Paradigm entropy as a measure of morphological simplicity

Workshop on Morphological Complexity
January 22, 2010

Rob Malouf
San Diego State University

Farrell Ackerman
UC San Diego

Our goal

"To exclude the evidence which their languages offer as to what the human mind can do is like expecting botanists to study nothing but food plants and hothouse roses and then tell us what the plant world is like." (Whorf 1940/1956:215)

Our goal is identify a dimension of simplicity the underlies the apparent "gratuitous complexity" (Baerman, et. al. 2009) of many morphological systems

What makes a language difficult to learn and use (not to describe)?

The issue is not complexity per se, but the nature of organization supporting that complexity (reflective of memory storage for words, patterns, and procedures for generalization)

Principle of Low Paradigm Entropy: Paradigms tend to have low expected conditional entropy

The Paradigm Cell Filling Problem

"Don't you see that neither you nor anybody else has ever heard all of the nouns of the paradigm *fa'il* or *maf'ul*? You have heard some forms and then you have proceeded by analogy to produce others." (Langhade 1985:111, cited in Itkonen 2005:89)

Speakers of languages with complex morphology and multiple inflection classes must generalize beyond direct experience, since it's implausible to imagine they all will have encountered each form of every word

Paradigm Cell Filling Problem: Given exposure to an inflected wordform of a novel lexeme, what licenses reliable inferences about the other wordforms in its inflectional family?

The Paradigm Cell Filling Problem

Nom Pl	**Acc Pl**	**Gen Pl**	**Dat Pl**	**Loc Pl**	**Ins Pl**	
dela	dela	del	delam	delax	delam’i	‘thing’
zavodi	zavodi	zavodov	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	stran	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	kost’ej	kost’am	kost’ax	kost’am’i	‘bone’

Russian noun declensions (Baerman, et al. 2009)

The Paradigm Cell Filling Problem

Nom Pl	**Acc Pl**	**Gen Pl**	**Dat Pl**	**Loc Pl**	**Ins Pl**	
dela	dela	del	delam	delax	delam’i	‘thing’
zavodi	zavodi	zavodov	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	stran	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	kost’ej	kost’am	kost’ax	kost’am’i	‘bone’

NOM. SG. -*a* → GEN. PL. -∅

The Paradigm Cell Filling Problem

Nom Pl	Acc Pl	Gen Pl	Dat Pl	Loc Pl	Ins Pl	
dela	dela	del	delam	delax	delam’i	‘thing’
zavodi	zavodi	zavodov	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	stran	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	kost’ej	kost’am	kost’ax	kost’am’i	‘bone’

GEN. PL. -∅ → NOM. SG. *-a / -i*

The Paradigm Cell Filling Problem

Word-based proposals offer one solution: (Pauonen 1976; Wurzel 1989; Thymé 1994; Thymé, Ackerman & Elman 1994; Ackerman, Blevins, & Malouf 2009 among others in WP tradition)

Paradigms are networks of implicative relations among related wordforms

Inflectional classes are patterns of wordforms displaying distinctive implicational relations

No matter how complex they look, languages must be simple enough to allow speakers to solve the Paradigm Cell Filling Problem

How can we measure this?

Information Theory

Shannon's (1948) Information Theory gives us a way to quantify the uncertainty in a random variable

