Paradigms:

The Low entropy conjecture

Workshop on Morphology and Formal Grammar

July 8, 2010

Rob Malouf
San Diego State University

Farrell Ackerman
University of California at San Diego

The basic background

Inflectional morphology exhibits spectacular complexity in:

(i) syntagmatic and supersegmental structure of individual words

(ii) paradigmatic organization that systems of words participate in.

Languages can have enormous inventories of morphosyntactic distinctions and many ways to formally express them:

West Greenlandic (Fortescue 2002:264)

aju-nngit-su-liur-vigi-nnit-tuar-tu-u-nngil-aq
be.good-NEG-PART-make-have.as.place.of-ANTIPASS-all.the.time-PART-be-NEG-3S.INDIC
'He is not (much of) a benefactor.'

This is the **External Complexity** or **E-complexity** of a morphological system

Our goal

Identify a dimension of simplicity that underlies the apparent "gratuitous complexity" (Baerman, et. al. 2009) of many morphological systems

Our guiding intuition: morphological systems **must** be simple in ways that allow them to be learned and used by native speakers, irrespective of how complex words and paradigms may appear according to external measures.

Speakers must generalize beyond their direct experience:

> Morphological systems must permit speakers to make accurate guesses about unknown forms of lexemes based on only a few known forms.

This is the **Internal Simplicity** or **I-simplicity** of a system

Our hypothesis: I-simplicity

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

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

I-simplicity is measurable and quantifiable

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

Organization

1. **The Paradigm Cell Filling Problem: Modern Greek**
2. Paradigm entropy: A language sample
3. A surprisingly simple case: Chiquihuitlán Mazatec
4. Testing entropy claims

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 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? (Ackerman, Blevins, & Malouf 2009)

The Paradigm Cell Filling Problem

Modern Greek nominal paradigms (Ralli 1994, 2002; cf. Sims 2010)

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	*-os*	*-u*	*-on*	*-e*	*-i*	*-on*	*-us*	*-i*
2	*-s*	*-∅*	*-∅*	*-∅*	*-es*	*-on*	*-es*	*-es*
3	*-∅*	*-s*	*-∅*	*-∅*	*-es*	*-on*	*-es*	*-es*
4	*-∅*	*-s*	*-∅*	*-∅*	*-is*	*-on*	*-is*	*-is*
5	*-os*	*-u*	*-o*	*-o*	*-a*	*-on*	*-a*	*-a*
6	*-∅*	*-u*	*-∅*	*-∅*	*-a*	*-on*	*-a*	*-a*
7	*-os*	*-us*	*-os*	*-os*	*-i*	*-on*	*-i*	*-i*
8	*-∅*	*-os*	*-∅*	*-∅*	*-a*	*-on*	*-a*	*-a*

Paradigm entropy

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*)

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

The **declension entropy**, the uncertainty in guessing the declension of a lexeme, for Modern Greek is $\log_2 8 = 3$ bits

Entropy can also measure the uncertainty in choosing a realization for a single paradigm cell

Paradigm entropy

Expected entropy is **1.621 bits**, equivalent to a choice among $2^{1.621} \approx 3$ equally likely declensions

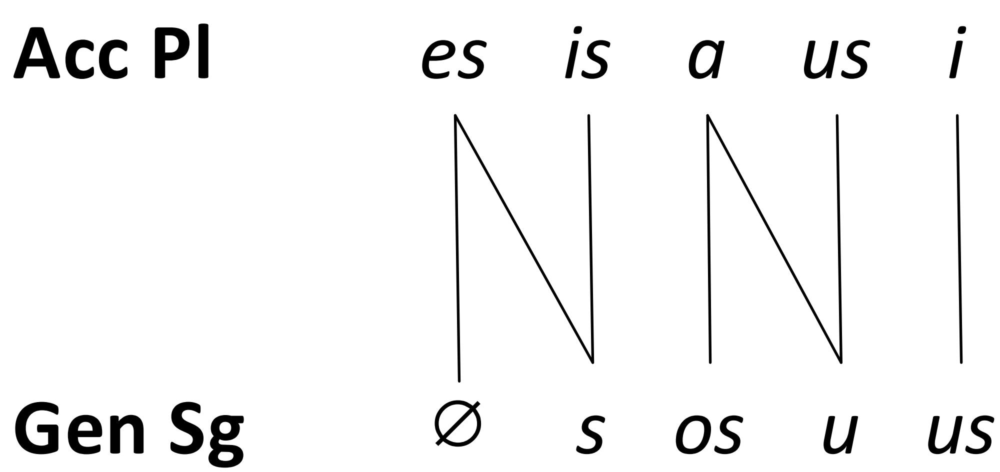
	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	*os*	*u*	*on*	*e*	*i*	*on*	*us*	*i*
2	*s*	∅	∅	∅	*es*	*on*	*es*	*es*
3	∅	*s*	∅	∅	*es*	*on*	*es*	*es*
4	∅	*s*	∅	∅	*is*	*on*	*is*	*is*
5	*o*	*u*	*o*	*o*	*a*	*on*	*a*	*a*
6	∅	*u*	∅	∅	*a*	*on*	*a*	*a*
7	*os*	*us*	*os*	*os*	*i*	*on*	*i*	*i*
8	∅	*os*	∅	∅	*a*	*on*	*a*	*a*
Entropy	1.750	2.156	1.549	1.549	1.906	0.000	2.156	1.906

