Paradigmatics, syntagmatics, and the agglutinative analogue

"In 1953 Floyd Lounsbury tried to tell us what we were doing with our clever morphophonemic techniques. We were providing alternations by devising an 'agglutinative analog' of the language and formulating rules that would convert expressions in that analog into the shapes in which they are actually uttered. Of course, even such an agglutinative analog , with its accompanying conversion rules, could be interpreted merely as a descriptive device. But it was not in general taken that way; instead, it was taken as a direct reflection of reality. We seemed to be convinced that, whatever might superficially appear to be the case, every language is 'really' agglutinative." (Hockett 1987:84)

Rob Malouf
San Diego State University

Farrell Ackerman
University of California at San Diego

General overview

Basic Question: Does agglutination, and the morpheme-based assumptions it presupposes, have any privileged status in explaining morphology?

1. Competing approaches to Morphology:

   Morphemic versus Word and Paradigm Models

2. Recent results from Word and Paradigm Implicative Models:

   Paradigm Cell Filling Problem and Low Entropy

3. Reflections on achieving these results:

   Objects of analogy w/o internal structure

4. Operationalizing analogy for objects w/ internal structure

   Many ways for word structure to yield low entropy

Morphology: Morphemic approach

Morphology: "Morphology is at the conceptual centre of linguistics. This is not because it is the dominant subdiscipline, but because morphology is the study of word structure, and words are at the interface between phonology, syntax, and semantics." (Spencer & Zwicky 1998:1)

Definition 1: "Morphology is the study of the combination of morphemes to yield words" (Haspelmath & Sims 2010:2)

Finnish: *lase-*   *i-*   *ssa*
            | | |
         cup-   PLURAL- INESSIVE
         'in the cups'

Morphology as the "syntax of morphemes" (Blevins 2007; Booij 2010:1)

Morphology: Morphemic approach

Syntagmatic: linear (and hierarchical) coordination of bi-unique form-meaning mappings, i.e. morphemes elements.

Compositional: derives the meaning of the whole word from the meanings of its isolable morphemic subparts.

Vertical: derivational, positing an underlying or surface structure form from which a target structure can be asymmetrically derived, i.e., accounts for all (regular) forms exhibiting the same morphosyntactic properties.

Morphology: Morphemic approach

Agglutinative Hypothesis: All complex words in all language are structured like Finnish, even if they don't seem so on the surface

Consequences:

Language particular surface variance and deviation from agglutination is instructive about underlying invariance.

Words and paradigms are epiphenomena, i.e., they disappear when we discover the rules responsible for piecing words together out of morphemes.

Recent morphemic approaches

"...words don't exist, they are the phlogiston of linguistics. Everyone must say that sentences are built out of morphemes... We... expect a high degree of isomorphism, of the type expressed by Baker's (1985) Mirror generalization; although various affixal properties may lead to readjustment rules that end up masking syntactic structures; ... The extent of such readjustment rules **enriches** the morphological component of the grammar, and makes it difficult to order vocabulary insertion with respect to linearization." (Boeckx, *Bare Syntax* 2008:63-64)

(3)   A Distributed Morphology Derivation:

	Operation		Output
a.	**Syntax:** Construct Numeration by selecting feature (bundles) (List 1).	a.	{BE, [+1, +sg, +f]$_D$, [+past]$_T$, TALK, [+Prog]}
b.	**Syntax:** Construct interpretable sentence structure by Merge, Move of feature (bundles). (The output of this step is sent to LF for semantic interpretation, and to PF for Spell-Out.)	b.	TP tree: D° [+1][+sg][+f] and T' ; T° [+past][+1][+sg][+f][BE] , ProgP ; Prog° , VP ; V°$_i$ [TALK] Prog° [+Prog] V°$_i$ [TALK]
c.	**Morphology:** Manipulate makeup of terminal bundles to conform to language-specific requirements (e.g. by Impoverishment, on which more anon).	c.	TP tree: D° [+1][+sg] and T' ; T° [+past][+sg][BE] , ProgP ; Prog° , VP ; V°$_i$ [TALK] Prog° [+Prog] V°$_i$ [TALK]
d.	**Morphology:** Realize (or 'discharge') the terminal nodes of the syntactic tree by inserting Vocabulary Items (List 2) into them, giving them phonological content.	d.	[[/aj/]$_D$ [[/wʌz/]$_T$° [[[/tɑk/]$_V$[/ɪŋ/]$_{Prog}$]$_{Prog}$]$_T$']$_{TP}$
e.	**Phonology:** Make morphophonological and phonological alterations to input as necessary to arrive at the optimal phonological form	e.	[¹ajwəz¹tʰɑkin]