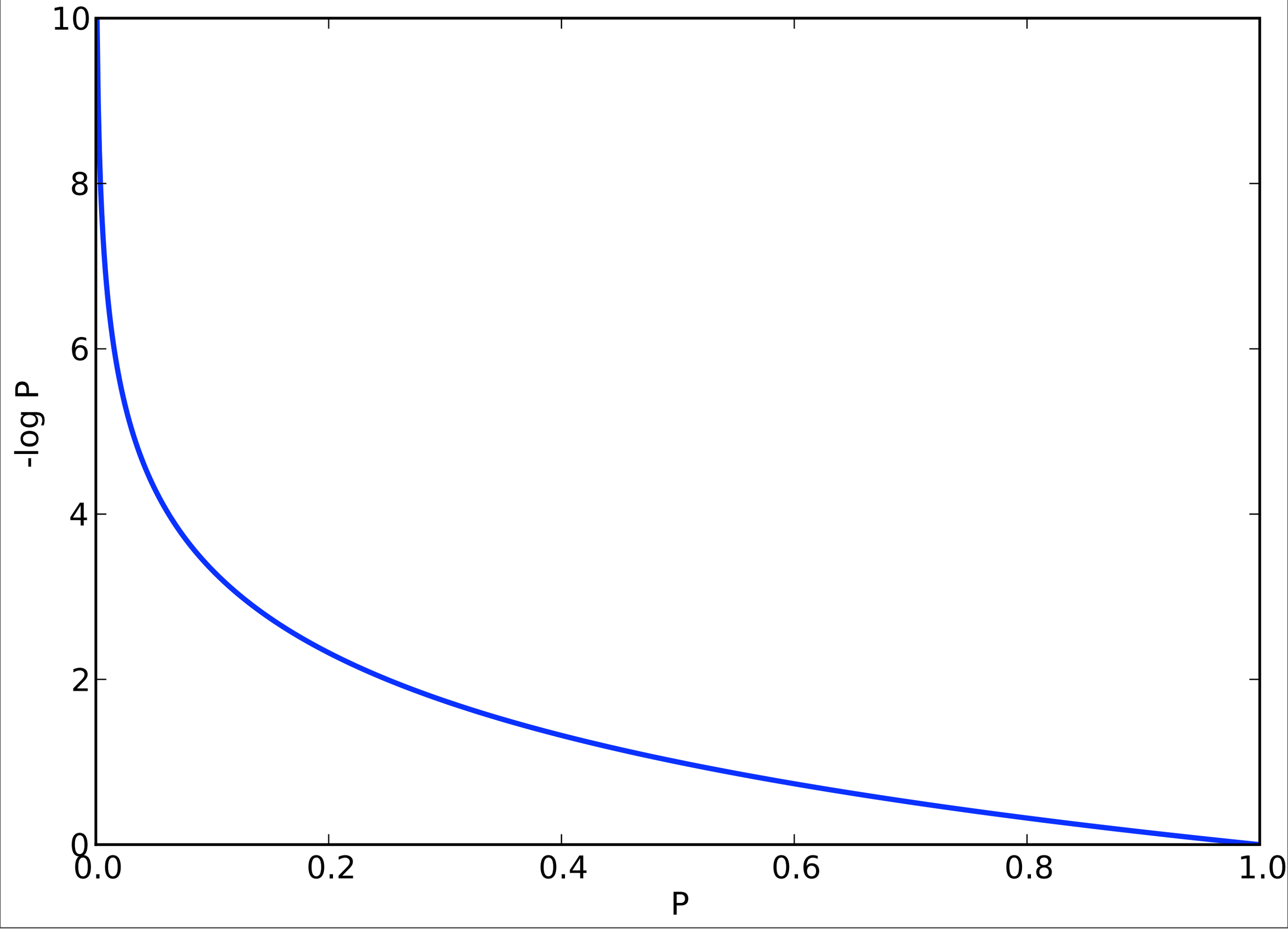
The key concept is **information entropy**

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

The entropy is the average number bits required to store the value of *X*

Also, the average number of yes-or-no questions you'd have to ask to guess the value of *X*

This is the expected value of the **surprisal** for each possible value


10
8
6
4
2
0
-log P
0.0
0.2
0.4
0.6
0.8
1.0
P

Entropy

Nom Pl	Acc Pl	Gen Pl	Dat Pl	Loc Pl	Ins Pl	
dela	dela	**del**	delam	delax	delam’i	‘thing’
zavodi	zavodi	**zavodov**	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	**stran**	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	**kost’ej**	kost’am	kost’ax	kost’am’i	‘bone’

$$H(\text{gen.pl.}) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$
$$= 1.5 \text{ bits}$$

