Discovering Implicative Morphology

Olivier Bonami

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

Huitièmes Décembrettes

Structure

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 highy 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
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 underlie the idea of principal parts: sets of *n* cells from which a categorical implication exists to all other cells.

Implicative structure and morphotactic structure

- Paradigms have implicative structure
- Words have morphotactic structure
- Both structures are established through paradigmatic opposition: comparing words/paradigms to other words/paradigms
- Central theoretical debate in morphology: can the implicative structure of paradigms be deduced from the morphotactic structure of words?
- The Bloomfieldian answer: in can, and it should.
- The Word and Paradigm answer (from Matthews, 1965, on): it can't always.
  - Parasitic formations (a.k.a. 'morphomic stems' Aronoff, 1994)
  - Syncretism (e.g. Stump, 2001; Baerman et al., 2005)
- The radical WP approach (Blevins, 2006; Ackerman et al., 2009): even when it can, it shouldn't.

Implicative stucture as an empirical property

- This is an interesting theoretical debate, but I won't say anything about it.

☞ We don't know nearly enough on implicative structure to take an informed decision.

- Very few large scale empirical studies of implicative structures.
- Two notable exceptions:
  - Studies of Romance conjugation by Boyé and colleagues
    - (Bonami and Boyé, 2002; Boyé and Cabredo Hofherr, 2006; Bonami and Boyé, 2007; Bonami et al., 2008; Boyé, 2011; Montermini and Boyé, 2012)
    - Ultimately grounded in (Aronoff, 1994)'s view of stem allomorphs and (Morin, 1987)'s view of implicative relations
  - Studies of principal part systems by Finkel & Stump
    - (Finkel and Stump, 2007, 2009; Stump and Finkel, in press)
    - Focus on categorical implicative relations

Today's plan

- Research program laid out in (Ackerman et al., 2009):
  - Use of information-theoretic tools to model implicative structure
  - Further applied and developed in (Sims, 2010; Malouf and Ackerman, 2010; Bonami et al., 2011)
- We will use a (revision of) Ackerman's approach as a way of exploring implicative structure.
- The particular approach here is:
  - Unashamedly quantitative: type frequency is crucial.
  - Unashamedly symbolic: we are writing descriptions, not modelling what happens in the brain
  - Fully implemented (with help from Gilles Boyé and Delphine Tribout)
  - Applied to real-size datasets (thousands of lexemes)

☞ This talk is about instrumented descriptive morphology, not theoretical morphology or psycholinguistics.

☞ We try to discover implicative morphology not to justify or model it.

The dataset

- Based on flexique, a new inflectional lexicon of French (Bonami et al., in preparation)

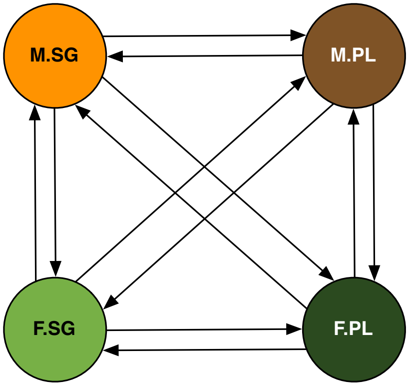
POS	lexemes	words
nouns	33,716	67,353
adjectives	11,420	45,680
verbs	5,325	271,575
total	50,461	384,608

- Design:
  - Based on Lexique 3 (New et al., 2007)
  - Hand-correction of phonemic transcriptions for principal parts
  - Automatic generation of predictable forms
  - Selective semi-automatic validation
- Limitations:
  - Limited support for phonetic alternations
  - Currently no support for overabundance
- Will be available within a few weeks; distributed as a free ressource

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 $\Rightarrow$ 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	$p_1 : X\text{al} \sim X\text{o}$	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 likelyhood 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.
  - Assumes 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 $\sim X$o	399	LOYAL	lwajal	lwajo
		$X \sim X$	29	BANAL	banal	banal
C_2	8797	$X \sim X$	8797	CALME	kalm	kalm

