boyé, 2002, 2007; Boyé, 2011) are wanting because of:
    1. existence of many alternative analyses within a single set of hypotheses
    2. nonquantitative take on regularity leading to untestable claims
    3. fitting of analyses to a particular system, leading to poor cross-linguistic extensibility
    4. lack of automation: poor reproducibility, exploration of alternatives tedious.
- Recent work (e.g. Finkel and Stump, 2007, 2009; Ackerman et al., 2009; Malouf and Ackerman, 2010) paves the way to a more direct approach.

Introduction

- Goals of this paper:
  - Motivate a refined version of the strategy of (Ackerman et al., 2009)
    - ☞ Conditional entropy of a pattern given properties of an input form
  - Apply it to a realistic dataset
    - ☞ 6440 French verb lexemes from BDLEX
  - Compare the results with what we learned from the previous generation of studies
    - ☞ Rediscovering (Bonami and Boyé, 2002)'s stem spaces

Structure

A simple dataset

- Suppose we have been exposed to exactly the following 10 pairs of singular and plural French nouns:

lexeme	SG	PL
CHEVAL 'horse'	ʃəval	ʃəvo
JOURNAL 'newpaper'	ʒuʁnal	ʒuʁno
BAL 'dance'	bal	bal
FESTIVAL 'festival'	fɛstival	fɛstival
OEIL 'eye'	œj	jø
SEUIL 'doorstep'	sœj	sœj
TABLE 'table'	tabl	tabl
CHAISE 'chair'	ʃɛz	ʃɛz
HOMME 'man'	ɔm	ɔm
FEMME 'woman'	fam	fam

- How can this knowledge guide us in inferring the form of the plural for new singular nouns?

Frequency of patterns

- If we agree on a deterministic strategy for inferring patterns of alternation, we have information on the frequency of each pattern:

pattern	example	#
$X \rightsquigarrow X$	HOMME	7
Xœj $\rightsquigarrow$ Xjø	ŒIL	1
Xal $\rightsquigarrow$ Xo	CHEVAL	2

- We can partition the singulars according to which set of patterns they could be input to, and observe the frequency of the partition cells.

form of SG	#	applicable patterns
Xal	4	$X \rightsquigarrow X$, Xal $\rightsquigarrow$ Xo
Xœj	2	$X \rightsquigarrow X$, Xœj $\rightsquigarrow$ Xjø
other	4	$X \rightsquigarrow X$

The distribution of conditional probabilities

- From this information we can deduce, for each class of singulars, the relative frequency of each applicable pattern in that class.

form of SG	#	pattern	relative frequency
*X*al	4	$X \rightsquigarrow X$	.5
		*X*al $\rightsquigarrow$ *X*o	.5
*X*œj	2	$X \rightsquigarrow X$	.5
		*X*œj $\rightsquigarrow$ *X*jø	.5
other	4	$X \rightsquigarrow X$	1

- If the sample of pairs is representative of the general situation, this can be used to estimate the distribution of conditional probabililities of a pattern given a singular.
- Arguably this is a good description of the form-based implicative relations in the paradigm: tells us how reliably we can guess the plural knowing the singular.

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.

$$\begin{aligned} H(\text{pattern} \mid \text{SG}) = & -\left(\frac{4}{10} (\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}) + \frac{2}{10} (\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}) + \frac{4}{10} \log_2 1 \right) \\ = & -\left(\frac{4}{10} \log_2 \frac{1}{2} + \frac{2}{10} \log_2 \frac{1}{2} + \frac{4}{10} \log_2 1 \right) \\ = & -\left(\frac{4}{10} \times -2 + \frac{2}{10} \times -2 + \frac{4}{10} \times 0 \right) \\ = & \; 0.6 \end{aligned}$$

More realistic data (lexique.org)

form of SG	#	pattern	relative frequency
*X*al	181	$X \rightsquigarrow X$	.35
		*X*al $\rightsquigarrow$ *X*o	.65
*X*œj	30	$X \rightsquigarrow X$	.96
		*X*œj $\rightsquigarrow$ *X*jø	.04
*X*man	103	$X \rightsquigarrow X$	.86
		*X*man $\rightsquigarrow$ *X*mɛn	.16
...	...	...	...
other	15910	$X \rightsquigarrow X$	1

$$H(\text{pat.} \mid \text{SG}) = 0,084$$

- Thus:
  - Very reliable implicative relation from SG to PL of French nouns.
  - A property of this system not captured by the entropy measure: pockets of uncertainty are easily identified as such.
  - This fits nicely with inflection errors of speakers.