Entropy

Nom Pl	Acc Pl	Gen Pl	Dat Pl	Loc Pl	Ins Pl	
dela	dela	del	delam	**delax**	delam’i	‘thing’
zavodi	zavodi	zavodov	zavodam	**zavodax**	zavodam’i	‘factory’
strani	strani	stran	stranam	**stranax**	stranam’i	‘country’
kost’i	kost’i	kost’ej	kost’am	**kost’ax**	kost’am’i	‘bone’

$$H(\text{loc.pl.}) = -\left(\frac{4}{4}\log_2\frac{4}{4}\right)$$
$$= 0 \text{ bits}$$

Entropy

An entropy of 1.5 bits for genitive plural corresponds to a choice among 2.8 equally likely declensions

This gives us an **upper bound** on the entropy of each paradigm cell

The actual entropy will be lower if:

Not all declensions are equally likely

Forms can be predicted in any way by external factors (phonology, semantics)

Speakers can generalize from two or more known forms

Entropy

Entropy measures the uncertainty in choosing a realization for a single paradigm cell

The expected entropy is the average entropy across all paradigm cells:

$$E[H] = \sum_{C \in P} p(C)\, H(C)$$

The expected entropy for the complete Russian noun paradigm (the average uncertainty in guessing the realization for a single cell of a single lexeme's paradigm) is 0.911 bits

The **declension entropy**, uncertainty in guessing the declension of a lexeme, is $\log_2 4 = 2$ bits

Conditional entropy

The **conditional entropy** is the uncertainty in one random variable on average, given that we know the value of another random variable

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x)$$
$$= H(X,Y) - H(X)$$

Information is anything that reduces entropy/uncertainty

We can now apply this to the PCFP to measure the degree of inter-prediction in a paradigm

Conditional entropy

Nom Pl	Acc Pl	Gen Pl	Dat Pl	Loc Pl	Ins Pl	
dela	dela	**del**	delam	delax	delam’i	‘thing’
zavodi	zavodi	zavodov	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	stran	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	kost’ej	kost’am	kost’ax	kost’am’i	‘bone’

$$H(\text{gen.pl.}|\text{nom.pl.} = \text{-}a) = -\left(\frac{1}{1}\log_2\frac{1}{1}\right)$$
$$= 0 \text{ bits}$$

Conditional entropy

Nom Pl	**Acc Pl**	**Gen Pl**	**Dat Pl**	**Loc Pl**	**Ins Pl**	
dela	dela	del	delam	delax	delam’i	‘thing’
zavodi	zavodi	**zavodov**	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	**stran**	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	**kost’ej**	kost’am	kost’ax	kost’am’i	‘bone’

$$H(\text{gen.pl.}|\text{nom.pl.} = \text{-}i) = -\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)$$
$$= 1.6 \text{ bits}$$

Conditional entropy

Nom Pl	Acc Pl	Gen Pl	Dat Pl	Loc Pl	Ins Pl	
dela	dela	**del**	delam	delax	delam’i	‘thing’
zavodi	zavodi	**zavodov**	zavodam	zavodax	zavodam’i	‘factory’
strani	strani	**stran**	stranam	stranax	stranam’i	‘country’
kost’i	kost’i	**kost’ej**	kost’am	kost’ax	kost’am’i	‘bone’

$$H(\text{gen.pl.}|\text{nom.pl.}) = -\left(\frac{1}{4}\times 0 + \frac{3}{4}\times 1.6\right)$$
$$= 1.2 \text{ bits}$$

Paradigm entropy

The conditional entropy of one cell given another is a measure of inter-predictability

To extend this to the whole paradigm, we calculate the expected conditional entropy or **paradigm entropy** (Ackerman, Blevins, & Malouf 2009)

$$E[H(C_1|C_2)] = \sum_{C_1, C_2 \in P} p(C_1, C_2)\, H(C_2|C_1)$$

The higher the paradigm entropy, the more difficult it is to predict an unknown wordform, given a known wordform.