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

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

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

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

Using conditional entropy as a summary of the distribution

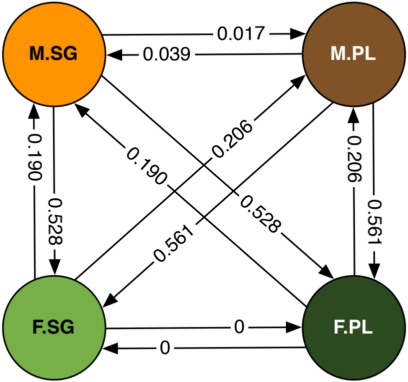
$$H(Y \mid X) = -\sum_{x \in X} P(x) \left(\sum_{y \in Y} P(y \mid x) \log_2 P(y \mid x) \right)$$

- 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:

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

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.
- The best overall predictor is the feminine (Durand, 1936)

Important caveats

1. 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.
2. All calculations are dependent on the way we classify data
   - There might more fine-grained ways of examining patterns
   - Other factors (e.g. morphosyntactic, semantic) might come into play
   - ☞ Our entropy values should be seen as upper bounds
3. We are just classifying a dataset
   - This probably corresponds to knowledge speakers use
   - However the exact shape and size of the lexicon varies considerably
   - We don't know how much information exactly speakers memorize

Illustrating caveat 1: [M.SG ⇒ F.SG]

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

Pattern		freq.
$\epsilon \to \epsilon$ /	__#	6153
$\epsilon \to$ ʁ /	{ʃ,ʒ,j}{e,ɛ}__#	110
$\epsilon \to$ t /	[+son,−lat]__#	1178
$\epsilon \to$ z /	[+voc,−cons,−nas]__#	506
$\epsilon \to$ d /	[−cons,−high]__#	133
$\epsilon \to$ s /	__#	22
$\epsilon \to$ 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
œ̃ → yn /	[+voice][+cons,−high]__#	7

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

Illustrating caveat 1: [M.SG ⇒ F.SG]

class	size	patterns	frees	examples		
				lexeme	M.SG	M.PL
C_1	3439	$\epsilon \to \epsilon$	3439	LAVABLE 'washable'	lavabl	lavabl
C_2	1591	$\epsilon \to \epsilon$	1113	GAI 'joyful'	gɛ	gɛ
		$\epsilon \to$ z	381	NIAIS 'stupid'	njɛ	njɛz
		$\epsilon \to$ t	79	PRÊT 'ready'	pʁɛ	pʁɛt
		$\epsilon \to$ d	11	LAID 'ugly'	lɛ	lɛd
		$\epsilon \to$ s	7	ÉPAIS 'thick'	epɛ	epɛs
C_3	913	$\epsilon \to$ t	876	CONTENT 'happy'	kɔ̃tɑ̃	kɔ̃tɑ̃t
		ɑ̃ → an	24	PERSAN 'persian'	pɛʁsɑ̃	pɛʁsan
		$\epsilon \to \epsilon$	9	ARGENT 'silver'	aʁʒɑ̃	aʁʒɑ̃
		$\epsilon \to$ d	4	GRAND 'large'	gʁɑ̃	gʁɑ̃d
		$\epsilon \to$ s	0	—	—	—
⋮	⋮	⋮	⋮	⋮ ⋮	⋮	
C_{41}	1	k → ʃ	1	SEC 'dry'	sɛk	sɛʃ
		$\epsilon \to \epsilon$	0	—	—	—

Illustrating caveat 1: an artificial dataset

	patterns	classes	entropy
[M.SG ⇒ M.PL]	2	2	0.017
[M.SG ⇒ F.SG]	26	41	0.528

- Now imagine a language *K* where [M.SG ⇒ F.SG] for adjectives is as follows:

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

- Clearly *K* is very different from French. Yet:

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

Illustrating caveat 2: the role of gender

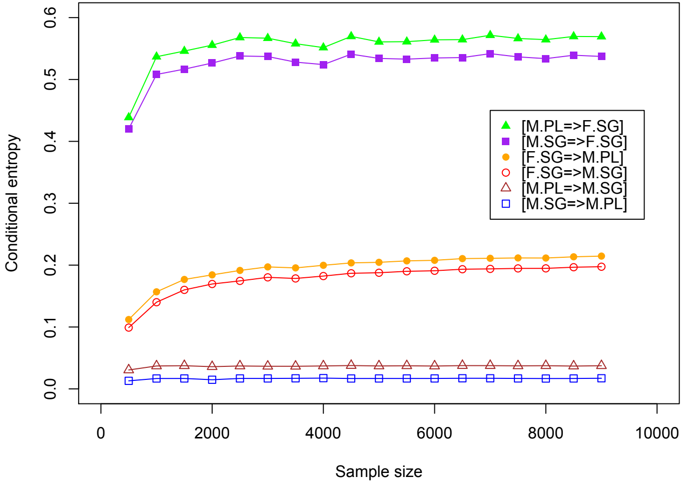
- Other possibly relevant factors: semantic classes (Baayen and Moscoso del Prado Martín, 2005), morphosyntactic properties, ...
- Gender of French nouns:

dataset	size	$H([\text{SG} \Rightarrow \text{PL}])$	$H([\text{PL} \Rightarrow \text{SG}])$
masc. nouns	19600	0.0152	0.0317
fem. nouns	14036	0.0000	0.0000
all nouns, gender ignored	33636	0.0120	0.0193
all nouns, with gender	33636	0.0089	0.0185

- All the uncertainty in [SG ⇒ PL] occurs on masculine nouns, mostly those ending in -al (*tribunal* vs. *festival*) or aj (*éventail* 'fan' vs. *vantail* 'casement')
- But there are also feminine nouns in -al (e.g. *cavale* '') and aj (e.g. *paille* 'straw')
- If gender is ignored, these nouns raise the uncertainty.

Illstrating caveat 3: influence of dataset

- Back to French adjectives
- Average entropy over 50 random samples of size 500, 1000, . . . , 9000
  - ☞ Sampling favors high token frequency, using data from `Lexique 3`



Binary implicative relations

- For now we have focused on unary implicative relations: the antecedent of the implication is a single cell.
- In the following toy example, intuitively:
  - From the INF, one can not be sure of the PST.PTCP.
  - From the PRS.2PL, one can not be sure of the PST.PTCP.
  - Yet from joint knowledge of the INF and the PRS.2PL, the PST.PTCP is known for sure.

lexeme	INF	PRS.2PL	PST.PTCP
TAPIR 'to hide'	tapiʁ	tapise	tapi
POURRIR 'to rot'	puʁiʁ	puʁise	puʁi
MOURIR 'to die'	muʁiʁ	muʁe	mɔʁ
TAPISSER 'to overlay'	tapise	tapise	tapise

- We are looking for a binary implication array: information on the likelihood of a PST.PTCP

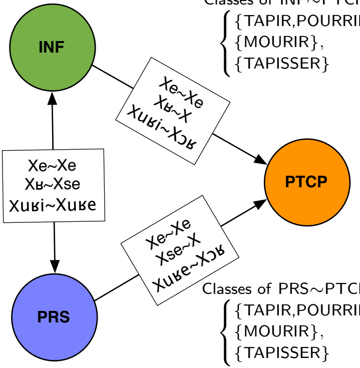
Deducing binary implication arrays

- Summing up all we know on INF and PRS.2PL from the unary arrays:

Classes of INF: $A = \left\{ \begin{array}{l} \{\text{TAPIR}\}, \\ \{\text{MOURIR,POURRIR}\}, \\ \{\text{TAPISSER}\} \end{array} \right\}$