Astonishing Hypothesis: Deviations from expectation count as discoveries that "enrich" our understanding of the morphology-syntax interface. Presumably the greater the departure from expectations, the more profound the enrichment. Couldn't the need for such enrichment count as evidence against the basic program?

Does not explain the actual rarity of systematic and widespread agglutination, but it's good to know that this rarity is illuminating for theory construction.

Problems with the agglutinative ideal

The agglutinative ideal: each morpheme expresses one meaning, each meaning is expressed by a single morpheme and all meanings are in a complex word are expressed as morphemes.

To preserve a one-to-one mapping between forms and meanings, morphology often requires covert morphemes, as well as (essentially arbitrary claims about) sequences of such elements.

Finnish:
lase-	*i-*	*ssa*	*lasi-*	*ø-*	*ø*
|	|	|	|	|	|
cup-	PLURAL-	INESSIVE	cup-	SINGULAR-	NOMINATIVE

'in the cups'

'the cup'

Problems with the agglutinative ideal

"I know of no compensating advantage for the modern descriptive reanalysis of traditional paradigmatic formulations in term of morpheme sequences. This [= morphemic analysis - FA/RM] 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)

On reflection

But,

(i) 46 years have passed

(ii) Many paradigmatic approaches have been developed – in fact, they are preferred over morpheme-based models by working morphologists – and applied with rigor and precision over wide varieties of phenomena and broad arrays of languages,

(iii) so, it no longer "difficult to say anything more definite."

What is the traditional alternative?

Morphology: Word and Paradigm

Finnish: *laseissa*   *lasi*

         | |

         cup.PLURAL.INESSIVE   cup.SINGULAR.NOMINATIVE

         'in the cups'   'cup'

Definition 2: "Morphology is the study of the systematic covariation in the form and meaning of words" (Haspelmath & Sims 2010:2)

Morphology is concerned with both the internal organization of words (how wordshapes are distinctively associated with meaning) and the external organization of word (how related words participate in similar patterns).

Morphology: Word and Paradigm

Paradigmatic: identifies (sets of) patterns and subpatterns that whole words participate in.

Configurative: the meaning of a word is not necessarily a straightforward composition of individually meaningful parts; it is the total arrangement of co-occurring elements of words that are conjointly associated with meaning.

Horizontal: relations among networks of inflected surface forms (i.e., whole words) which share a lexeme as well as the schematic patterns of implications abstracted from them.

[see Blevins (2006) for a more comprehensive introduction to W&P morphology]

Old intuitions: associations and analogies

Proportional analogies for words:

pain : *painful* :: *disdain* : ?

"A unit like *painful* decomposes into two subunits *(pain-ful)*, but these units are not two independent parts that are simply lumped together *(pain + ful)*. The unit is a product, a combination of two interdependent elements that acquire value only through their reciprocal action in a higher unit *(pain × ful)*. The suffix is non-existent when considered independently; what gives it a place in the language is series of common terms like *delight-ful, fright-ful,* etc…. The whole has value only through its parts, and the parts have value by virtue of their place in the whole." (Saussure 1966:128) [see Baayen & Hay 2005 for similar insights]

Old intuitions: associations and analogies

"Thus it is that the different uses, which have come to be associated with a word or a phrase, associate themselves with each other. Thus to the different cases of the same noun, the different tenses, moods and persons of the same verb, the different derivatives of the same root, associate themselves, thanks to the relationships between their sounds and the meaning; ....further forms of different words with similar functions – e.g., all plurals, all genitives, ...all masculines which form their plural by means of umlaut as contrasted with those that form it otherwise...*These associations may one and all arise and operate without consciousness, and they must not be confounded with grammatical categories, which are the result of conscious abstraction, though they not infrequently cover the same ground.*" (Paul 1890:6 [in 1970 trans.])