For the Russian nouns, the paradigm entropy is 0.538 bits

Paradigm entropy

Paradigms vary a lot in their apparent morphological complexity

For all these paradigms, the paradigm entropy is much lower than either the expected entropy or the declension entropy

Language	Declensions	Cells	Realizations	Max realizations	Possible declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	24	3	31	14	1,008	4.585	2.882	1.105
Arapesh	26	2	41	26	390	4.700	4.071	0.630
Burmeso	2	12	24	2	4,096	1.000	1.000	0.000
Fur	19	12	80	10	6,096,384,000	4.248	2.395	0.517
Kwerba	4	12	26	4	3,072	2.000	0.864	0.428
Ngiti	10	16	68	5	6,400,000,000	3.322	1.937	0.484
Nuer	16	6	12	3	48	4.000	0.864	0.793
Russian	4	12	26	3	3,888	2.000	0.911	0.538

Paradigm organization

Some entropy-lowering strategies:

Small number of cells, forms, inflection classes

Paradigm Economy Principle (Carstairs 1984), No Blur Principle (Carstairs-McCarthy 1994)

Language	Declensions	Cells	Realizations	Max realizations	Possible declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	24	3	31	14	1,008	4.585	2.882	1.105
Arapesh	26	2	41	26	390	4.700	4.071	0.630
Burmeso	2	12	24	2	4,096	1.000	1.000	0.000
Fur	19	12	80	10	6,096,384,000	4.248	2.395	0.517
Kwerba	4	12	26	4	3,072	2.000	0.864	0.428
Ngiti	10	16	68	5	6,400,000,000	3.322	1.937	0.484
Nuer	16	6	12	3	48	4.000	0.864	0.793
Russian	4	12	26	3	3,888	2.000	0.911	0.538

Paradigm Economy Principle

Paradigm Economy Principle (Carstairs 1984): There are as many inflection classes as there are realizations of the paradigm cell with the most realizations

Hungarian present indefinite verbs (factoring out phonologically conditioned allomorphy):

	INDICATIVE		SUBJUNCTIVE	
SING 1	*-ok*	*-ik*	*-ak*	*-am*
2	*-ol*	*-om*	*-∅*	
3	*-∅*	*-ik*	*-on*	*-ék*
PL 1	*-unk*		*-unk*	
2	*-tok*		*-tok*	
3	*-nak*		*-nak*	

Paradigm Economy Principle

Mystery if PEP is correct: Syntagmatic organization (the composition of wordforms) constrains paradigmatic organization (the number of classes)

PEP, at best, accounts for the partitioning of paradigms into classes, but not for how to calculate relatedness of wordforms within and across classes.

Is the explanation really in terms of the influence of internal structure on external organization?

Paradigm Economy Principle

Hypothesis: Languages with low paradigm entropy (and, hence, transparency among classes of related wordforms) permit speakers to generalize beyond their limited experience.

The PEP holds (to the extent that it does) because it is one strategy for restricting paradigm entropy, while paradigm entropy accounts for more than PEP effects. (Malouf & Ackerman 2010)

Size of paradigms (in terms of class number and wordform complexity) doesn't matter, it's how they're organized that matters.

