Inferential organization in grammar systems: Word structure, paradigm organization and learning

Farrell Ackerman

University of California at San Diego

(Rob Malouf
San Diego State University)

I

Minél különösebb valami, annál kevésbe rejtélyes -
Sherlock Holmes kalandjai

The big question: How do we demonstrate that the more curious a phenomenon seems, the less mysterious it actually is?

Q1: What ensures the learnability of complex inflectional systems?
H1: (Partly) the inferential organization that emerges and inheres in complex inflectional systems: I(ntegrative) complexity as constrained by the Low Entropy Conjecture.

Contributory questions:

Q2: What is the nature of word internal structure?
Q3: What is the nature of paradigm organization?
Q4: What is the nature of learning as it relates to paradigm organization?

Outline of the presentation

And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.
T. Eliot Little Gidding V

1. The empirical challenge and the problems it raises: Finnish

2. Words and Paradigms as Part/Whole relations: Uralic and Western Nilotic

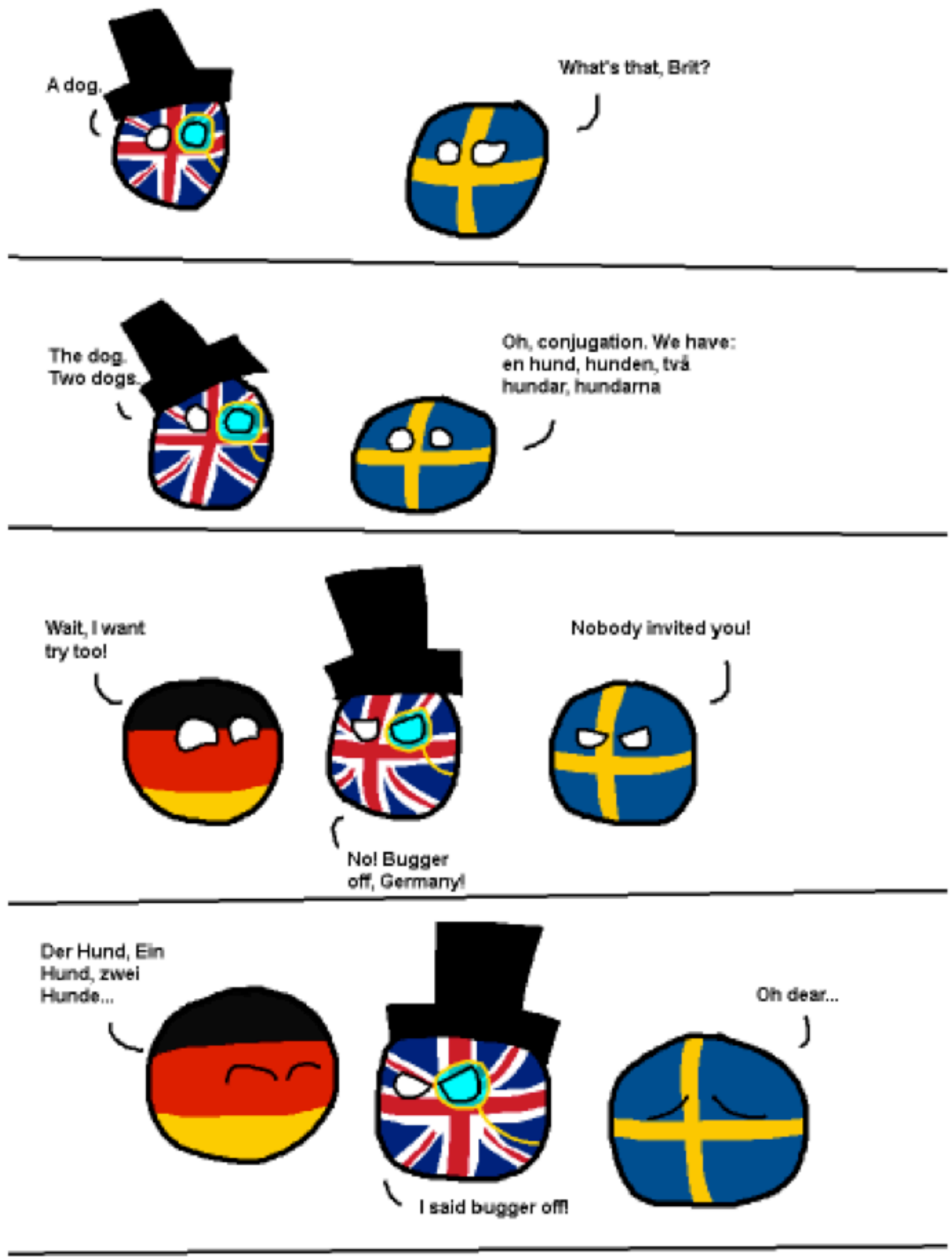
3. Quantifying relations between words: Pite Saami

4. Back to Finnish

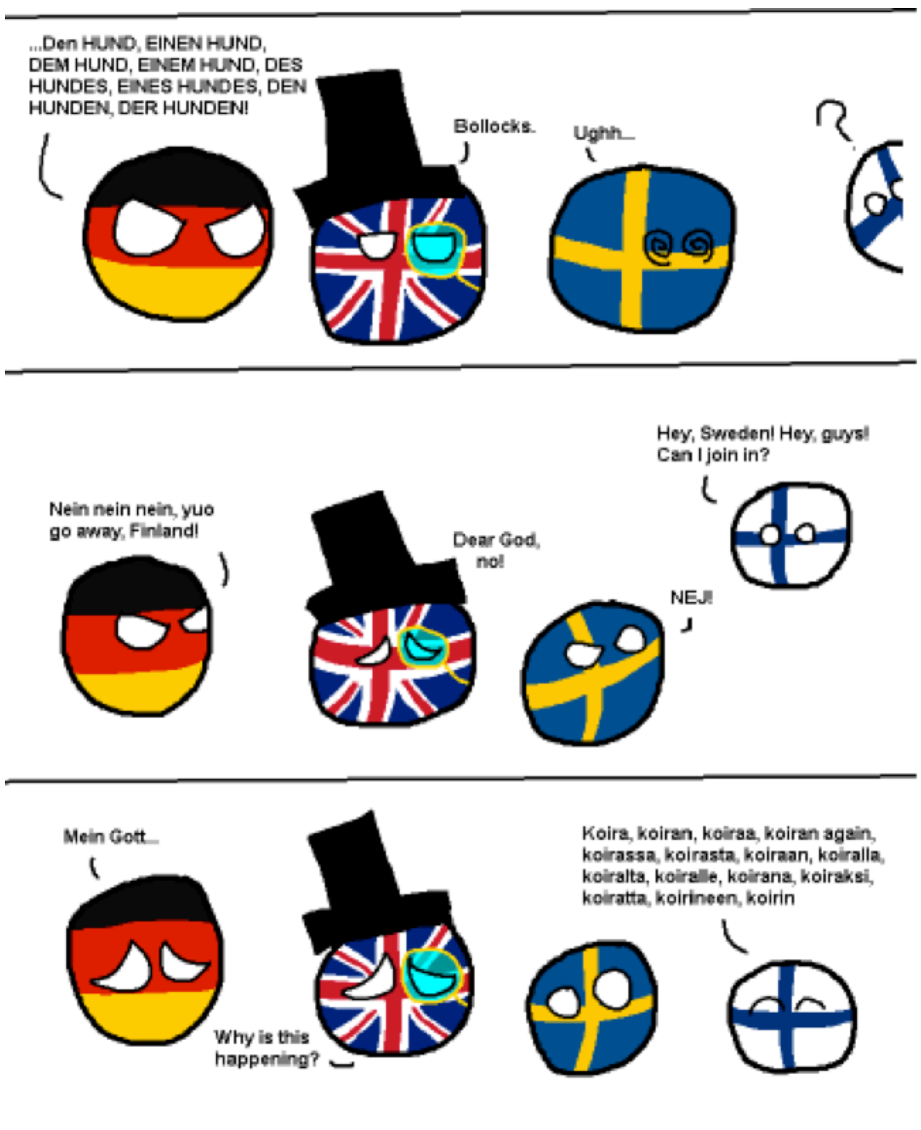
5. Conclusions

1. The empirical challenge and the problems it raises

A fundamental learnability problem



A fundamental learnability problem



6

A fundamental learnability problem

The basic challenge

Q1: How can a Finnish child reliably predict e.g., the singular translative 1sg person possessor **koirakseni `turn into my dog'** given knowledge of e.g., the nominative singular **koira** for the LEXEME cat, as well as all correctly produce all other forms of this lexeme?