Classes of INF∼PTCP: $D = \left\{ \begin{array}{l} \{\text{TAPIR,POURRIR}\}, \\ \{\text{MOURIR}\}, \\ \{\text{TAPISSER}\} \end{array} \right\}$

Classes of INF∼PRS: $B = \left\{ \begin{array}{l} \{\text{TAPIR,POURRIR}\}, \\ \{\text{MOURIR}\}, \\ \{\text{TAPISSER}\} \end{array} \right\}$



Classes of PRS: $C = \left\{ \begin{array}{l} \{\text{TAPIR,TAPISSER}\}, \\ \{\text{MOURIR}\}, \\ \{\text{POURRIR}\} \end{array} \right\}$

Classes of PRS∼PTCP: $E = \left\{ \begin{array}{l} \{\text{TAPIR,POURRIR}\}, \\ \{\text{MOURIR}\}, \\ \{\text{TAPISSER}\} \end{array} \right\}$

- We can combine these classifications to get a joint classification of patterns and a joint classification of input forms.

Deducing binary implication arrays

Classes of INF: $A = \left\{\begin{array}{l}\{\text{TAPIR}\},\\ \{\text{MOURIR,POURRIR}\},\\ \{\text{TAPISSER}\}\end{array}\right\}$

Classes of INF$\sim$PRS: $B = \left\{\begin{array}{l}\{\text{TAPIR,POURRIR}\},\\ \{\text{MOURIR}\},\\ \{\text{TAPISSER}\}\end{array}\right\}$

Classes of PRS: $C = \left\{\begin{array}{l}\{\text{TAPIR,TAPISSER}\},\\ \{\text{MOURIR}\},\\ \{\text{POURRIR}\}\end{array}\right\}$

Classes of INF$\sim$PTCP: $D = \left\{\begin{array}{l}\{\text{TAPIR,POURRIR}\},\\ \{\text{MOURIR}\},\\ \{\text{TAPISSER}\}\end{array}\right\}$

Classes of PRS$\sim$PTCP: $E = \left\{\begin{array}{l}\{\text{TAPIR,POURRIR}\},\\ \{\text{MOURIR}\},\\ \{\text{TAPISSER}\}\end{array}\right\}$

- Classification of pairs of patterns:
  $\{X \cap Y \mid \langle X, Y\rangle \in D \times E\} \setminus \emptyset = \Big\{\{\text{TAPIR,POURRIR}\},\{\text{MOURIR}\},\{\text{TAPISSER}\}\Big\}$
- Classification of pairs of input forms:
  $\{X \cap Y \cap Z \mid \langle X, Y, Z\rangle \in A \times B \times C\} \setminus \emptyset = \Big\{\{\text{TAPIR}\},\{\text{POURRIR}\},\{\text{MOURIR}\},\{\text{TAPISSER}\}\Big\}$

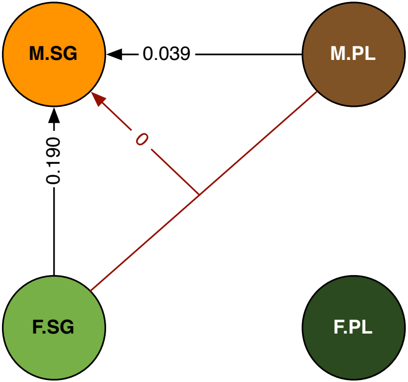
Deducing binary implication arrays

- We can now examine the conditional probability of a pair of patterns given a pair of input forms:

class	size	patterns	freqs
{TAPIR}	1	{TAPIR,POURRIR} : $\langle X$ʁ $\sim X, X$se $\sim X\rangle$	1
{POURRIR}	1	{TAPIR,POURRIR} : $\langle X$ʁ $\sim X, X$se $\sim X\rangle$	1
{MOURIR}	1	{MOURIR} : $\langle X$uʁiʁ $\sim X$oʁ$, X$uʁe $\sim X$oʁ$\rangle$	1
{TAPISSER}	1	{MOURIR} : $\langle X$e $\sim X$e$, X$e $\sim X$e$\rangle$	1