Inflectional systems with more classes than the PEP would predict may be organized in ways that reduce paradigm entropy (Stump & Finkel 2007)

Paradigm organization

Fur and Russian paradigms look very different, but have similar paradigm entropies

Principal parts (Finkel & Stump 2007)

Language	Declensions	Cells	Realizations	Max realizations	Possible declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	24	3	31	14	1,008	4.585	2.882	1.105
Arapesh	26	2	41	26	390	4.700	4.071	0.630
Burmeso	2	12	24	2	4,096	1.000	1.000	0.000
Fur	19	12	80	10	6,096,384,000	4.248	2.395	0.517
Kwerba	4	12	26	4	3,072	2.000	0.864	0.428
Ngiti	10	16	68	5	6,400,000,000	3.322	1.937	0.484
Nuer	16	6	12	3	48	4.000	0.864	0.793
Russian	4	12	26	3	3,888	2.000	0.911	0.538

	‘thing’	‘factory’	‘country’	‘bone’
NOM SG	del-o	zavod	stran-a	kost′
ACC SG	del-o	zavod	stran-u	kost′
GEN SG	del-a	zavod-a	stran-i	kost′-i
DAT SG	del-u	zavod-u	stran-e	kost′-i
LOC SG	del-e	zavod-e	stran-e	kost′-i
INS SG	del-om	zavod-om	stran-oj	kost′-ju
NOM PL	del-a	zavod-i	stran-i	kost′-i
ACC PL	del-a	zavod-i	stran-i	kost′-i
GEN PL	del	zavod-ov	stran	kost′-ej
DAT PL	del-am	zavod-am	stran-am	kost′-am
LOC PL	del-ax	zavod-ax	stran-ax	kost′-ax
INS PL	del-am′i	zavod-am′i	stran-am′i	kost′-am′i

Figure 7: Russian inflectional classes (phonological transcription)

TABLE 16. Tonal and suffixal exponents of verb classes in Fur (Jakobi 1990:103-113)

Class	Agreement	Subjunctive	Perfect	Present	Example
I,1a	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	LH-o HH-o HH-òl LH-òl	LH-ò HH-ò HH-ùl LH-ùl	LH-èl HH-èl HH-èl-à/-ì LH-èl-à/-ì	buuN ‘descend’
I,1b	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	LH-o HH-o HH-òl LH-òl	LH-ò HH-ò HH-ùl LH-ùl	LF-Ø HF-Ø HH-è LH-è	jaan ‘wait’
I,1c	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	LH-o HH-o HH-òl LH-òl	LH-ò HH-ò HH-ùl LH-ùl	LH-ì HH-ì HH-è LH-è	irt ‘shake’
I,2a	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LL-o LL-òl HH-òl	HH-o LL-ò LL-ùl HH-ùl	HH-èl LL-èl LL-èl-à/-ì HH-èl-à/-ì	tall ‘chew’
I,2b	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LL-o LL-òl HH-òl	HH-o LL-ò LL-ùl HH-ùl	HF-Ø LL-Ø LL-è HH-è	fuul ‘spin’
I,2c	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LL-o LL-òl HH-òl	HH-o LL-ò LL-ùl HH-ùl	HH-ì LL-ì LL-è HH-è	kir ‘cook’
II,1a	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	LH-i HH-i HH-i-A(l) LH-i-A(l)	LH-i HH-i HH-i-è LH-i-è	LH-itì HH-itì HH-iti-A(l) LH-iti-A(l)	rii ‘snatch’
II,1b	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	LH-i HH-i HH-i-A(l) LH-i-A(l)	LH-i HH-i HH-i-è LH-i-è	LF-Ø HF-Ø HH-è LH-è	tiir ‘meet’
II,2a	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ì LL-i LL-i-A(l) HH-i-A(l)	HH-ì LL-i LL-i-è HH-i-è	HH-itì LL-itì LL-iti-A(l) HH-iti-A(l)	faul ‘open’
II,2b	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ì LL-i LL-i-A(l) HH-i-A(l)	HH-ì LL-i LL-i-è HH-i-è	HF-Ø LF-Ø LL-è HH-è	kaun ‘grind’
IIIa	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ì LH-ì LH-è HH-è	HH-à LH-à LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	arr ‘measure’
IIIb	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LH-ò LH-è HH-è	HH-ò LH-ò LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	awi ‘pound’
IIIc	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LF-Ø LH-è HH-è	HH-ò LH-ò LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	dus ‘tear’ (tr)
IIId	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HF-Ø LF-Ø LH-è HH-è	HH-à LH-à LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	kair ‘stop’ (itr)
IIIe	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HF-Ø LF-Ø LH-è HH-è	HH-à LH-ò LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	tai ‘hold, seize’
IVa	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HF-Ø LF-Ø LH-Al HH-Al	HH-ò LH-ò LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	jum ‘cover’
IVb	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-ò LH-ò LH-Al HH-Al	HH-ò LH-ò LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	bul ‘find’
IVc	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HF-Ø LF-Ø LH-Al HH-Al	HH-à LH-à LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	juuN ‘terrify’
IVd	1/2 SG/PL 3 SG 3 PL [–HUM] 3 PL [+HUM]	HH-à LH-à LH-Al HH-Al	HH-à LH-à LH-e HH-e	HH-èl LH-èl LH-èl-à HH-èl-à	kur ‘touch’