Q2: How does she extrapolate to the full set(s) of forms for all other LEXEMES?

Alternatively, how do speakers reduce the uncertainty of predicting one wordform from knowledge of another (pattern of) wordforms?

Expressions of despair (see Baerman To Appear on Seri for another big challenge)

Andersen on Dinka (Western Nilotic): (but, see Baerman on Nuer 2012 on related language Nuer)

From the very beginning of linguistic research on Dinka, it has been noted that number inflection of nouns in this language is irregular. Mitterutzner (1866:15) and Beltrame (1880:22–24) stated that **there is no general rule for forming the plural from the singular**, and both authors made observations about the types of phonetic differences existing between the singular form and the plural form of a noun… That number inflection of simple native nouns, such as those … above, is indeed irregular and unpredictable, has recently been established by Ladd et al. (2009). **The plural form cannot be predicted from the singular form, nor can the singular form be predicted from the plural form, and the number inflection may appear to be totally irregular**. 2014: 226

Carstairs-McCarthy on Polish:

If the Plural were taken into account, too, the amount of blurring could not decrease and might even increase, as is suggested by de Bray's gloomy comment (1980, p. 273): **"the student has to learn for each [noun], apart from the Nom. sing., at least the Gen. sing. and Nom. and Gen. p1. as well, and preferably also the Dat. and Loc. sing.** 2000:818

Vajda on Ket (Eurasian Isolate)

The lexical entry of each Ket finite verb therefore contains, in addition to its purely derivational morpheme shapes, a formula that predetermines the configuration of its actant agreement markers. This formula cannot be predicted in the grammar based on any overall set of syntactic functions, semantic roles, or other formal stem features.... This strategy necessitates the unprecedented claim that most inflection-bearing positions in the Ket verb are specified idiosyncratically by the lexical entry of each individual stem. **Though finite verb morphemes are agglutinative on a phonological level, with separator elements often appearing between them, semantically they exhibit a network of extended and multiple exponences that rivals the most fusional of languages**.

9

A simple(minded) answer

A: If each verb LEXEME has a single stem and each distinct morphosyntactic category e.g., case has a unique marker or exponent for each of its values, i.e., NOM, TRANS..., the solution is easy:

Morphotactic schema: [lexical stem$_N$ - **CASE** - **POSS**]

1. koira-**kse-ni**
   dog-**TRANS-1SG.POSS**
   `into my dog'

Reflects Corbett's operative notion of *canonical inflectional encoding*, i.e., the logical extreme in three dimensions of word encoding (Brown et. al. 2013, Corbett 2015)

a) unique morphotactics,
b) unique stem, and
c) unique morphological marker that models e.g. agglutinative systems like Turkish

Canonical inflectional encoding functions as a locus from the simplest encoding pattern between morphosyntactic categories and forms for tracking cross-linguistic deviations from it: it has **no normative (or universal) theoretical status, simply a heuristic taxonomic one and, thereby, differs from many morpheme-based proposals** (see Siddiqi & Harley eds. *Morphological Metatheory* (to appear) for competing views)

10

Problems with the simple answer

Straightforward **alignments of morphological feature CONTENT and FORM** are much rarer than often supposed and theoretically modeled (Matthews 1991, Brown et. al. 2013, Stump 2001, 2013, 2016, Blevins 2016, Harris (To Appear) on *multiple exponence*), Corbett 2015, Ackerman and Bonami (To Appear), among others) **and**,

requisite **segmentations of internal structure into morphemic units and structural constituents** are often more artifactual and indeterminate than recognized. (see Hockett 1987, Morpurgo-Davies 1978, Matthews 1991, Bochner 1993, Anderson 1990, Corbett 2009, 2015, Stump 2016, Blevins 2016, Bonami and Benjamin 2016, mong others).

Both reflect Lounsbury's (1953) *fictive agglutinative analogue* as a model for structuralist morphemic morphology.

Relatedly and worse, languages commonly

(1) have **multiple inflectional/declensional classes,** and

(2) display **Zipfian distributions of wordforms within and acrosss those classes**.

Multiple declension classes in Finnish (Paunonen 1976, Thymé 1993, Thymé et.al. 1994, Ackerman et. al. 2009)

Nom Sg	Gen Sg	Part Sg	Part PI	Iness PI	
ovi	oven	ovea	ovia	ovissa	'door' (8)
kieli	kielen	kielta¨	kieliä	kielissä	'language' (32)
vesi	veden	vetta¨	vesiä	vesissä	'water' (10)
lasi	lasin	lasia	laseja	laseissa	'glass' (4)
nalle	nallen	nallea	nalleja	nalleissa	'teddy' (9)
kirje	kirjeen	kirjettä	kirjeitä	kirjeissä	'letter' (78)

Patterns of form variation define different declensional classes: Compare (8), (32), and (10) with respect to Nom Sg, Gen Sg, and PartSg.

Q: Isn't all this variability worse than useless, and arguably an impediment to learnability?

1. The numbers in the gloss column refer to declension classes as presented in the Soome-eesti sõnaraamat (Finnish-Estonian Dictionary) Kalju Pihel & Arno Pikamäe (eds.) 1999: 758-771 Tallin: Valgus. An early exploration of PCFP as relates to Finnish is found in Pauonen 1976.

12

It gets worse: Zipfian distribution of forms (Zipf 1935, 1949, Kornai 1992, Yang 2004, Chan 2008, Ellis and O'Donnell 2011, Kurumada et. al. 2013, Bonami and Beniamine 2015, Blevins et. al. 2015, Finlay 2015, Lignos and Yang (to Appear), Yang (to appear) among others)

One of the most puzzling facts about human language is also one of the most basic: words occur according to a famously systematic frequency distribution such that there are few very high frequency words that account for most of the tokens in text (e.g. "a", "the", "I", etc.), and many low frequency words (e.g. "accordion", "catamaran", `ravioli". Piantadosi 2015: 1

No surprise that inflectional morphology follows the same distribution:

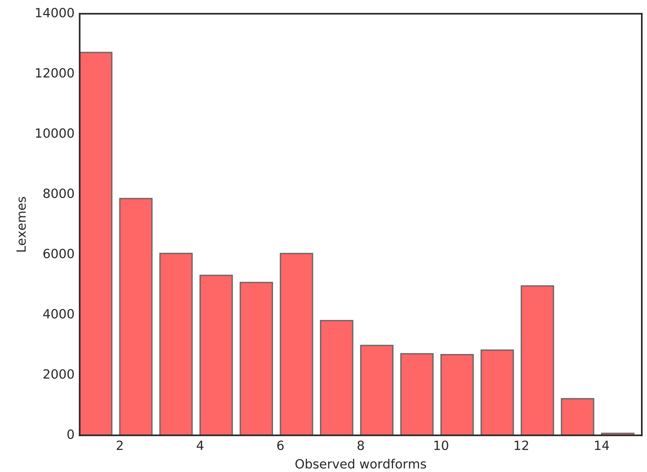
A few lexemes occur frequently, with skewed distributions of their inflected forms, and most lexemes and inflected forms occur rarely, if at all.

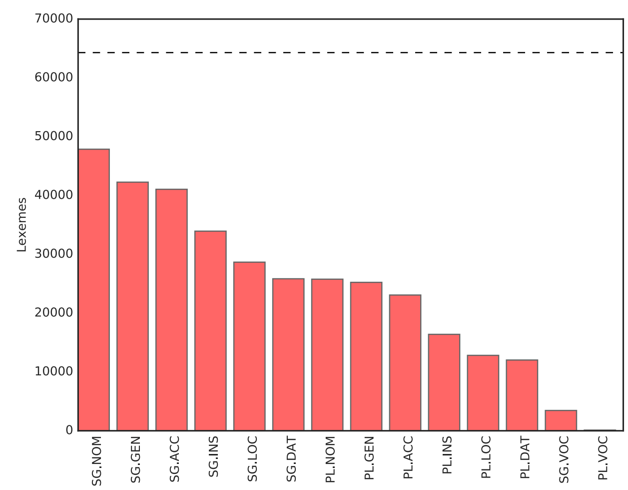
It gets worse: Zipfian distribution of forms (Yang 2004, Chan 2008, Ellis and O'Donnell 2011, Lingnos and Yang (to Appear), Kurumada et. al. 2013, Bonami and Beniamine 2015, Blevins et. al. 2015, Finlay 2015, among others)

Czech National Corpus SYN2010
(Olivier Bonami p.c.)