Traditional assumptions, recent results

"Paul's concept of analogy and of analogical proportion is a definite attempt at providing a generalized account at a certain level of detail of how language production occurs and of how the speaker and hearer can produce and analyze an infinite number of forms and sentences which they have not heard before." (Morpurgo Davies 1998:259)

Central role of (proportional) analogy in identifying the patterns of relatedness among words and how this facilitates mastery of morphological systems

"It is worth noticing...that the proportion is a neutral form of notation, i.e., it may be interpreted in cognitive or structural terms and may or may not be reformulated in terms of morphological or morphophemic rules. At the same time it offered an algorithm for a structurally based form of segmentation without making any claims about the status of the segments in questions." (Morpurgo Davies 1998:258)

Results are not dependent upon morpheme-based assumptions of word structure: a morphemic agglutinative strategy has no *a priori* advantage

Goals of the presentation

1. When analogy is appropriately defined and operationalized, a central organizing principle of morphological organization is relevant not only to patterned organization of (complex) words across paradigms but to the syntagmatics of words within single paradigms

   a. The Low Entropy Conjecture operates in paradigmatic relations

   b. The Low Entropy Conjecture operates in syntagmatic relations

   c. A single measure concerning patterned organization more broadly characterizes both paradigmatics and syntagmatics.

Goals of the presentation

2. The agglutinative strategy has no analytic priority (and morphemes no necessary theoretical status).

   a. The agglutinative ideal, when it (rarely) occurs, is merely a (common) stage of diachronic development for complex words.

   b. The explanation for its rarity is that it is only one particularly simple strategy associated with low entropy.

   c. It is low entropy that is a fundamental property of morphological systems, not the structural arrangement of bi-unique form-meaning mappings

   d. Theories based on the agglutinative ideal have reified an historically contingent state of language, mistaking it for an organizing principle of morphology.

The rest of the talk

1. The Paradigm Cell Filling Problem and Low Entropy

2. The Low Entropy Conjecture: Analogy without segmentation

3. Syntagmatics: Analogy with segmentation

3. Concluding remarks

Paradigm Cell Filling Problem

It is implausible that speakers of languages with complex morphology and multiple inflection classes encounter every inflected form of every word (Paul 1890/1970; Anttila 1977; Bybee 1985; Itkonen 2005)

Paradigm Cell Filling Problem: Given exposure to a novel inflected word form, what licenses reliable inferences about the other word forms in its inflectional family? (Paunonen 1976, among others)

Complex paradigms are organized in a way to make the PCFP tractable (Ackerman, *et al.* 2009; Stump & Finkel 2009; Bonami *et al.* 2010)

As languages depart from the agglutinative ideal, predicting forms by analogy becomes increasingly important: Paradigms are networks of implicative relations among related wordforms and inflectional classes are patterns of wordforms displaying distinctive implicational relations. (Paul 1890/1970:101)

Paradigm Cell Filling Problem

Finnish partial nominal paradigms (following classification in Pihel & Pikimäe 1999:758-771)

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	*oven*	*ovea*	*ovia*	*ovissa*	'door' (8)
kieli	*kielen*	*kieltä*	*kieliä*	*kielissä*	'language' (32)
vesi	*veden*	*vettä*	*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)

To deduce the Finnish NOM.SG. for *pantti* 'pants', it is enough to encounter the PART.SG. *panttia* (on analogy with *lasi~lasia* 'cup'), since no other PART.SG. ends in *-ia*

The PART.PL. *pantteja* restricts class membership to either 4 or 9, i.e., either *pantti* or *pantte*, since PART.PL. *-eja* occurs in both

Encountering the INESS.PL. *pantteissa* restricts class membership to 4, 9, or 78, since all three share the segments *-eissa*

Measuring implicative relations

With others (among others: Sims 2010; Bonami *et al.* 2011) we measure implicative relations in terms of conditional entropy within Information Theory.