Ackerman et al.'s (2009) claims

- Claims from (Ackerman et al., 2009; Malouf and Ackerman, 2010):
  1. Knowledge of implicative relations from cell *A* to cell *B* reduces to knowledge of the distribution of conditional probabilities of patterns from *A* to *B* given the form of *A*.
  2. Conditional entropy then gives a good overall measure of the complexity of inferring *B* from *A*.
  3. We can use conditional entropy to evaluate the strength of implicative relations between different pairs of cells within a system
     - ☞ Limiting case: in a system with segregated principal parts (Finkel and Stump, 2007), null entropy from the principal parts to the predicted cells.
  4. The complexity of different systems can be compared by comparing mean conditional entropy across all pairs of cells in each system.
- This defines a promising research program.
- However methodological problems with Ackerman *et al.*'s pilot studies mask the interest of the approach.

Structure

Four methodological issues

- Main claim: Ackerman *et al.*'s strategy makes sense only if we start from:
  - Reasonably good statistical knowledge on the inflectional system
  - Morphologically unbiased representations
  - Phonologically unbiased representations
- In addition, we need an unbiased way of inferring patterns from cell A to cell B.

Type frequency information is crucial

- In the absence of type frequency information, the only meaningful contrast is between null and non-null entropy.

- Assume an inflection system with
  - 2 paradigm cells
  - 2 exponents for cell A
  - 4 exponents for cell B
  - A strong preference of one exponent in cell B

IC	A	B	type freq.
1	-i	-a	497
2	-i	-e	1
3	-i	-u	1
4	-i	-y	1
5	-o	-a	497
6	-o	-e	1
7	-o	-u	1
8	-o	-y	1

- Results:

	A	B
A	—	2
B	1	—

$H(\text{col}|\text{row})$, without frequency

	A	B
A	—	0.0624
B	1	—

$H(\text{col}|\text{row})$, with frequency

No segmentation

- Example: two types of French infinitives ending in -iʁ, segmented according to the traditional classification.

IC	INF	IPFV.3SG	lexeme	trans.
1	sɔʁt-iʁ	sɔʁtɛ	sortir	go out
2	amɔʁti-ʁ	amɔʁtisɛ	amortir	cushion

- If we know whether iʁ is a single morph, we know what the IPFV will be.
- However a speaker inflecting a novel lexeme can't know that.
  - ☞ Speakers can't hear morph boundaries.
  - ☞ The form of the IPFV is precisely what motivates the segmentation of the infinitive.
- Any entropy figure computed from a grammatical description in terms of exponence, rather than actual forms, tells us something on the description but little on the data.

No phonological abstraction

- As morphologists, we are used to abstracting away automatic phonology.
- This makes implications more reliable by suppressing neutralizations.

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

- However this is unadvisable here:
  - For the speaker, the uncertainty stemming from phonology is just as important as the uncertainty stemming from morphology.
  - The fact that the inflection system tolerates such a situation contributes to its complexity.
  - If we want to make comparisons across languages, no easy way to check whether we made similar abstractions in all languages.

Choosing the right patterns

- The results depend heavily on what method is used to classify the patterns of alternation between forms.
- Practical case:
  - When examining the paradigm of OEIL, do we infer the pattern Xœj $\rightsquigarrow$ Xjø or the pattern #œj $\rightsquigarrow$ jø?
  - Previously we chose Xœj $\rightsquigarrow$ Xjø, and found that $H(\text{pattern} \mid \text{SG}) = 0.6$.
  - Had we made the other choice, we would have found that $H(\text{pattern} \mid \text{SG}) = 0.4$.
- A general, cross-linguistically valid method for inferring patterns is not forthcoming.
- Therefore we should choose a method that:
  - is sophisticated enough to deal with the morphology of the languages at hand;
  - is simple enough that linguists can criticize it and evaluate whether it biases results.