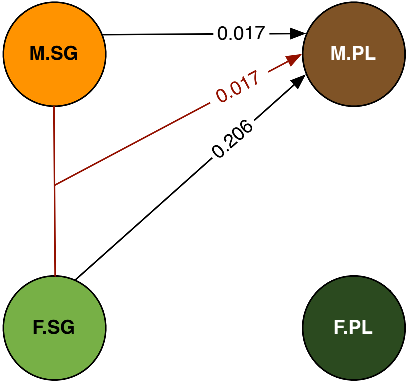
- In this particular (toy) example we end up with conditional entropy 0
- This procedure
  - is fully general
  - does not depend on any new inference of patterns
  - generalizes trivially to *n*-ary implicative relations
- The entropy of a binary array is at most as high as that of the most predictive unary array

More realistic examples



- The binary array is a lot more predictive that both unary arrays
  - All the uncertainty in [M.PL ⇒ M.SG] is due to lexemes with a M.PL in -o: is the M.SG in -al or -o?
  - The F.SG always disambiguates this: all lexemes with a M.SG in -al are also in -al in the F.SG.

More realistic examples



- The binary array is exactly as predictive as [M.SG ⇒ M.PL]
  - All the uncertainty in [M.SG ⇒ M.PL] is due to lexemes with a M.SG in -al. For those lexemes the F.SG provides no extra information.
  - The uncertainty in [F.SG ⇒ M.PL] is due to the same lexemes as that in [F.SG ⇒ M.SG]. Thus knowing the M.SG suppresses that uncertainty.

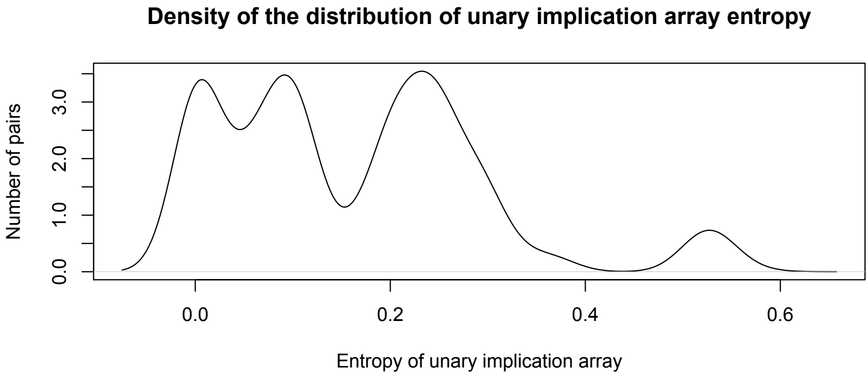
Conclusion on the implicative structure of French adjectives

- F.SG and F.PL are related by mutual 0 entropy arrays
  - They form an inflection zone (Bonami and Boyé, 2003), an alliance of forms (Ackerman et al., 2009), a distillation (Stump and Finkel, in press).
  - For computational efficiency, one of them can be dropped from further calculations.
- F is the best overall predictor of the rest of the paradigm.
- Uncertainty between M.SG and M.PL is due to a single pocket of lower predictibility
- Uncertainty between M.SG and F.SG is due to scattered idiosyncrasies
- These are the only two sources of uncertainty: no *specific* uncertainty between F.SG and M.PL
- F and M.PL constitute the only set of static principal parts (Finkel and Stump, 2007) for the rest of the paradigm.