100 million morphologically tagged words
64,302 distinct noun lexemes
561,668 distinct noun wordforms
900,228 possible wordforms (7 cases, 2 numbers)
Only 66 lexemes occur with full paradigms
No single form is observed for every lexeme
Only 110 lexemes occur in the voc.pl (but more frequent in spoken language, same form as nom.pl)

Larger studies reveal similar distributions in French and Portuguese.





As corpora increase in size, they do not converge on uniformly populated paradigms. Instead, they reinforce previously attested forms and classes while introducing progressively fewer new items. This distribution reflects the fact that inflected variants of open-class items obey Zipf 's law at all observed sample sizes. Blevins et. al. To Appear.

Old recognition and solution for multiple classes and prediction from skewed distributions (Paul 1891, Anttila 1989, Wurzel 1989, Fertig 2013, among others)

Principal parts = $_{\textbf{def}}$ The principal parts of a lexeme L are a set of cells in L's realized paradigm from which one can categorically deduce the remaining cells in L's realized paradigm.

Conjugation	σ_1 = {1sg pres ind active}	σ_2 = {1sg perf ind active}	σ_3 = {first supine}	σ_4 = {pres active inf}
1st	⟨*laudō*, σ_1⟩	⟨*laudāvī*, σ_2⟩	⟨*laudātum*, σ_3⟩	⟨*laudāre*, σ_4⟩
2nd	⟨*moneō*, σ_1⟩	⟨*monuī*, σ_2⟩	⟨*monitum*, σ_3⟩	⟨*monēre*, σ_4⟩
3rd	⟨*dūcō*, σ_1⟩	⟨*dūxī*, σ_2⟩	⟨*dūctum*, σ_3⟩	⟨*dūcere*, σ_4⟩
3rd (-*iō*)	⟨*capiō*, σ_1⟩	⟨*cēpī*, σ_2⟩	⟨*captum*, σ_3⟩	⟨*capere*, σ_4⟩
4th	⟨*audiō*, σ_1⟩	⟨*audīvī*, σ_2⟩	⟨*audītum*, σ_3⟩	⟨*audīre*, σ_4⟩

All 4 Latin declension classes are partitioned into 4 Principal Parts

		Names of the tenses
I. From -*o* are formed,	-*bam*,	Imperf. Indic.
	-*bo*,	Fut Indic. of the 1st and 2d Conjugation.
	-*am*,	Pres. Subj. of the 2d; Pres. Subj. and Fut Indic. of 3d and 4th.
	-*em*,	Pres. Subj. of the 1st.
	-*ns*,	the Present participle.
	-*dus*,	the Fut. Participle, Passive.
	-*dum*, -*di*, -*do*,	the Gerunds.
II. From the -*i* are formed,	-*ram*,	the Plup. Indic.
	-*rim*,	the Perf. Subj.
	-*ro*,	the Fut. Subj. [= future perfect; cf. pp. 83f]
	-*ssem*,	the Plup. Subj.
	-*sse*,	the Perf. Infinit.
III. From -*um* are formed,	-*u*,	the second Supine.
	-*us*,	the Perf. Participle, Passive.
	-*rus*,	the Future Participle.
IV. From the infinitive, whether in -*re*, -*le*, or -*se*, are formed the imperative, by cutting off the final syllable; and the imperfect of the subjunctive, by adding -*m* to it.		

Principal Parts are (categorically) predictive of all inflected forms of a given lexeme.

(Stump and Finkel 2013: 11-13)

15

A hypothesis: Principal Parts in the light of Information Theory

Morphological systems consist of words that are informative about other words (Wurzel 1987 on implicational relations and "paradigm structure constraints")

Informativeness is quantifiable in terms of conditional entropy in Information-Theory: words are informative about other words to a specifiable degree.

The categoricality of predictiveness definitional of Principal Parts is simply a reflex of the possibility that in some systems some words are associated with no unpredictability, i.e. are fully predictive, with respect to their relations to other (forms of) words. (see Ackerman and Malouf 2013, 2015, Bonami and Benjamin (to appear) and Blevin (to appear for discussion)

16

General formulation of the implication problem: The Paradigm Cell Filling Problem (PCFP)

Speakers of languages with complex morphology and multiple inflection classes confront a large learning task whose solution raises fundamental questions about the structure of words, and the organization of morphological systems. This task receives a general formulation as the Paradigm Cell Filling Problem (PCFP) in Ackerman et al. (2009):

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 and derivational) family?

Not much a problem for languages like English, with simple morphology, or Turkish with agglutinative morphology and straightforward mapping between morphosyntactic properties and forms.

A central problem of morphology and learnability, though much of the literature in morphology and learnability focuses on word (and word internal morpheme) segmentation, leaving somewhat mysterious what to do with respect to productivity once many inflected words have been learned. (Saffran et, al. 1996, Fedzechkina et. al. 2011, Finley and Wiemers 2013, 2015, Gerken 2005, Gerken and Knight 2015, Gerken et. al. 2015,Gagliardi and Lidz 2014, Lidz and Gagliardi 2015, and many others.)

Uncertainty reduction

How do speakers reliably resolve/diminish uncertainties in the selection of the appropriate form for a previously unencountered word?

The problem seems increasingly difficult:

(i) the larger the number of morphosyntactic properties a language contains,

(ii) the greater the number of allomorphic variants it uses to encode them, and

(iii) the more extensive the conjugation classes and subclasses, i.e., distinctive patterns, over which words can be distributed.

Morphological complexity is commonly calculated by considering these factors

Ackerman & Malouf (2013) refer to this perspective on morphological typology as **Enumerative Complexity** or **E-complexity**: this is the classification and quantification of morphological phenomena by reference to factors (i)–(iii) above, which all figure in the formal shapes of words. (see Carstairs-Mcarthy's Paradigm Economy Principle and the No Blur Principle for a highly articulated form of this view.)

18

Getting to a solution for PCFP (for solutions within a different theoretical perspective see Allen et. al. , Allen

Hypothesis$_1$: The cross-linguistically profuse shape variation in the internal structure of wordforms and the external patterns of relatedness between wordforms are constrained (in part) by *implicative organization* (= **I(ntegrative) Complexity**) among words as quantified in *low conditional entropy values* (= **Low Conditional Entropy Conjecture (LCEC)**. (Ackerman and Malouf 2013, 2015)

Hypothesis$_2$: The LCEC is one strategy of addressing the learnability of (complex) morphological systems given the sparseness of the data reflected in Zipfian distributions of inflectional wordforms.

Hypothesis$_3$: Given the demonstrable Sparseness of Stimuli and the Richness of the Stimuli (= the curse of dimensionality (Aslin and Newport 2009:17)) morphological systems must be organized in ways that facilitate learnability.

Hypothesis$_4$: To understand morphological learning, one must understand inflectional organization and the LCEC is an emergent constraint shaping (many) morphological systems. (for similar directions see Racz et. al. 2015, Pater and Moreton 2012, Allen and Hanson 2016, Allen 2016, Allen and Becker ms, Pulleyblank and Archangeli (To Appear), Wedel 2007, 2009).

A general solution harking back to traditional WP

Q*:* What should morphologists model?

Hypothesis: Words and their patterns of organized relations (paradigm structure) are the primary objects of morphological analysis.

Hockett (1967) asks: Is a compactly elegant IA analysis of complex morphological compositions in Yawalmani usefully and instructively replaced by the less compact WP?

To cover the complex alternations of Yawelmani by principal-parts-and-paradigms would take much more space than is occupied in the first sections of this paper by the morphophoneme-and-rewrite-rule presentation. But there would be a net gain in realism, for the student of the language would now be required to produce new forms in exactly the way the native user of the language produces or recognizes them-by analogy. There would remain this difference: the situation for the student is artificially simplified. He is enabled to operate, in his analogizing, in terms of a neat minimal set of reference paradigms and a fixed point-of-departure set of principal parts. **The native user of the language, of course, does not do this. He operates in terms of all sorts of internally stored paradigms, many of them doubtless only partial; and he may first encounter a new basic verb in any of its inflected forms. For the native user, the forms that we have for convenience selected to be our 'principal parts' have no such favored position. They are as likely to be created analogically, as needed, as are any of the other forms. Hockett 1967:221**

A general perspective harking back to traditional WP

This concurs with Chomsky's favorable view of WP over IA and his argued dispreference for the "ill-advised" constructs required in *morpheme-based* proposals:

"I know of no compensating advantage for the modern descriptive reanalysis of traditional paradigmatic formulations in term of morpheme sequences. **This [= morphemic analysis - FA] seems, therefore to be an ill-advised theoretical innovation...** It seems that in inflectional systems, the paradigmatic analysis has many advantages and is to be preferred ... **It is difficult to say anything more definite, since there have been so few attempts to give precise and principled description of inflectional systems in a way that would have some bearing on the theoretical issues involved here.**" (Chomsky 1965:174)

But,

a) 51 years have passed.

b) Many paradigmatic approaches have been developed and applied with rigor and precision over wide varieties of phenomena and broad arrays of languages, so

c) it is no longer "difficult to say anything more definite."