A is a morphophoneme realized as [o] after a high vowel and otherwise as [a]; see Jakobi (1990:80f) for details.

Paradigm organization

The implicational structure of the paradigm is crucial to reducing paradigm entropy in Fur

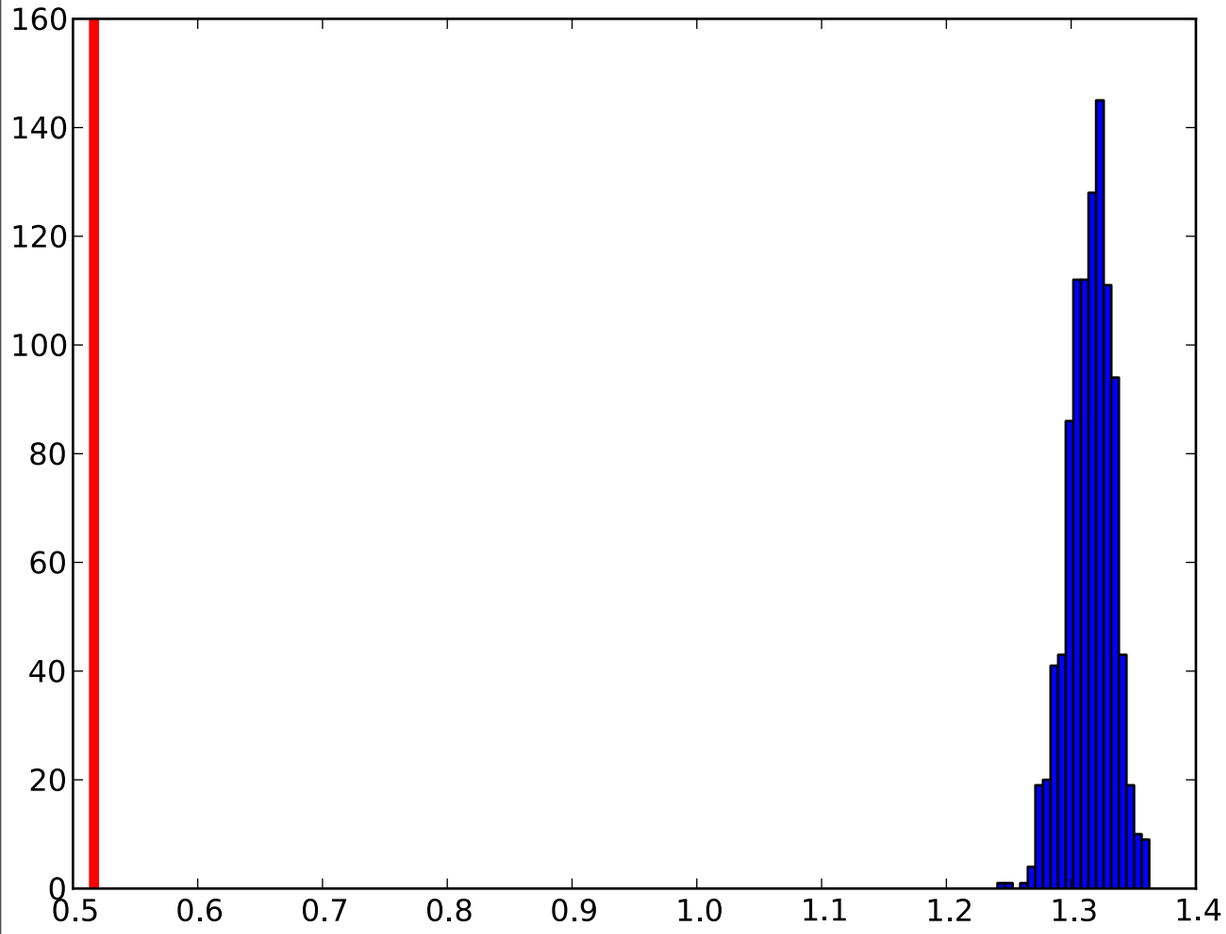
Paradigm organization is much less important in Russian, where the simplicity of the forms and morphosyntactic property system limits paradigm entropy