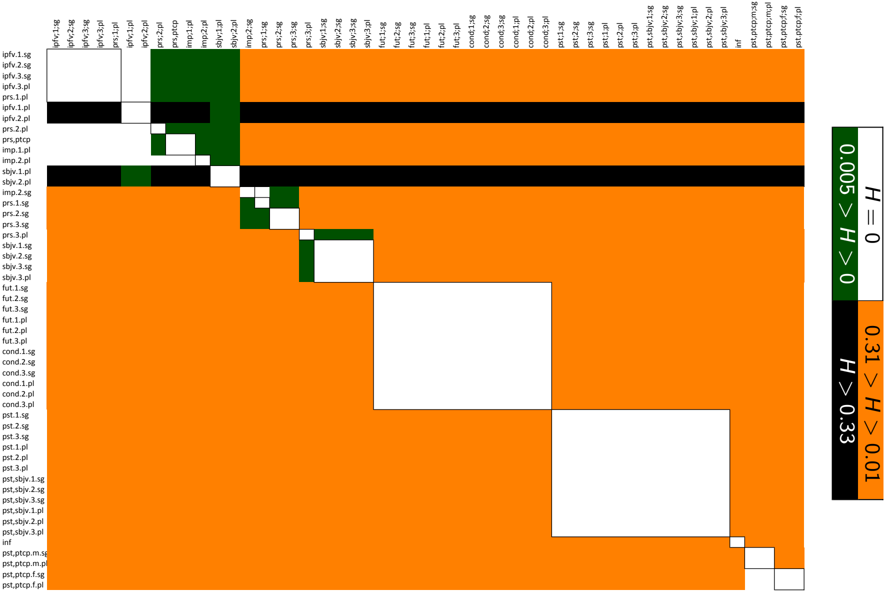
Structure

Unary arrays: the big picture

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



Unary arrays: basic classification



Alliances of forms

- We uncover 16 zones of perfect interpredictibility:

Finite forms

TEMPS	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						
COND			13			
PST						
PST.SBJV			13			

Nonfinite forms

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

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 distillation of the paradigm:

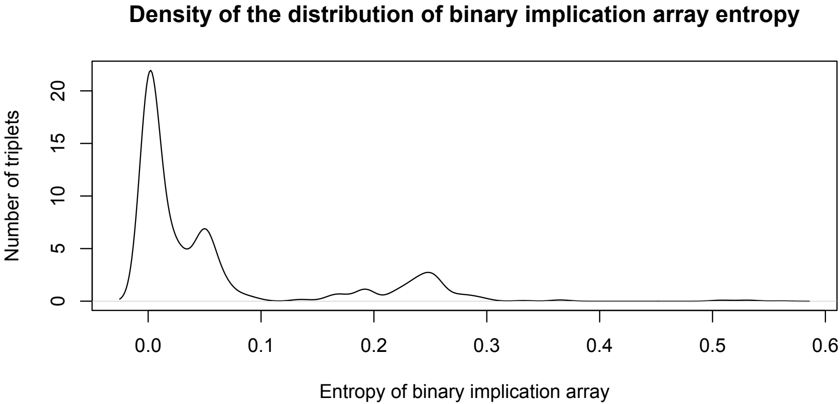
	PRS.1.SG	PRS.2.SG	PRS.1.PL	PRS.2.PL	PRS.3.PL	IPFV.1.PL	IMP.2.SG	IMP.2.PL	SBJV.1.SG	SBJV.1.PL	FUT.1.SG	PST.1.SG	INF	PRS.PTCP	PST.PTCP.M.SG	PST.PTCP.F.SG
PRS.1.SG	---	0,0011	0,2582	0,2558	0,234	0,2401	0,0008	0,2573	0,2447	0,2395	0,0839	0,2434	0,2786	0,2599	0,2166	0,2365
PRS.2.SG	0,0004	---	0,2681	0,2743	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.

Binary arrays

- Focussing on the distillation, there are $\frac{16*15*14}{2} = 1680$ binary arrays to consider
- Mean entropy on binary arrays: 0.0584
  - Compare: on unary arrays: 0.1618



Principal parts?