Structure

Practical matters

- The dataset:
  - 6440 nondefective verbal lexemes from BDLEX (de Calmès and Pérennou, 1998)
  - Hand correction of:
    - random errors
    - unwarranted phonological abstractions
- The program:
  - Simple (600 lines) Python 3 script
  - Efficient enough: mean run over 6440 pairs of cells my laptop is 18.2*s* (to be repeated $48 \times 47 = 2256$ times)
  - Verbose mode detailing all intermediate classification in a linguist-friendly format

Strategy for inferring patterns

- We borrow the strategy of the Minimal Generalization Learner (Albright, 2002).
  - Assume a decomposition of segments into distinctive features.
  - Assumes that forms are related by SPE-style rules (Chomsky and Halle, 1968).
  - For each ⟨INPUT, OUTPUT⟩ 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

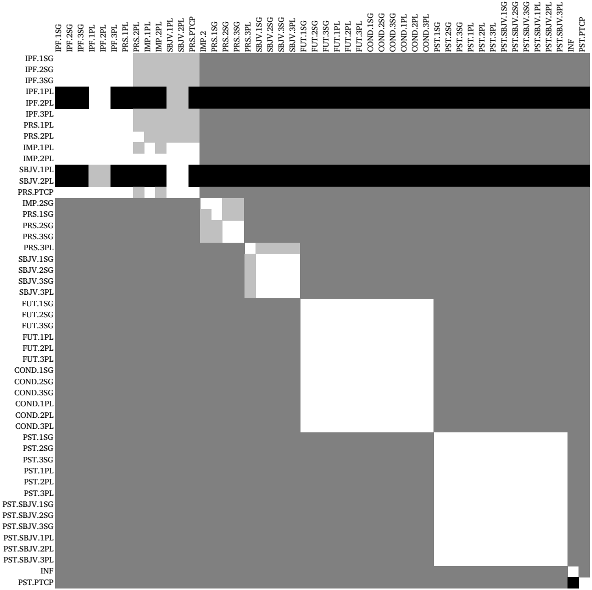
    $$r = A \rightarrow B/(\#|X)[\text{feat}^{+}]^{*}\text{seg}^{*}\_\_\_\text{seg}^{*}[\text{feat}^{+}]^{*}(Y|\#)$$

    such that all rules in R are specializations of r.
- Unlike the MGL's, this is a tractable computation: for n structural changes, $n-1$ rule comparisons in the worst case.

Sample run

```
*****************************************
        ipf.4 ==> ipf.1
*****************************************
Inferring rules...
jô-->E/X[p,t,k,b,d,g,f,s,S,v,z,Z,m,n,J,j,l,r,w,H,i,y,E,6,u,o,ê,û,ô]___#
ô-->E/X[J,j]___#
ijô-->E/X[p,t,k,b,d,g,f,s,S,v,z,Z][l,r]___#
-----
class 1 ( ElwaJô ~ ElwaJE ):  228 members
ô -> E / X[J,j] ___ # : 228 (éloigner, etc.)
local conditional entropy: -0.0
-----
class 2 ( pwavrijô ~ pwavrE ):  319 members
jô -> E / X[p,t,k,b,d,g,f,s,S,v,z,Z,m,n,J,j,l,r,w,H,i,y,E,6,u,o,ê,û,ô] ___ # : 0
ô -> E / X[J,j] ___ # : 30 (vriller, etc.)
ijô -> E / X[p,t,k,b,d,g,f,s,S,v,z,Z][l,r] ___ # : 289 (poivrer, etc.)
local conditional entropy: 0.44982565416940956
-----
class 3 ( anordisjô ~ anordisE ):  5893 members
jô -> E / X[p,t,k,b,d,g,f,s,S,v,z,Z,m,n,J,j,l,r,w,H,i,y,E,6,u,o,ê,û,ô] ___ # : 5429
ô -> E / X[J,j] ___ # : 464 (orthographier, etc.)
local conditional entropy: 0.3977149431671207
conditional entropy of ipf.1 given ipf.4: 0,3862156123856960
```

Global results



black: $H > 0.3$
dark gray: $0.3 > H > 0.01$
light gray: $0.004 > H > 0$
white: $H = 0$