How can we test this?

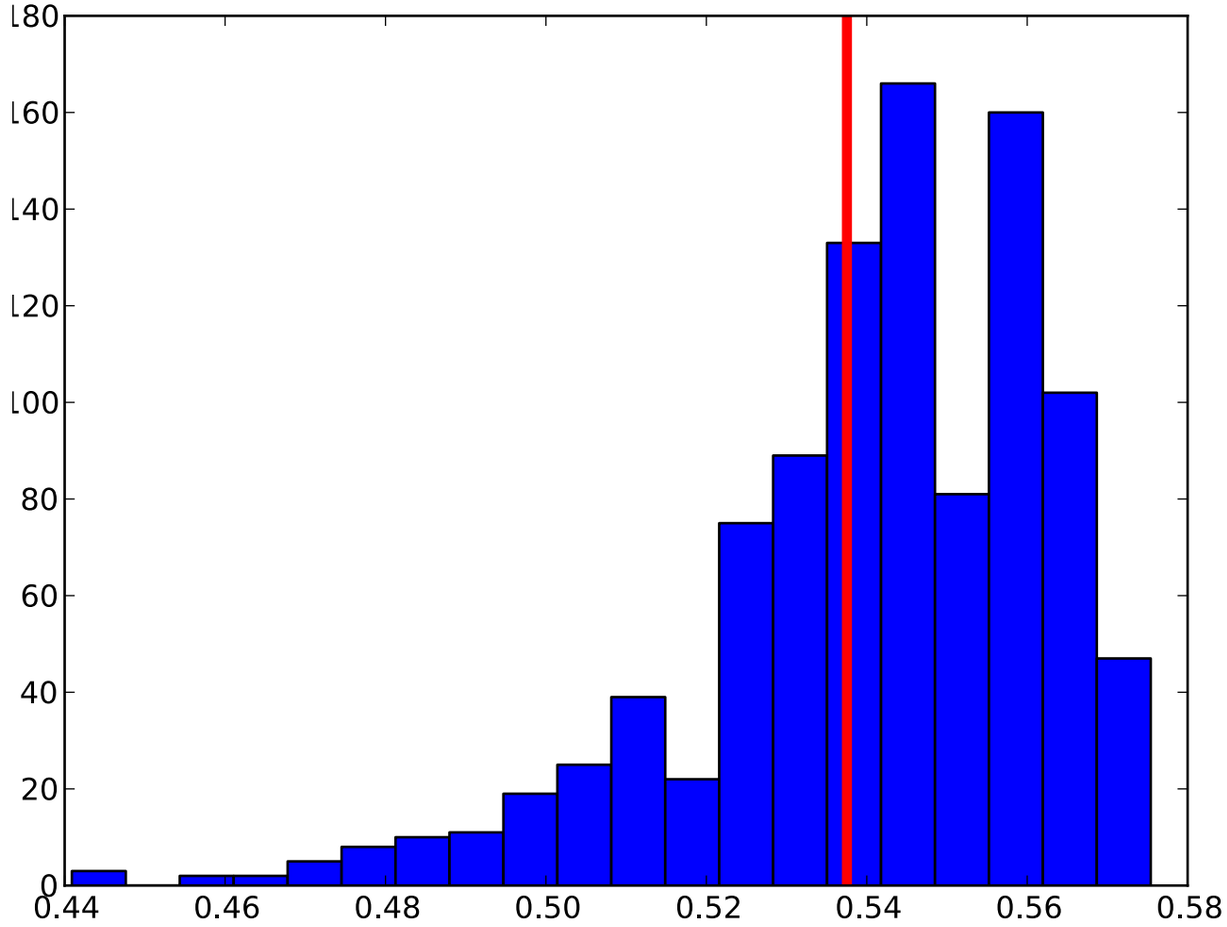
Null hypothesis: Paradigm entropy of language *L* is independent of paradigm organization