…words are not merely wholes made up of parts, but are themselves construable as parts with respect to systems of forms in which they participate.
Matthews (1991:204)

2. Words and Paradigms as Part/Whole relations: Uralic and Western Nilotic

Part/whole relations: The internal structure of words

Internal composition of words: words are commonly the smallest units of meanings (see Robins 1959, Matthews 1991, Blevins 2016).

Cross-linguistic importance of internal structure for morphology is not in the identification of exponents for meaningful bits, i.e., morphemes and often not in constituent structure, but in ways that the organization of exponents facilitate patterns of discriminability that help to distinguish and relate (classes of) words.

Sometimes, this involves (semi-)classic morphemic composition, but more commonly it requires considering words as recombinant gestalts, i.e., wholes consisting of configurations of redeployed elements (segmental, suprasegmental) that each alone do not contribute invariant meanings independent of the word contexts in which they occur.

External patterns of word distribution: relations between words as instructive as instructive about morphological organization.

Part/whole relations$_1$: Internal composition of words

Hungarian (Uralic) nominal morphology:

Stem$_{adj}$-NOMLZ-POSS-CASE Morphotactics

1. bátor- ság- om- ról
   'about my bravery'

Hungarian: (classic) agglutination of morphemic suffixes to lexical stems

24

Part/whole relations$_1$: Internal composition of words

Mari (Uralic) verbal morphology: singular 2nd past paradigms for *kol* ‘die’.

	1st past affirmative		1st past negative	
1	kolə̂-**š̆-ə̂m**	\`I died	**š̆-ə̂m** kolə̂	\`I didn’t die’
2	kolə̂-**š̆-ə̂c**	\`you died‘	**š̆-ə̂c** kolə̂	\`you didn’t die’
3	kolə̂-**š̆**	\`s/he died‘	**ə̂š̆** kolə̂	\`s/he didn’t die’
	Synthetic expression		**Periphrastic expression**	

The same pieces deployed in different morphological configurations convey different polarity values for verbs: they take on different functions in the word context in which they occur.

The organization (gestalt) of the elements is as important as the elements themselves.

Mari is morphologically more E-complex than English, since it has morphological markers for several past tenses and person/number, as well as different paradigms for affirmative and negative polarity.

25

Part/whole relations$_1$: Internal composition of words

Agar dialect of Dinka (Eastern Nilotic) nominal morphology: Andersen (2014) (see also Remijsen 2005, Ladd et. al. 2009, Storch 2005, among others)

Case and number distinguished by word internal interactions among four parameters: (i) vowel length, (ii) tone, (iii) voice quality of the vowel, and (iv) **vowel quality alternation grade.**

SG	PL	
dí̱t	djɛ̱̀ɛt	'bird'
kɔ̱̀ɔɔr	kà̱ar	'elbow'
rjɛ̱̀ɛm	rî̱m	'blood'
cjé̤ec	cí̤c	'bee'
lá̤j	là̤aj	'animal'
mà̱ac	mɛ̤̂ɛc	'fire'
dò̱m	dṳ̂um	'field'
tò̤oɲ	tô̤oɲ	'pot'
t̪wô̤ooɲ	t̪ṳ́uɲ	'ember'
ɰà̤am	ɰɔ̤̀ɔm	'thigh'

Mirror image patterning:

Singular for 'elbow' has triple length for its vowels and low tone while this is the word internal pattern for the plural of 'thigh'.

The double length vowels and low tone for the singular of 'thigh' parallels the same pattern for the plural of 'elbow'.

Contrasts between (pairs of) words disclose the patterned nature of morphological organization.

Dinka is morphologically more E-complex than English, since it has many more strategies for expressing number contrasts.

26

Shilluk (Western Nilotic) (Remijsen & Ayoker 2014)

Shilluk presents a rich system of morphological marking with a small segmental footprint. Morphological exponence is characterized not by the concatenation of discrete morphemes, but rather by a stacking of morphological operations within a confined domain, consisting of the stem and a limited window of affixes. In such a system, morphological marking is restricted by the fact that the stem syllable can have only one specification each for tone, length, vowel quality, ATR, and the stem-final consonant, and these specifications need to convey both lexical and morphological information. One way in which the use of these resources is maximized is through distributed exponence, whereby a morphophonological pattern realizes two morphosyntactic values. Remijen et. al. To Appear: 17

a. **càak á-mấat̪**
milk:PL PST-drink
'Somebody drank milk.'

b. **càak á-mát̪-ì**
milk:PL PST-drink-ITER
'Somebody drank milk repeatedly.'

c. **lwɔ̂ɔl á-mấaat̪ càak**
cup:SG PST-drink:APPL milk:PL
'Somebody drank milk using a gourd.'

d. **lwɔ̂ɔl á-mát̪-ì càak**
cup:SG PST-drink-ITER milk:PL
'Somebody drank milk repeatedly using a gourd.'

Some Shilluk verb classes (Remijsen & Ayoker 2014)

Remijsen et. al. To Appear: 17

Verb classes	Fixed Short		Short with Grade		Long		
	Low	Fall	Low	Fall	Low	Fall	High Fall
Example	|ŋɔl|	|lɛŋ|	|cam|	|mʌl|	|keel|	|lʊʊɲ|	|lʊʊɲ|
	'cut'	'drum'	'eat'	'roast'	'spear'	'pluck'	'take turns'
PST	á-ŋɔ̂́l	á-lɛ̂́ŋ	á-cấm	á-mʌ̂́l	á-kếel	á-lʊ̂́ʊɲ	á-lʊ̂́ʊɲ
PST 2SG	á-ŋɔ̀l	á-lɛ̂ŋ	á-càaam	á-mʌ̂ʌʌl	á-kèeel	á-lʊ̂ʊʊɲ	á-lʊ̀ʊʊɲ
PST APPL	á-ŋɔ̄l	á-lɛ̂́ŋ̄	á-cāaam	á-mʌ̂́ʌʌl̄	á-kēeel	á-lʊ̂́ʊʊɲ̄	á-lʊ̄ʊʊɲ
PST APPL 2SG	á-ŋɔ̌l	á-lɛ́ŋ	á-cǎaam	á-mʌ́ʌʌl	á-kěeel	á-lʊ́ʊʊɲ	á-lʊ̌ʊʊɲ
SPAT/FUG	á-ŋɔ́l̀	á-lɛ́ŋ̀	á-cáaam̀	á-mʌ́ʌʌl̀	á-kéeel̀	á-lʊ́ʊʊɲ̀	á-lʊ̂ʊʊɲ

28

Part/whole relations$_1$: Internal composition of words

Nom Sg	kukk	lukk	pukk	sukk
Part(itive) Sg	**kukke**	**lukku**	**pukki**	**sukka**
Gen Sg	kuke	luku	puki	suka
Illa(tive)2 Sg	kukke	lukku	pukki	sukka
	'rooster'	'lock'	'trestle'	'stocking'

Distinctive forms

First declension partitives in Estonian (Blevins 2016:83)

All of the forms are discriminably distinct, exhibiting the same patterns of distinctiveness, including patterns of syncretism for Part Sg and Ill Sg.

Strong Stem	kukk	lukk	pukk	sukk
Weak Stem	kuk	luk	puk	suk
Theme Vowel	e	u	i	a

First declension stems and theme vowels

There are no pieces of form that are uniquely associated with morphosyntactic case and number properties, but there is a set of phonological resources that get reused to create distinctive wordshapes.

29

Part/whole relations$_1$: Summary

Internal structure: discriminative function, not a (necessarily) compositional function: distinguishing words from one another or via syncretisms identifying patterns of identity.