Conditional entropy provides a measure of how difficult it is to reliably guess e.g., one form of a word given another: the lower the entropy, the easier the accessibility to the (set of) licensed guess(es)

Measuring implicative relations

The conditional entropy of one cell given another is a measure of inter-predictability of individual words within paradigms (Moscoso del Prado Martin *et al.* 2004; Ackerman, Blevins, and Malouf 2009)

To extend this to the whole paradigm, we calculate the paradigm entropy: the *average* conditional entropy in predicting an unknown wordform given a known wordform

Ackerman et. al. 2009 found that for the Finnish partial paradigm, the paradigm entropy is **0.663 bits.**

Despite the kaleidoscopic complexity of Finnish nominal inflection there are many reliable cues to predict which pattern a novel word belongs to.

Extending language particular results

Guiding intuition (Malouf & Ackerman 2010): morphological systems must be simple in ways that allow them to be learned and used by native speakers, irrespective of how profusely complicated words and paradigms may appear according to external measures.

External complexity (E-complexity) of a morphological system: the surface exponence for words in all of their variety, i.e. the number and types of morphosyntactic categories and the strategies for encoding them.   (e.g., Finnish has 15 cases and two numbers; Tundra Nenets has 7 cases and three numbers) - some Ls may be agglutinative, some may not.

The Finnish morphological system permits 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

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

The issue is not simplicity or complexity per se (E-complexity), but the nature of organization supporting that complexity (reflective of memory storage for words, patterns, and procedures for generalization): Explanation is in terms of I-simplicity.

I-simplicity is measurable and quantifiable

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

Paradigm entropy: Beyond Finnish

Paradigms vary tremendously in their E-complexity (Malouf & Ackerman 2010)

Mazatec has morphosyntactic property combinations that define 6 distinct cells; there are 356 different forms distributed across 109 different classes.

Language	Cells	Realizations	Declensions	Paradigm entropy
Arapesh	2	41	26	0.630
Burmeso	12	24	2	0.000
Fur	12	80	19	0.517
Greek	8	12	8	0.644
Kwerba	12	26	4	0.428
Mazatec	6	356	109	0.709
Ngiti	16	68	10	0.484
Nuer	6	12	16	0.793
Russian	12	26	4	0.538

The paradigm entropy is quite low in all these languages.

Pausing to reflect

An evident advantage: the internal structure of words seems irrelevant to the presence of low entropy, since many of these languages depart dramatically from the agglutinative ideal.

Independence of syntagmatics and paradigmatics is a property previously observed to obtain in WP models (Itkonen 2005)

However, Ackerman, *et al*. solved the PFCP by keying analogy to inflection class membership: which inflection classes are consistent with the known wordforms of a lexeme (cf. Finnish *pantti*)?

Pausing to reflect

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
lasi	*lasin*	?	*laseja*	*laseissa*	'glass'
pantti	?	*panttia*	*panteja*	?	'pants'

↓

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
lasi	*lasin*	*lasia*	*laseja*	*laseissa*	'glass'
pantti	*panttin*	*panttia*	*panteja*	*pantteissa*	'pants'

Pausing to reflect

Two possible problems for this conclusion:

1) Crucially assumes inflection classes as theoretical primitives: the task was to identify the most likely class assignment for a lexeme, and use this to guess new wordforms, i.e., guessing the right class implicates selection of likely new form.

2) The method of analogy used ignored internal shape of the words that entered into analogical relations, since the words were coded as distinctive atomic units.

New challenge

Low entropy results are independent of the specific type of structure imputed to words; as long as contrasts are identifiable the computation can proceed.

For any particular language, though, the paradigm entropy does depend on the choice of representation

Furthermore, Malouf & Ackerman's (2010) hypothesis is that languages develop low entropy paradigms over generations as a side effect of the way class-based analogy works

Can we still explain the low entropy effects if we operationalize analogy in a way which keys into the form of words?