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 or forms can be predicted in any way by external factors

Paradigm entropy

The GenSg and AccPl each have five distinct realizations and an entropy of 2.156 bits

Neither form is predictable from the other, but only 7 of the 25 possible pairings occur



Knowing one of these forms provides a lot of information about the other

Paradigm entropy

The **conditional entropy** is a measure of inter-predictability: the uncertainty in one random variable on average, given that we know the value of another random variable

To extend this to the whole paradigm, we calculate the expected conditional entropy or **paradigm entropy** (Ackerman, Blevins, & Malouf 2009; Malouf & Ackerman 2010)

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

Paradigm entropy

Paradigm entropy for [this fragment of] Modern Greek is **0.644 bits**, equivalent to a choice between 1.6 equally likely alternatives

H(col | row)	NomSg	GenSg	AccSg	VocSg	NomPl	GenPl	AccPl	VocPl	E[row]
NomSg	—	1.000	0.250	0.250	0.750	0.000	1.000	0.750	0.571
GenSg	0.594	—	0.594	0.594	0.594	0.000	0.594	0.594	0.509
AccSg	0.451	1.201	—	0.000	0.951	0.000	0.951	0.951	0.644
VocSg	0.451	1.201	0.000	—	0.951	0.000	0.951	0.951	0.644
NomPl	0.594	0.844	0.594	0.591	—	0.000	0.250	0.000	0.411
GenPl	1.750	2.156	1.549	1.549	1.906	—	2.156	1.906	1.853
AccPl	0.594	0.594	0.344	0.344	0.000	0.000	—	0.000	0.268
VocPl	0.594	0.844	0.594	0.594	0.000	0.000	0.250	—	0.411
E[col]	0.719	1.120	0.561	0.561	0.736	0.000	0.879	0.736	0.664

Organization

1. The Paradigm Cell Filling Problem: Modern Greek
2. **Paradigm entropy: A language sample**
3. A surprisingly simple case: Chiquihuitlán Mazatec
4. Testing entropy claims

Paradigm entropy

Paradigms vary a lot in their apparent E-complexity.

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

Language	Cells	Realizations	Max realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	3	31	14	24	4.585	2.882	1.105
Arapesh	2	41	26	26	4.700	4.071	0.630
Burmeso	12	24	2	2	1.000	1.000	0.000
Fur	12	80	10	19	4.248	2.395	0.517
Greek	8	12	5	8	3.000	1.621	0.644
Kwerba	12	26	4	4	2.000	0.864	0.428
Mazatec	6	356	94	109	6.768	4.920	0.709
Ngiti	16	68	5	10	3.322	1.937	0.484
Nuer	6	12	3	16	4.000	0.864	0.793
Russian	12	26	3	4	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	Cells	Realizations	Max realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	3	31	14	24	4.585	2.882	1.105
Arapesh	2	41	26	26	4.700	4.071	0.630
Burmeso	12	24	2	2	1.000	1.000	0.000
Fur	12	80	10	19	4.248	2.395	0.517
Greek	8	12	5	8	3.000	1.621	0.644
Kwerba	12	26	4	4	2.000	0.864	0.428
Mazatec	6	356	94	109	6.768	4.920	0.709
Ngiti	16	68	5	10	3.322	1.937	0.484
Nuer	6	12	3	16	4.000	0.864	0.793
Russian	12	26	3	4	2.000	0.911	0.538

Paradigm organization

Some entropy-lowering strategies:

Implicational relations (Wurzel 1989)

Principal parts (Stump & Finkel 2007)

Language	Cells	Realizations	Max realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy
Amele	3	31	14	24	4.585	2.882	1.105
Arapesh	2	41	26	26	4.700	4.071	0.630
Burmeso	12	24	2	2	1.000	1.000	0.000
Fur	12	80	10	19	4.248	2.395	0.517
Greek	8	12	5	8	3.000	1.621	0.644
Kwerba	12	26	4	4	2.000	0.864	0.428
Mazatec	6	356	94	109	6.768	4.920	0.709
Ngiti	16	68	5	10	3.322	1.937	0.484
Nuer	6	12	3	16	4.000	0.864	0.793
Russian	12	26	3	4	2.000	0.911	0.538

Paradigm organization

Principal parts are a small set of wordforms that are diagnostic of inflection class membership for lexemes; some wordforms are diagnostic, while others are not.