The word is frequently the smallest meaningful unit and the patterns that characterize and distinguish (types of) words makes words discriminable from one another:

Words are made up of reused pieces reconfigured to serve different purposes and to facilitate discriminability among words: Gurevich (2006:44) on Georgian

The meaning of the whole word licenses the exponents to be used, but there is no precondition that the meanings of the exponents have to combine to comprise the meaning of the whole. **Compositionality may, indeed, emerge, but as a side product rather than a central principle, or perhaps as an effective learning strategy. The whole itself may contribute meaning to the meanings of the parts, or may override the meanings of the parts.**

The adoption of words as independent and necessary units of analysis also permits words, in turn, to be parts of larger systems.

Consequence of permitting words to be contrasted with words: the possibility of discovering morphological organization (paradigmatic systems or information niches) in the systems of relations between words.

30

Part/Whole relations$_2$: External patterns of word distribution

Blevins 2006: 549

		NONE		QUANTITATIVE				QUALITATIVE	
	GRADE	SING	PLUR	SING	PLUR	SING	PLUR	SING	PLUR
GRAMMATICAL	NOMINATIVE	*pesa*	**pesad**	*`kool*	**koolid**	*`kukk*	**kuked**	*pidu*	**`peod**
	GENITIVE	**pesa**	*pesade*	**kooli**	*`koolide*	**kuke**	*`kukkede*	**`peo**	*pidude*
	PARTITIVE	*pesa*	*pesasid*	*`kooli*	*`koolisid*	*`kukke*	*`kukkesid*	*pidu*	*pidusid*
	STEM PARTITIVE		*pesi*		*`koole*		*`kukki*		—
	SHORT ILLATIVE	*`pessa*		*`kooli*		*`kukke*		*`pittu*	
SEMANTIC	ILLATIVE	**pesasse**	*pesadesse*	**koolisse**	*`koolidesse*	**kukesse**	*`kukkedesse*	**`peosse**	*pidudesse*
	INESSIVE	**pesas**	*pesades*	**koolis**	*`koolides*	**kukes**	*`kukkedes*	**`peos**	*pidudes*
	ELATIVE	**pesast**	*pesadest*	**koolist**	*`koolidest*	**kukest**	*`kukkedest*	**`peost**	*pidudest*
	ALLATIVE	**pesale**	*pesadele*	**koolile**	*`koolidele*	**kukele**	*`kukkedele*	**`peole**	*pidudele*
	ADESSIVE	**pesal**	*pesadel*	**koolil**	*`koolidel*	**kukel**	*`kukkedel*	**`peol**	*pidudel*
	ABLATIVE	**pesalt**	*pesadelt*	**koolilt**	*`koolidelt*	**kukelt**	*`kukkedelt*	**`peolt**	*pidudelt*
	TRANSLATIVE	**pesaks**	*pesadeks*	**kooliks**	*`koolideks*	**kukeks**	*`kukkedeks*	**`peoks**	*pidudeks*
	TERMINATIVE	**pesani**	*pesadeni*	**koolini**	*`koolideni*	**kukeni**	*`kukkedeni*	**`peoni**	*pidudeni*
	ESSIVE	**pesana**	*pesadena*	**koolina**	*`koolidena*	**kukena**	*`kukkedena*	**`peona**	*pidudena*
	ABESSIVE	**pesata**	*pesadeta*	**koolita**	*`koolideta*	**kuketa**	*`kukkedeta*	**`peota**	*pidudeta*
	COMITATIVE	**pesaga**	*pesadega*	**kooliga**	*`koolidega*	**kukega**	*`kukkedega*	**`peoga**	*pidudega*
		‘nest’		‘school’		‘rooster’		‘party’	

Blevins 2006: 551

PARTITIVE SG
→ NOMINATIVE SG
↔ PARTITIVE PL
↔ GENITIVE PL
↔ SEMANTIC PL

GENITIVE SG
↔ NOMINATIVE PL
↔ SEMANTIC SG

Pervasive patterns of implicative interdependencies

Cool data, but so what?

Pesetsky (2009:464) expresses appropriate skepticism about similar “cabinets of curiosities” identified in Evans and Levinson’s (2009) against language universals, since the importance of unusual variants of familiar constructions w/o analysis is hard to evaluate.

So, what sort of analysis can be developed?

32

3. Quantifying relations between words: Pite Saami

33

Pite Saami

All natural languages show a certain degree of what Baerman et al. (2010:2) call "gratuitous" morphological complexity and Wurzel (1986:76) describes as "ballast" in the linguistic system.

Pite Saami (Wilbur 2014: 109)

class		NOM.SG	NOM.PL	ACC.SG	GEN.PL	ILL.SG	ELAT.SG	
I	a	*luakkt-a*	*luokt-a*	*luokt-a-v*	*luokt-a-j*	*luakkt-a-j*	*luokt-a-st*	'bay'
	b	*mánn-á*	*mán-á*	*mán-á-v*	*mán-á-j*	*mánn-á-j*	*mán-á-st*	'child'
	c	*bäbbm-o*	*biebm-o*	*biebm-o-v*	*biebm-o-j*	*bäbbm-o-j*	*biebm-o-st*	'food'
	d	*skåvvl-å*	*skåvl-å*	*skåvl-å-v*	*skåvl-å-j*	*skåvvl-å-j*	*skåvl-å-st*	'school'
	e	*guoll-e*	*guol-e*	*guol-e-v*	*gul-i-j*	*guoll-á-j*	*guol-e-st*	'fish'
		vágg-e	*vágg-e*	*vágg-e-v*	*vägg-i-j*	*vágg-á-j*	*vágg-e-st*	'valley'
		sábm-e	*sám-e*	*sám-e-v*	*säm-i-j*	*sábm-á-j*	*sám-e-st*	'Saami'
II		*båts-oj*	*buhts-u*	*buhts-u-v*	*buhts-u-j*	*buhts-u-j*	*buhts-u-st*	'reindeer'
		ålm-aj	*ålm-a*	*ålm-a-v*	*ålm-a-j*	*ålm-a-j*	*ålm-a-st*	'man'
III	a	*sabek*	*sabeg-a*	*sabeg-a-v*	*sabeg-i-j*	*sabeg-i-j*	*sabeg-i-st*	'ski'
		vanás	*vadnás-a*	*vadnás-a-v*	*vadnás-i-j*	*vadnás-i-j*	*vadnás-i-st*	'boat'
	b	*bena*	*bednag-a*	*bednag-a-v*	*bednag-i-j*	*bednag-i-j*	*bednag-i-st*	'dog'
		gáma	*gábmag-a*	*gábmag-a-v*	*gábmag-i-j*	*gábmag-i-j*	*gábmag-i-st*	'shoe'

Pite Saami has eight nominal declensions showing distinct grade and suffix patterns.

Since the assignment of lexical items to particular declensions
is largely arbitrary (though influenced by phonological factors), these classes add complexity to the inflectional system in a way that serves no (evident) communicative purpose.

34

Extracted pattern of form variation distribution

Pite Saami (Wilbur 2014: 102)

	SINGULAR	PLURAL
NOM	*bärrgo*	*biergo*
GEN	*biergo*	*biergoj*
ACC	*biergov*	*biergojd*
ILL	*bärrgoj*	*biergojda*
INESS	*biergon*	*biergojn*
ELAT	*biergost*	*biergojst*
COM	*biergojn*	*biergo*
ABESS	*biergodak*	*biergodahta*
ESS	*bärrgon*	

	SINGULAR	PLURAL
NOM	ä+str	
GEN		
ACC		
ILL	ä+str	ie+wk
INESS		
ELAT		
COM		
ABESS		
ESS	ä+str	

Pattern of stem and affix variation (stems partition the inflectional feature space (Bonami and Boyé 2006, 2007))

35

Information-Theoretic insights

Q: What potentiates the learnability of the system?

A: Systemic relations among words organized in terms of low conditional entropies, i.e., the predictability of an unknown form, given knowledge of another form of that word.

H: Quantify "prediction" or as a reduction in uncertainty, or information entropy (Shannon 1948)

Complex inflectional systems are organized in ways that facilitate the reliable guessing from known forms to unknown forms and this organization can be made visible by using Information-Theoretic measures of **Entropy** and **Conditional Entropy**.

What we need to know:

1) The **Paradigm Cell Entropy** for each cell i.e., how the choices of exponents for each cell can be calculated in terms of the degree to which a choice between them is uncertain.

36

Information-Theoretic insights[3]

(2) the **Conditional Entropy** between pairs of cells, so that we can determine how much the knowledge of the form in one cell reduces the uncertainty associated with selecting a paired cell, $H(Y|X)$

How much information does the presence of X have for predicting Y, or how surprised are we that given X we get Y?