If this is true, then L_0, a version *L* with the same forms and the same classes but a different organization, should have more or less the same paradigm entropy

Bootstrap test: sample with replacement from the space of possible L_0's, and compare to the observed *L*


160
140
120
100
80
60
40
20
0
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4

Fur


180
160
140
120
100
80
60
40
20
0
0.44
0.46
0.48
0.50
0.52
0.54
0.56
0.58

Russian

Language	Declensions	Cells	Realizations	Declension entropy	Expected entropy	Paradigm entropy	Bootstap Avg	Bootstrap *p*
Amele	24	3	31	4.585	2.882	1.105	1.327	0.001
Arapesh	26	2	41	4.700	4.071	0.630	0.630	1.000
Burmeso	2	12	24	1.000	1.000	0.000	0.000	1.000
Fur	19	12	80	4.248	2.395	0.517	1.316	0.001
Kwerba	4	12	26	2.000	0.864	0.428	0.523	0.001
Ngiti	10	16	68	3.322	1.937	0.484	1.019	0.001
Nuer	16	6	12	4.000	0.864	0.793	0.811	0.160
Russian	4	12	26	2.000	0.911	0.538	0.541	0.383

External factors

Amele (Roberts 1987) is described in WALS as having 31 different classes of possessive suffixes plus a postposition

Hein and Müller (2009) argue that factoring out phonologically predictable alternations reduces this to 23 suffixed classes

H & M's paradigms have an entropy of 1.105 bits!

But, some facts:

- Possessive suffixes only apply to a closed class of 109 inalienably possessed nouns
- A combination of almost (but not quite) categorical semantic and phonological patterns generate most of the classes
- Many classes have only a single member

External factors

Three lessons from Amele:

We need a way of quantifying the contribution of non-morphological information to entropy reduction

Unexpectedly high entropy can be an indicator that something important has been missed in the description or analysis

Beware of paradigms on handouts!

Prospects

Paradigm entropy measures the complexity of a paradigm with respect to the Paradigm Cell Filling Problem

There are many ways that morphological systems can be complex, but (perhaps) only oneway that they can be rendered simple

Questions:

- What is the range of paradigm entropies in real typologically-diverse languages?
- What are the ways that paradigms can be organized to manage complexity (and keep paradigm entropy low)?
- Are there other aspects of morphological simplicity that can be quantified?

"Why complexity?" becomes "Why not?"

Language sources

Amele	Hein & Müller (2009), Roberts (1987)
Arapesh	Hein & Müller (2009)
Burmeso	Baerman et al. (2009)
Fur	Finkel & Stump (2007)
Kwerba	Finkel & Stump (2007)
Ngiti	Finkel & Stump (2007)
Nuer	Baerman et al. (2009)
Russian	Baerman et al. (2009)