This misses the essential insight that all wordforms contribute in some measure to implicational networks of relatedness

Organizing paradigms around a small set of principal parts is merely one way that I-simplicity can be achieved

Organization

1. The Paradigm Cell Filling Problem: Modern Greek

2. Paradigm entropy: A language sample

3. **A surprisingly simple case: Chiquihuitlán Mazatec**

4. Testing entropy claims

A surprisingly simple case

In Chiquihuitlán Mazatec, verbs are marked for person and aspect by a combination of tones, final vowel, and stem formative (Jamieson 1982, Capen 1996, Baerman & Corbett 2010)

Positive paradigm for *ba³se²* ‘remember’

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		*ča²sẽ²*		*ča²sẽ⁴²*
1	*ba³sæ¹*	*ča²sĩ²⁴*	*kua³sæ¹*	*ča⁴sĩ²⁴*
2	*ča²se²*	*ča²sũ²*	*ča⁴se²*	*ča⁴sũ²*
3	*ba³se²*		*kua⁴se²*	

-s- ‘remember’

Tone class B31

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		2-2		4-42
1	3-1	2-24	3-1	4-24
2	2-2	2-2	4-2	4-2
3	3-2		4-2	

Final vowel -e

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		*-ẽ*		*-ẽ*
1	*-æ*	*-ĩ*	*-æ*	*-ĩ*
2	*-e*	*-ũ*	*-e*	*-ũ*
3	*-e*		*-e*	

Stem-formative 11

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		*ča-*		*ča-*
1	*ba-*	*ča-*	*kua-*	*ča-*
2	*ča-*	*ča-*	*ča-*	*ča-*
3	*ba-*		*kua-*	

Implicational relations

Each of these separate inflectional systems show considerable complexity

Language	Cells	Realizations	Max realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy
Neutral tones	6	16	4	6	2.585	1.622	0.264
Final vowel	6	11	9	10	3.322	1.333	0.775
Stem formative	4	32	16	18	4.170	2.369	0.099

Each lexical item is a member of some conjugation in each of these three systems

There are potentially 6×10×18=1,080 meta-conjugations

Baerman & Corbett report that 109 are attested in Capen (1996)

Implicational relations

Stem formatives, final vowels, and tone patterns appear to be independent systems associated with inflection classes

By Baerman & Corbett's count, most meta-conjugations have only one or two members; the most frequent has 22

Knowing which class a lexeme belongs to in one dimension provides relatively little information about another dimension:

> Expected entropy for choosing a class in a dimension is **2.469 bits**
>
> Expected conditional entropy for choosing a class in a dimension knowing the class in another dimension is **2.154 bits**

Jamieson offers diachronic explanations for the development of this complexity, but how is it maintained?

Implicational relations

Consider a less abstract problem: given the stem formative, final vowel, and tone pattern of a wordform, guess the stem formative, final vowel, and tone pattern for some other wordform

This turns out to be much easier: for the positive neutral forms, the expected entropy is **4.920 bits** but the paradigm entropy is only **0.709 bits**

Every word form provides information about all three dimensions

Jamieson's inflection classes show a high degree of inter-paradigm syncretism, so listing lexemes by class greatly overstates the variation

Compared to Modern Greek, writing a dictionary of Chiquihuitlán Mazatec is significantly harder (E-complexity), but speaking it isn't (I-simplicity)

Organization

1. The Paradigm Cell Filling Problem: Modern Greek
2. Paradigm entropy: A language sample
3. A surprisingly simple case: Chiquihuitlán Mazatec
4. **Testing entropy claims**

Some caveats

Entropy calculations depend on many, many assumptions

- Identification and enumeration of forms
- Frequencies of lexemes and wordforms
- Choice of (sub-)paradigms
- Generalizing from a single, randomly selected form