What's important is *surprisal:* measure of the amount of information expressed by a particular outcome measured in bits, where 1 bit is a choice between 2 equiprobable outcomes.[3]

The intuition: Outcomes which are less probable (harder to predict and more uncertain) have higher surprisal,

Surprisal is 0 bits for outcomes which always occur ($p(x) = 1$) and approaches ∞ for very unlikely events (as $p(x)$ approaches 0.

3) The **Average Conditional Entropy** as the averaged sum of all of the pairwise conditional entropies provides us with a general "interpredictablity between forms" measure.

3. For present purposes we will assume all of the tasks below reflect equiprobable distributions. This actually provides the upper bound of entropy values with frequency information and/or semantic and phonological conditions likely reducing entropy (see references for explorations along these lines)

37

Calculating Paradigm Cell Entropy

class		NOM.SG	NOM.PL	ACC.SG	GEN.PL	ILL.SG	ELAT.SG	
I	a	*luakkt-a*	*luokt-a*	*luokt-a-v*	*luokt-a-j*	*luakkt-a-j*	*luokt-a-st*	'bay'
	b	*mánn-á*	*mán-á*	*mán-á-v*	*mán-á-j*	*mánn-á-j*	*mán-á-st*	'child'
	c	*bäbbm-o*	*biebm-o*	*biebm-o-v*	*biebm-o-j*	*bäbbm-o-j*	*biebm-o-st*	'food'
	d	*skåvvl-å*	*skåvl-å*	*skåvl-å-v*	*skåvl-å-j*	*skåvvl-å-j*	*skåvl-å-st*	'school'
	e	*guoll-e*	*guol-e*	*guol-e-v*	*gul-i-j*	*guoll-á-j*	*guol-e-st*	'fish'
		vágg-e	*vágg-e*	*vágg-e-v*	*vägg-i-j*	*vágg-á-j*	*vágg-e-st*	'valley'
		sábm-e	*sám-e*	*sám-e-v*	*säm-i-j*	*sábm-á-j*	*sám-e-st*	'Saami'
II		*båts-oj*	*buhts-u*	*buhts-u-v*	*buhts-u-j*	*buhts-u-j*	*buhts-u-st*	'reindeer'
		ålm-aj	*ålm-a*	*ålm-a-v*	*ålm-a-j*	*ålm-a-j*	*ålm-a-st*	'man'
III	a	*sabek*	*sabeg-a*	*sabeg-a-v*	*sabeg-i-j*	*sabeg-i-j*	*sabeg-i-st*	'ski'
		vanás	*vadnás-a*	*vadnás-a-v*	*vadnás-i-j*	*vadnás-i-j*	*vadnás-i-st*	'boat'
	b	*bena*	*bednag-a*	*bednag-a-v*	*bednag-i-j*	*bednag-i-j*	*bednag-i-st*	'dog'
		gáma	*gábmag-a*	*gábmag-a-v*	*gábmag-i-j*	*gábmag-i-j*	*gábmag-i-st*	'shoe'

Cells can be realized by different numbers of allomorphs: Illative singular has 5 allomorphs.

Given equiprobability assumptions, the more the number of choices, the higher the entropy. (for more realistic assumptions see Ackerman et. al. 2009, Bonami 2015, Sims 2016, among others)

Each member of a realization set is associated with an entropy value reflecting the degree of surprise of its selection.

38

An (un)realistic task and its entropies: Paradigm Cell Entropy

A speaker's task: Guess the wordform for a specific cell.

This will depend on the number of possible realizations for each cell:

Illative singular only has five possible realizations and an entropy of 2.250 bits, while the most diverse cells have an entropy at 3.00 bits (= 8 choices)

NOM.SG	GEN.SG	ACC.SG	ILL.SG	INESS.SG	ELAT.SG	COM.SG
3.000	2.406	2.406	2.250	2.750	2.750	2.750

NOM.PL	GEN.PL	ACC.PL	ILL.PL	INESS.PL	ELAT.PL	COM.PL
2.406	2.750	2.750	2.750	2.750	2.750	2.750

Average entropy across all cells is 2.66; this average is a measure of how difficult it is for a speaker to guess the realization of any one wordform of any particular lexeme in the absence of any information about that lexeme's declension = **Paradigm cell entropy**

An entropy of 2.66 bits is equivalent to selecting among only $2^{2.26}$ = 6.31 equally likely alternatives:

Pite Saami has eight declensions, but selecting the realization for a particular wordform of a lexeme is as difficult as a choice among a little more than six equally likely alternatives.

A realistic speaker's task: Conditional Entropy

Quantifying the predictability of one form given the other: measure the size of the surprise associated with these forms using conditional entropy H (Y|X), the uncertainty in the value of Y given that we already know the value of X:

$$
\begin{aligned}
H(Y|X) &= H(X,Y) - H(X) \\
&= \sum_{x \in X} \sum_{y \in Y} P(x,y) \log_2 P(y|x)
\end{aligned}
$$

It measures how much entropy remains for a given form in a given cell if a form in another cell is already known.

The point: The smaller H(Y|X) is, the more predictable Y is on the basis of X, i.e., the less surprised one is that Y is selected given knowledge of X.

Where X completely determines Y, the conditional entropy X(Y|X) is 0 bits: given the value of X, there is no question remaining as to what the value of Y is. (Think of Principal Parts)

But, if X gives us no information about Y at all, the conditional entropy X(Y|X) is equal to H(Y): given the value of X, we are just as uncertain about the value of Y as we would be without knowing X at all.

40

Average Conditional Entropies in Pite Saami

	NOM.SG	GEN.SG	ACC.SG	ILL.SG	INESS.SG	ELAT.SG	COM.SG
NOM.SG	—	0.000	0.000	0.000	0.000	0.000	0.000
GEN.SG	0.594	—	0.000	0.344	0.344	0.344	0.344
ACC.SG	0.594	0.000	—	0.344	0.344	0.344	0.344
ILL.SG	0.750	0.500	0.500	—	0.500	0.500	0.500
INESS.SG	0.250	0.000	0.000	0.000	—	0.000	0.000
ELAT.SG	0.250	0.000	0.000	0.000	0.000	—	0.000
COM.SG	0.250	0.000	0.000	0.000	0.000	0.000	—
NOM.PL	0.594	0.000	0.000	0.344	0.344	0.344	0.344
GEN.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000
ACC.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000
ILL.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000
INESS.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000
ELAT.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000
COM.PL	0.250	0.000	0.000	0.000	0.000	0.000	0.000

	NOM.PL	GEN.PL	ACC.PL	ILL.PL	INESS.PL	ELAT.PL	COM.PL
NOM.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GEN.SG	0.000	0.344	0.344	0.344	0.344	0.344	0.344
ACC.SG	0.000	0.344	0.344	0.344	0.344	0.344	0.344
ILL.SG	0.500	0.500	0.500	0.500	0.500	0.500	0.500
INESS.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ELAT.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
COM.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOM.PL	—	0.344	0.344	0.344	0.344	0.344	0.344
GEN.PL	0.000	—	0.000	0.000	0.000	0.000	0.000
ACC.PL	0.000	0.000	—	0.000	0.000	0.000	0.000
ILL.PL	0.000	0.000	0.000	—	0.000	0.000	0.000
INESS.PL	0.000	0.000	0.000	0.000	—	0.000	0.000
ELAT.PL	0.000	0.000	0.000	0.000	0.000	—	0.000
COM.PL	0.000	0.000	0.000	0.000	0.000	0.000	—

Table 2: Conditional entropies $H(\text{col}|\text{row})$ for Pite Saami noun paradigms in Table 1

For Pite Saami example, H(P) is 0.116 bits, equivalent to a choice among only $2^{0.116} = 1.08$ equally likely declensions.

While Pite Saami has eight nominal declensions from the point of view of a lexicographer trying to describe the language, for a speaker trying to use the system it has on average only slightly more than one: this is the I(ntegrative)-complexity of this paradigm.

Provides a quantitative measure of the descriptions and insights about paradigm structure (Wurzel 1989, Matthews 1991) and a way to calculate I-complexity of system.