- There is no set of principal parts for French with cardinality 2
  - Not surprising: Stump and Finkel (in press) arrive at 5
- However there are some near principal part sets:
  - 4 pairs of cells from which the average entropy is below 0.001

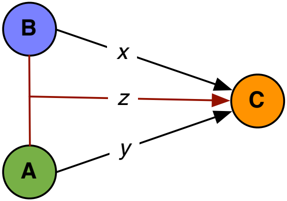
PRS.3PL	PST.PTCP.M.SG	0.00064
SBJV.3SG	PST.PTCP.M.SG	0.00061
PRS.3PL	PST.PTCP.F.SG	0.00046
SBJV.3SG	PST.PTCP.F.SG	0.00042

  ☞ Only a handful of lexemes are not predicted by these pairs.

☞ Predicting the last few lexemes is very hard, but is it very important?

Informativeness of binary implications

- A binary array is informative if its entropy is lower than the entropy of both corresponding unary arrays.
  - That is, $z < x$ and $z < y$



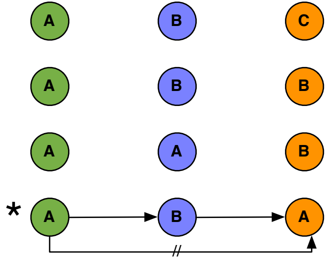
- In the French conjugation data:
  - 88% of binary arrays are informative.
  - 16% of binary arrays bring entropy down to 0.
  - For 53% of arrays, $z < \frac{1}{2}\min(x, y)$
    That is, the binary arrays shaves off at least half of the uncertainty.

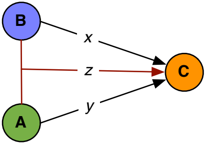
☞ There is a lot of implicative structure in the system that unary implications can not capture.

Binary implicative relations and stem graph

- Uninformative binary arrays relate to the central analytic technique in Morin (1987), (Boyé, 2000) and later work.

☞ Directional patterns in the distribution of unexpected forms





- Directional patterns emerge when the binary array is uninformative
- But most binary arrays are informative

☞ A graph of informative unary directional patterns is much more connected than (Boyé, 2011) suggests

Structure

Conclusions

- Implicative structure exists and should be studied *en elle-même*.
  - ☞ Whether it is reducible to something else is an important but separate matter
- I have motivated a particular way of investigating it
  - Builds on Ackerman et al. (2009) and later work in the same tradition
  - Allows for easy, fast computations
  - Arguably, more principled than previous approaches such as (Bonami and Boyé, 2002) or (Stump and Finkel, in press)
- I have illustrated the fruitfulness of automated analysis over semi-exhaustive datasets in inflection
  - We are working on small finite domains. For well-documented languages, there is no excuse for not exploring them thouroughly.
- Related projects
  - Studying quantitatively the complexity of Creole morphology (Bonami et al., 2011, 2012)
  - Portuguese conjugation (Bonami, Boyé, Luís & Tribout)
  - . . . any large enough dataset that is available

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

- Farrell Ackerman
- Jim Blevins
- Gilles Boyé
- Gauthier Caron
- Fabiola Henri
- Ana R. Luís
- Clément 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.

Baayen, H. and Moscoso del Prado Martín, F. (2005). 'Semantic density and past-tense formation in three Germanic languages'. *Language*, 81:666–698.

Baerman, M., Brown, D., and Corbett, G. G. (2005). *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

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.

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

——— (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. (2000). *Problèmes de morpho-phonologie verbale en français, espagnol et italien*. Ph.D. thesis, Université Paris 7.

References II

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.

Durand, M. (1936). *Le genre grammatical en français parlé à Paris et dans la réguion parisienne*. Paris: Bibliothèque du "français moderne".

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.

Matthews, P. H. (1965). 'The inflectional component of a word-and-paradigm grammar'. *Journal of Linguistics*, 1:139–171.

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.

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

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 Boyé and 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

TEMPS	1SG	2SG	3SG	1PL	2PL	3PL
PRS	3					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	3A	3B		1B	1C	2
IPFV	1B			1A		1B
IMP	—	5	—	6A	6B	—
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	12A		12B	

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