The numbers should be interpreted with this in mind

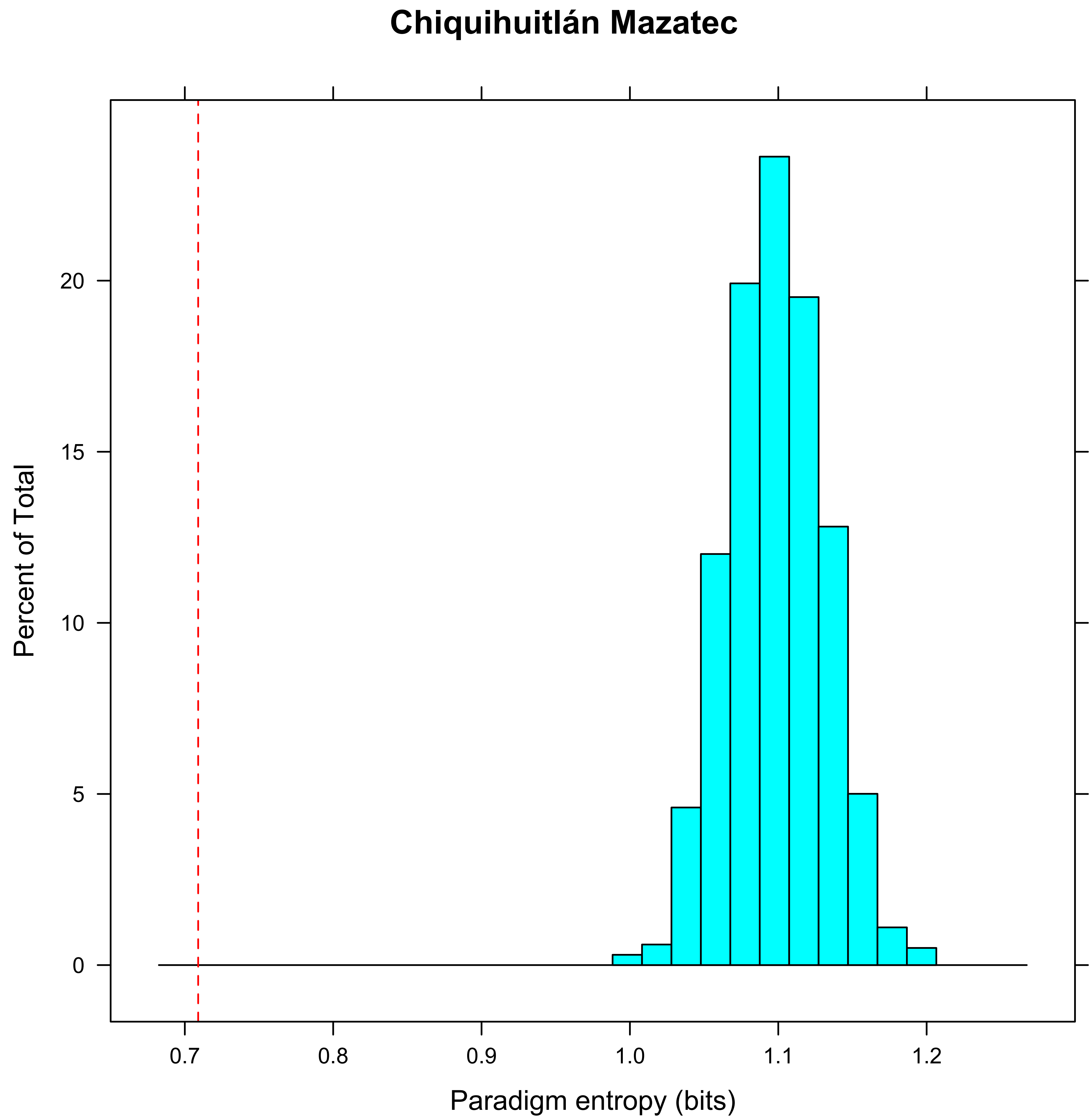
What is clear, however, is that paradigm entropies are much lower than they could be

Testing entropy claims

The implicational structure of the paradigms is crucial to reducing 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*


Chiquihuitlán Mazatec
Percent of Total
0
5
10
15
20
0.7
0.8
0.9
1.0
1.1
1.2
Paradigm entropy (bits)

Language	Cells	Realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy	Bootstap Avg	Bootstrap *p*
Amele	3	31	24	4.585	2.882	1.105	1.327	0.001
Arapesh	2	41	26	4.700	4.071	0.630	0.630	1.000
Burmeso	12	24	2	1.000	1.000	0.000	0.000	1.000
Fur	12	80	19	4.248	2.395	0.517	1.316	0.001
Greek	8	12	8	3.000	1.621	0.644	0.891	0.001
Kwerba	12	26	4	2.000	0.864	0.428	0.523	0.001
Mazatec	6	356	109	6.768	4.920	0.709	1.100	0.001
Ngiti	16	68	10	3.322	1.937	0.484	1.019	0.001
Nuer	6	12	16	4.000	0.864	0.793	0.811	0.160
Russian	12	26	4	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

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 E-complex, but (perhaps) only one basic principle of I-Simplicity, though many ways to get there.

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?