41

Identifying degrees of informativity

NOM.SG	GEN.SG	ACC.SG	ILL.SG	INESS.SG	ELAT.SG	COM.SG
0.368	0.038	0.038	0.079	0.118	0.118	0.118
NOM.PL	**GEN.PL**	**ACC.PL**	**ILL.PL**	**INESS.PL**	**ELAT.PL**	**COM.PL**
0.038	0.118	0.118	0.118	0.118	0.118	0.118

NOM.SG	GEN.SG	ACC.SG	ILL.SG	INESS.SG	ELAT.SG	COM.SG
0.000	0.311	0.311	0.519	0.019	0.019	0.019
NOM.PL	**GEN.PL**	**ACC.PL**	**ILL.PL**	**INESS.PL**	**ELAT.PL**	**COM.PL**
0.311	0.019	0.019	0.019	0.019	0.019	0.019

Columns averages E[col] are a measure of **predictedness**, or how difficult it is to guess the realization of a cell (on average) given knowledge of some other cell

Row averages indicate a cell's **predictiveness**: the average uncertainty in another paradigm cell given knowledge of that cell.

The nominative singular is very predictive but harder to predict: on its basis all other forms are completely predictable, making it a principal part in the classical sense.

I-complexity generalizations across languages?[2]

Language	Total Cells	Total Real	Max Real	Decl Class	Decl	Entropy Paradigm	Avg Cond
Amele	3	30	14	24	4.585	2.882	1.105
Arapesh	2	41	26	26	4.700	4.071	0.630
Burmeso	12	6	2	2	1.000	1.000	0.000
Fur	12	50	10	19	4.248	2.395	0.517
Greek	8	12	5	8	3.000	1.621	0.644
Kwerba	12	9	4	4	2.000	0.864	0.428
Mazatec	6	356	94	109	6.768	4.920	0.709
Ngiti	16	7	5	10	3.322	1.937	0.484
Nuer	6	12	3	16	4.000	0.864	0.793
Russian	12	14	3	4	2.000	0.911	0.538

The Low Entropy Conjecture is the prediction that languages will show low average conditional entropy along the lines discovered for these languages.

Ackerman & Malouf (2013) refer to this measure as calculating the **I**(-ntegrative)-**complexity** of a morphological system: it is measure of transparency of the relations of patterns of words in paradigm.

2. Additional confirming empirical results: Pite Saami (Ackerman and Malouf above), Estonian (Blevins, Baerman), Tundra Nenets (Ackerman et al.), Tlingit (Cable), Murrinhpatha (Mansfield and Nordlinger), Portuguese (Bonami and Luis), French (Bonami and Beniamine), Nuer (Baerman), Voro (Baerman), Palantla Chinantec and Kadiweu (Sims and Parker).

43

Observations

The LCEC reflects a strategy for how complex morphological systems are organized in a way to insure learnability, despite luxuriant E-complexity (many different forms, many different classes) and skewed Zipfian distribution of sparse stimuli.

Represents the discovery of a robust principle, a statistical language universal, of cross-linguistic morphological organization that only becomes clear when the internal structure of words and their external distributions are recognized as primary theoretical objects of analysis - words are not epiphenomena.

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 small sample of known forms.

This is the **Integrative Complexity** of a system: the relative informativity associated with each form and how this defines language particular patterns of interpredictability between forms.

An important observation concerning word internal structure

Results do not depend on (universal) assumptions about particular formal properties of words, only on whether a speaker can discriminate between two forms:

specific form type, either synthetic or periphrastic, and the manner of discrimination (whether by affixes, tones, stress, ablaut, or ensembles of these) is irrelevant, as long as for the forms are discriminable from one another.

An important observation concerning word internal structure

A (hypothetical) straightforwardly agglutinative language has an average conditional entropy of 0 bits, as expected:

CLASS	NOM.SG	NOM.PL	ACC.SG	ACC.PL
I	-a	-am	-aj	-ajm
II	-o	-om	-oj	-ojm

as does a hypothetical fusional language.

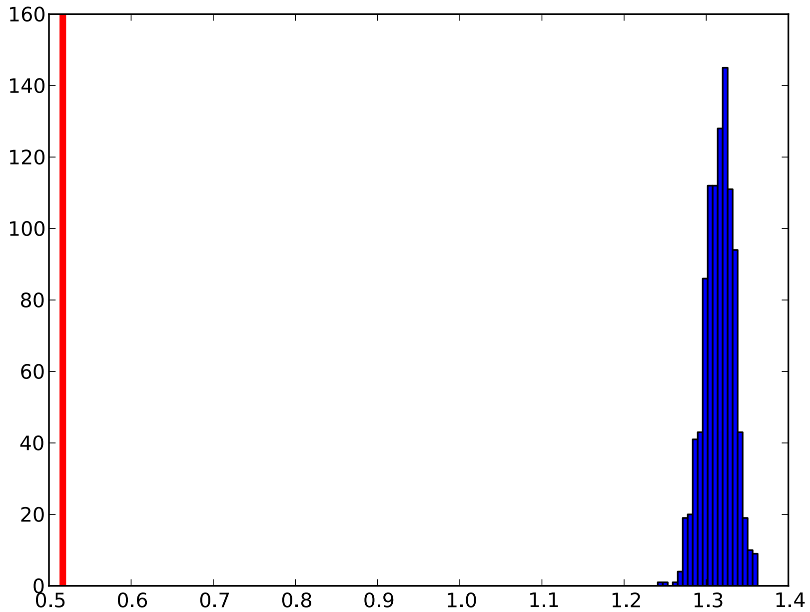
CLASS	NOM.SG	NOM.PL	ACC.SG	ACC.PL
I	-am	-ij	-im	-ux
II	-it	-os	-un	-ad

Organized or accidental mappings? Monte Carlo Simulations

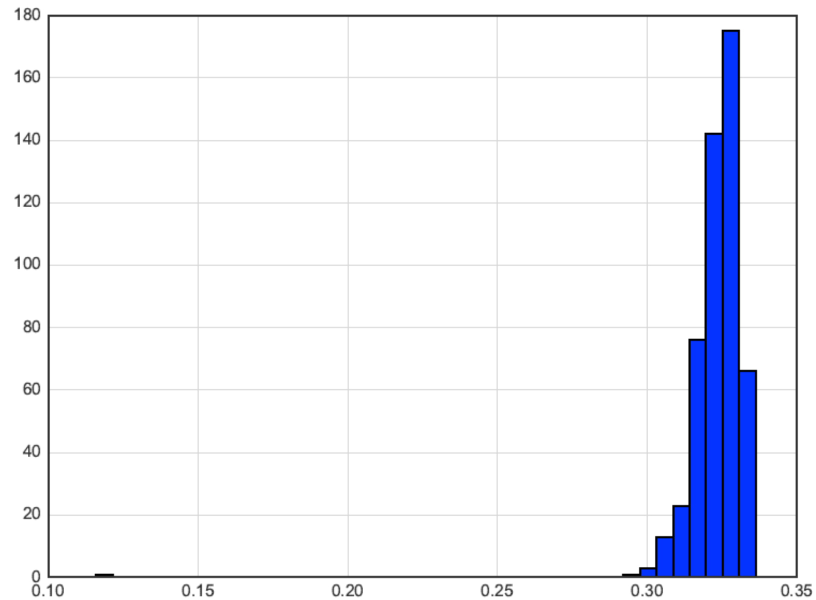
Simulations compare the entropies associated with actual

Language	Declensions	Cells	Realizations	Paradigm entropy	Bootstap Avg	Bootstrap *p*
Fur	19	12	80	0.517	1.316	0.001
Pite-Saami	8	14	70	0.116	0.322	0.001
Russian	4	12	26	0.538	0.541	0.383

Fur



Pite-Saami



Results based on uniform type frequencies (Malouf and Ackerman 2010, 2013, 2015)

4. Back to Finnish

Finnish implicative organization (Ackerman et. al. 2009)

Class	NOM.SG	GEN.SG	PART.SG	PART.PL	INESS.PL	
4	*lasi*	*lasin*	*lasia*	*laseja*	*laseissa*	'glass'
8	*ovi*	*oven*	*ovea*	*ovia*	*ovissa*	'door'
9	*nalle*	*nallen*	*nallea*	***nalleja***	*nalleissa*	'teddy'
32	*kuusi*	*kuusen*	*kuusta*	*kuusia*	*kuusissa*	'six'
10	*kuusi*	*kuuden*	*kuutta*	*kuusia*	*kuusissa*	'spruce'

Given stimulus *tuohta* 'birchbark (part.sg)', there is correct assignment to class 32 based on the analogical proportion **kuusta : tuohta :: kuusi : TUOHI** .

If the stimulus is a non-diagnostic wordform, correct class assignment is underdetermined. Stimulus *nuken* 'puppet (gen.sg)' could be assigned either to class 9 or class 8, based on the competing analogical proportions **nallen : nuken :: nalle : NUKKE** versus **oven : nuken :: ovi : NUKKI**.