Zooming in: extreme values

	IPFV.1SG	IPFV.2SG	IPFV.3SG	IPFV.1PL	IPFV.2PL	IPFV.3PL	PRS.1PL	PRS.2PL	IMP.1PL	IMP.2PL	SBJV.1PL	SBJV.2PL
IPFV.1SG	—	0	0	0	0	0	0	0.0012	0.0003	0.0015	0.0003	0.0003
IPFV.2SG	0	—	0	0	0	0	0	0.0012	0.0003	0.0015	0.0003	0.0003
IPFV.3SG	0	0	—	0	0	0	0	0.0012	0.0003	0.0015	0.0003	0.0003
IPFV.1PL	0.3862	0.3862	0.3862	—	0	0.3862	0.3862	0.3870	0.3865	0.3873	0.0003	0.0003
IPFV.2PL	0.3862	0.3862	0.3862	0	—	0.3862	0.3862	0.3870	0.3865	0.3873	0.0003	0.0003
IPFV.3PL	0	0	0	0	0	—	0	0.0012	0.0003	0.0015	0.0003	0.0003
PRS.1PL	0	0	0	0	0	0	—	0.0012	0.0003	0.0015	0.0003	0.0003
PRS.2PL	0	0	0	0	0	0	0	—	0.0003	0.0003	0.0003	0.0003
IMP.1PL	0	0	0	0	0	0	0	0.0012	—	0.0012	0	0
IMP.2PL	0	0	0	0	0	0	0	0	0	—	0	0
SBJV.1PL	0.3871	0.3871	0.3871	0.0011	0.0011	0.3871	0.3871	0.3881	0.3877	0.3887	—	0
SBJV.2PL	0.3871	0.3871	0.3871	0.0011	0.0011	0.3871	0.3871	0.3881	0.3877	0.3887	0	—

Zooming in: central values

	IPFV.3SG	PRS.3PL	PRS.3SG	PRS.PTCP	IMP.2SG	IMP.2PL	SBJV.3SG	SBJV.2PL	INF	FUT.3SG	PST.3SG	PST.PTCP
IPFV.3SG	—	0.0532	0.2406	0.0003	0.2318	0.0015	0.0582	0.0003	0.2872	0.2838	0.2433	0.2527
PRS.3PL	0.0697	—	0.1837	0.0555	0.1837	0.0709	0.0029	0.0662	0.2883	0.2215	0.2822	0.2679
PRS.3SG	0.2493	0.2086	—	0.2372	0.0003	0.2235	0.2090	0.2330	0.2498	0.0682	0.2095	0.2182
PRS.PTCP	0.0000	0.0583	0.2351	—	0.2339	0.0012	0.0601	0.0000	0.2807	0.2672	0.2419	0.2502
IMP.2SG	0.2295	0.2086	0.0003	0.2294	—	0.2236	0.2091	0.2331	0.2498	0.0682	0.2094	0.2181
IMP.2PL	0.0000	0.0529	0.2305	0.0000	0.2334	—	0.0606	0.0000	0.2787	0.2788	0.2405	0.2487
SBJV.3SG	0.0743	0.0034	0.1823	0.0743	0.1823	0.0761	—	0.0697	0.2881	0.2232	0.2821	0.2676
SBJV.2PL	0.3871	0.3774	0.5099	0.3871	0.5103	0.3887	0.3526	—	0.5833	0.5405	0.5407	0.5580
INF	0.0653	0.1159	0.0476	0.0630	0.0476	0.0613	0.0938	0.0686	—	0.0372	0.0129	0.0307
FUT.3SG	0.2161	0.1803	0.0179	0.1985	0.0179	0.2055	0.1757	0.2053	0.1861	—	0.1875	0.1696
PST.3SG	0.0908	0.1494	0.0762	0.0907	0.0760	0.0854	0.1276	0.0750	0.0453	0.0813	—	0.0413
PST.PTCP	0.0564	0.1146	0.0663	0.0677	0.0660	0.0601	0.0927	0.0572	0.3428	0.0484	0.0213	—

Motivating the stem space

- (Bonami and Boyé, 2002) develops an analysis of French conjugation heavily influenced by (Aronoff, 1994, chapter 2)
  - ☞ Similar in spirit to Brown (1998), Pirelli and Battista (2000), (Stump, 2001, chapter 6), etc.
- Design features
  - Each lexeme comes equipped with a collection of indexed stems.
    - Crucially the same stem may occur multiple times with different indices.
  - Stem indices are related by default implicative relations.
    - Often this relation is one of identity.
  - Irregular conjugation mostly consists of violating these defaults
  - Assumption that there are no affixal inflection classes in Romance conjugation: any variation in affixal exponence is a lexeme-specific exception.

Inferring the stem space

- For each paradigm cell, find the longest suffix that is common to all lexemes (except possibly a handful of deep irregulars)
  - ☞ In many instances this suffix is empty.
- For each lexeme, substract from each cell its suffix to find the corresponding stem.
- Determining stem indices:
  - Two cells α and β select the same stem index iff α's stem is identical to β's stem for all lexemes.
- This results in a partition of the paradigm.
- The partition is intended to capture perfect interpredictibility between cells.