If the stimuli comprise the pair *nuken* 'puppet (gen.sg)' and *nukkeja* 'puppet (part.pl)', then there correct assignment of this word to class 9. (= joint entropy (Bonami and Beniamine 2015,

1. The numbers in the Class column refer to declension classes as presented in the Soome-eesti sõnaraamat (Finnish-Estonian Dictionary) Kalju Pihel & Arno Pikamäe (eds.) 1999. Tallin: Valgus.

Finnish implicative organization (Ackerman et. al. 2009)

Most Predicted

	Nom Sg	Gen Sg	Part Sg	Part Pl	Ines Pl	E[row]
Nom Sg	—	1.333	1.667	0.874	0.541	1.104
Gen Sg	0.459	—	0.459	0.459	0.459	0.459
Part Sg	0.333	0.000	—	0.333	0.333	0.250
Part Pl	0.333	0.792	1.126	—	0.000	0.563
Ines Pl	0.459	1.252	1.585	0.459	—	0.939
E[col]	0.396	0.844	1.209	0.531	0.333	0.663

Most Predictive (Part Sg row) — Most Predictive (E[row] 0.250)

Most Predicted

50

5. Conclusions

Current directions

(1) identify larger data sets supplemented with frequency information to serve as objects of measurement to replace the measurement of forms derived from descriptive grammars, as well as developing appropriate tools for their measurement (Bonami 2014, Boye ms., Sims 2015, Bonami & Beniamine 2015, Baerman on Seri (to appear), among others)

(2) more carefully explore the nature of the phonological/phonetic stimuli constitutive of word internal structure and, more generally, inquire about the appropriate forms that words as objects of analysis should take = morphophonetics (Lehiste 1972, Kemps et al. 2005, Blazej & Cohen-Goldberg 2015, Seyfarth et al. 2015, Plag et. al. 2015, among others)

(3) explore how analogical inference may rely upon implicative organization in the learning of complex morphological systems (Baayen & Ramscar 2015, Ramscar et al. 2015), and

(4) identify cross-linguistic generalizations concerning possible constraints on the organization of morphological systems (Stump & Finkel 2013, Baerman et al. 2015, Ackerman & Malouf 2015, Sims & Parker to appear).

52

And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.
T. Eliot Little Gidding V

The big question:

Q1: What ensures the learnability of complex inflectional systems?
A1: (Partly) the inferential organization that emerges and inheres in complex inflectional systems: I-complexity as constrained by the Low Entropy Conjecture.

Contributory questions:

Q2: What is the nature of word internal structure?
A1: Discriminative
Q3: What is the nature of paradigm organization?
A3: Implicative
Q4: What is the nature of learning as it relates to paradigm organization?
A4: Analogical (implicit in today's presentation, but see pattern generalization literature on statistical learning)

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 *ba*^{3}*se*^{2} ‘remember’

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		*ča*^{2}*sẽ*^{2}		*ča*^{2}*sẽ*^{42}
1	*ba*^{3}*sæ*^{1}	*ča*^{2}*sĩ*^{24}	*kua*^{3}*sæ*^{1}	*ča*^{4}*sĩ*^{24}
2	*ča*^{2}*se*^{2}	*ča*^{2}*sũ*^{2}	*ča*^{4}*se*^{2}	*ča*^{4}*sũ*^{2}
3	*ba*^{3}*se*^{2}		*kua*^{4}*se*^{2}	

-*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	Average entropy	Avg. cond. 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

Table 5: Average conditional entropies for the Chiquihuitlán Mazatec inflectional systems

Each lexical item is a member of some conjunction 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

These appear to be independent systems of inflection classes

By Baerman & Corbett's count, most joint 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)

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

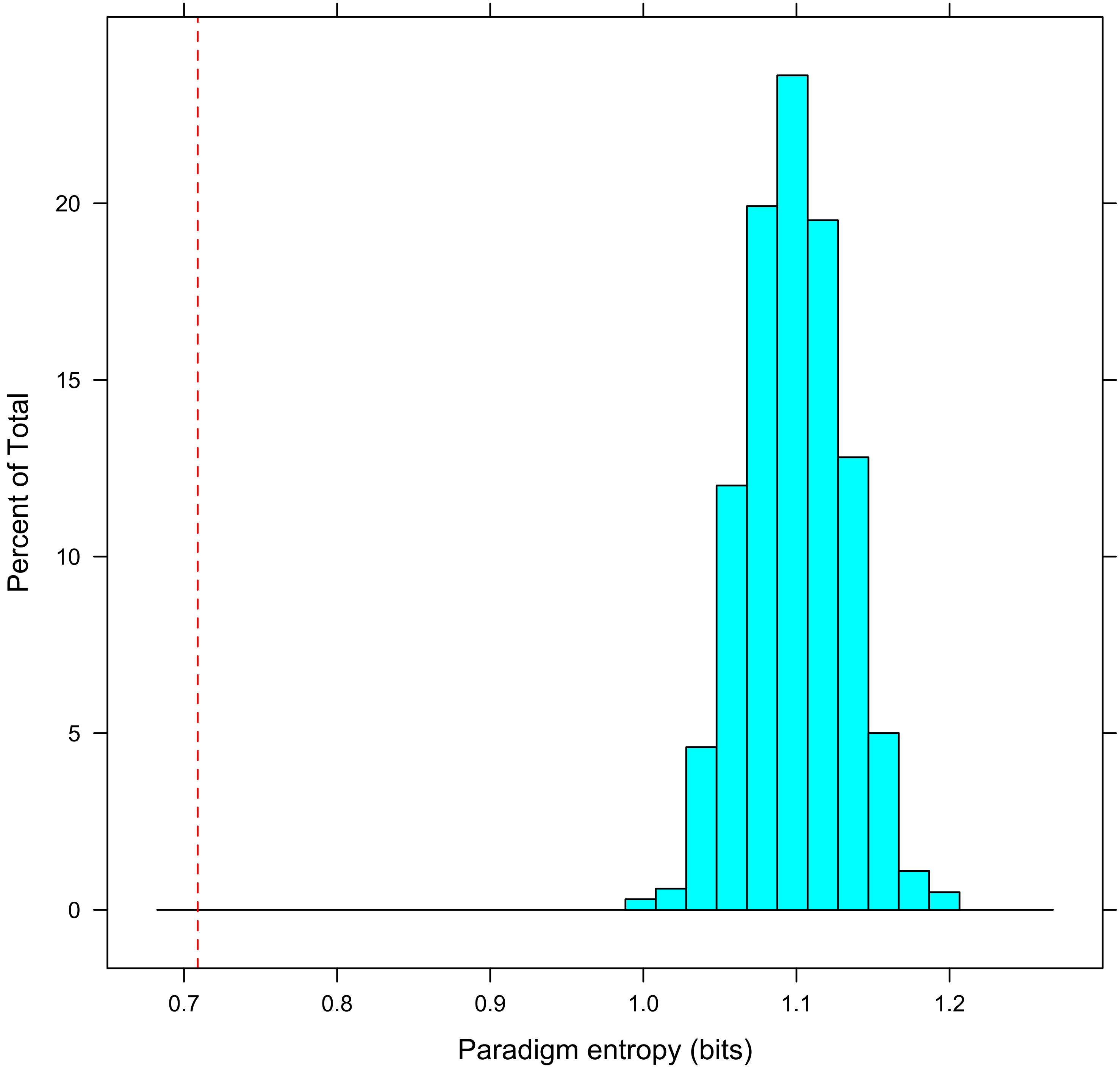
How can we test this?

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



Language	Cells	Realizations	Declensions	Declension entropy	Expected entropy	Paradigm entropy	Bootstap Avg	Bootstrap *p*
Amele	3	31	24	4.585	2.88213807	1.10533067	1.32749714	0.001
Arapesh	2	41	26	4.700	4.07053803	0.62990168	0.62990168	1
Burmeso	12	24	2	1.000	1	0	0	1
Fur	12	80	19	4.248	2.39476245	0.51688084	1.31619569	0.001
Greek	8	12	8	3.000	1.621	0.644	0.891	0.001
Kwerba	12	26	4	2.000	0.86397239	0.42803030	0.52267333	0.001
Mazatec	6	356	109	6.768	4.92	0.709	1.1	0.001
Ngiti	16	68	10	3.322	1.93664164	0.48399289	1.01926185	0.001
Nuer	6	12	16	4.000	0.86367582	0.79261309	0.81094024	0.16
Russian	12	26	4	2.000	0.91115286	0.53752563	0.54118103	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?