The resulting 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" rowspan="2">1</td><td>2</td></tr>
<tr><td>IPFV</td><td colspan="3"></td><td></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>

Comparison 1: bidirectional null entropy

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>1C</td><td>1D</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="4">12</td></tr>
</table>

Comparison 2: monodirectional null entropy

Stem space based partition

Finite forms

TEMPS	1SG	2SG	3SG	1PL	2PL	3PL
PRS	3			1		2
IPFV	1					
IMP	—	5	—	6		—
PRS.SBJV	7	7	7	8		7
FUT / COND	10					
PST / PST.SBJV	11					

Nonfinite forms

INF	PRS.PTCP	PST.PTCP M.SG	PST.PTCP F.SG	PST.PTCP M.PL	PST.PTCP F.PL
9	4	12			

Entropy-based partition

Finite forms

TEMPS	1SG	2SG	3SG	1PL	2PL	3PL
PRS	3			1B	1C	2
IPFV	1B			1A		1B
IMP	—	5	—	6		—
PRS.SBJV	7	7	7	8		7
FUT / COND	10					
PST / PST.SBJV	11					

Nonfinite forms

INF	PRS.PTCP	PST.PTCP M.SG	PST.PTCP F.SG	PST.PTCP M.PL	PST.PTCP F.PL
9	4	12			

Comparison 3: monodirectional low entropy

- With a threshold at 0.004 bits:

Stem space based partition

Finite forms

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

Nonfinite forms

INF	PRS.PTCP	PST.PTCP M.SG	PST.PTCP F.SG	PST.PTCP M.PL	PST.PTCP F.PL
9	4	12			

Entropy-based partition

Finite forms

TEMPS	1SG	2SG	3SG	1PL	2PL	3PL
PRS	C					2
IPFV	A					
IMP	—	C	—			—
PRS.SBJV	7					7
FUT	10					
COND						
PST	11					
PST.SBJV						

Nonfinite forms

INF	PRS.PTCP	PST.PTCP M.SG	PST.PTCP F.SG	PST.PTCP M.PL	PST.PTCP F.PL
9	A	12			

Discussion

- In (Bonami and Boyé, 2002)'s analysis:
  - Exceptional forms such as DIRE 'say' PRS.2PL *dites* or ÊTRE 'be' PRS.1PL *sommes* are dismissed as suppletive inflected forms, because they are not segmentable.
  - Exceptional forms such as SAVOIR 'know' PRS.PTCP *sachant* are segmented because the suffix is regular.
  - Likewise, forms such as ÊTRE 'be' PRS.1SG *suis* are assimilated are dismissed as suppletive inflected forms despite the absence of irregular suffixal exponence, because they are isolated in the paradigm.
  - On the other hand, forms such as AVOIR 'have' IMP.2SG *aie* are assumed to participate in normal inflection because their stem occurs elsewhere in the paradigm.
- All these phenomena are of the same size (less than 12 lexemes involved) and thus lead to similarly low entropy.

Conclusions

- The distribution of conditional entropy by and large captures the properties encoded by Bonami & Boyé's stem space.
- Detailed examination of the results highlights disputable modelling choices:
  - Bonami & Boyé ignore the role of phonological neutralizations, but these contribute more unpredictability to the sytem than anything else.
  - Distinctions between low entropy zones due to:
    - Insistence on full segmentation
    - Stem-based implications only
- The method used here:
  - Definitely does not lead to compact descriptions of inflectional systems, but
  - Helps us discover previously unknown properties of inflectional systems.
  - Definitely can not serve as an overall measure of complexity, but
  - Goes a long way towards quantifying the complexity of implicative relations.

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.

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.

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*. Paris: Hermès, 41–68. Ms., U. de Bordeaux.

Brown, D. (1998). 'Stem Indexing and morphonological selection in the russian verb: a network morphology account'. In R. Fabri, A. Ortmann, and T. Parodi (eds.), *Models of Inflection*. Niemeyer, 196–224.

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

de Calmès, M. and Pérennou, G. (1998). 'BDLEX : a lexicon for spoken and written french'. In *Proceedings of the First International Conference on Language Resources and Evaluation*. Granada: ERLA, 1129–1136.

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, 13–53.

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

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

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

Stump, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Wurzel, W. U. (1989). *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.