Solving analogies: alternative strategy

Proportional analogies among wordforms (Yvon, *et al.* 2004; Stroppa & Yvon 2005; Langlais & Yvon 2008; Langlais, *et al.* 2009)

$$x : y :: z : t$$

The **shuffle** $u \bullet v$ of two words u and v contains all the words w which can be composed using all the symbols in u and v, subject to the condition that if a precedes b in u or v, then it precedes b in w

If u = *abc* and v = *def*, the words *abcdef*, *abdefc*, *adbecf* are in $u \bullet v$, but *abefcd* is not

Solving analogies

If *v* is a subword of *w*, the **complementary set** of *v* with respect to *w* (denoted by *w* \ *v*) is the set of subwords of *w* obtained by removing from *w*, in a left-to-right fashion, the symbols in *v*. When *v* is not a subword of *w*, *w* \ *v* is empty.

For example, *eea* is a complementary subword of *xmplr* with respect to *exemplar*

Shuffle and complementary set are inverses:

$$w \in u \bullet v \Leftrightarrow u \in w \setminus v$$

Solving analogies

An **analogical proportion**

$$x : y :: z : t$$

holds if symbols in *x* and *t* are also found in *y* and *z*, and appear in the same relative order

That is:

$$\forall x, y, z, t \in \Sigma^\star, x : y :: z : t \Leftrightarrow x \bullet t \cap y \bullet z \neq \varnothing$$

A corollary:

$$t \text{ is a solution of } x : y :: z : ? \Leftrightarrow t \in \{y \bullet z\} \setminus x$$

Solving analogies

An **analogical proportion** $x : y :: z : t$ holds if symbols in x and t are also found in y and z, and appear in the same relative order

$$\forall x, y, z, t \in \Sigma^\star, x : y :: z : t \Leftrightarrow x \bullet t \cap y \bullet z = \varnothing$$

$$t \text{ is a solution of } x : y :: z : ? \Leftrightarrow t \in \{y \bullet z\} \setminus x$$

seeing : unseeable :: watching : unwatchable

Shuffle   u n s e e a b l e • w a t c h i n g

Solving analogies

An **analogical proportion** $x : y :: z : t$ holds if symbols in x and t are also found in y and z, and appear in the same relative order

$$\forall x, y, z, t \in \Sigma^\star, x : y :: z : t \Leftrightarrow x \bullet t \cap y \bullet z = \varnothing$$

$$t \text{ is a solution of } x : y :: z : ? \Leftrightarrow t \in \{y \bullet z\} \setminus x$$

seeing : unseeable :: watching : unwatchable

Shuffle   u n w s e a e t c h a i b n g l e

Solving analogies

An **analogical proportion** $x : y :: z : t$ holds if symbols in x and t are also found in y and z, and appear in the same relative order

$$\forall x, y, z, t \in \Sigma^\star, x : y :: z : t \Leftrightarrow x \bullet t \cap y \bullet z = \varnothing$$

t is a solution of $x : y :: z : ? \Leftrightarrow t \in \{y \bullet z\} \setminus x$

seeing : unseeable :: watching : unwatchable

Complement   u n w s e a e t c h a i b n g l e

Solving analogies

An **analogical proportion** $x : y :: z : t$ holds if symbols in x and t are also found in y and z, and appear in the same relative order

$$\forall x, y, z, t \in \Sigma^\star, \; x : y :: z : t \Leftrightarrow x \bullet t \cap y \bullet z = \emptyset$$

$$t \text{ is a solution of } x : y :: z : ? \Leftrightarrow t \in \{y \bullet z\} \setminus x$$

seeing : unseeable :: watching : unwatchable

Complement   u n w a t c h a b l e   s e e i n g

Solving analogies

Many paths, many solutions

unseeable • watching
unwseaetchaibngle
unwseaetchaibngle\seeing
unwatchable

unseeable • watching
unwatchseeaibngle
unwatchseeaibngle\seeing
unwatchable

unseeable • watching
unseeablewatching
unseeablewatching\seeing
uneablwatch

Analogical inference

Linguistic knowledge *L* consists of a set of morphosyntactic words and their realizations

More specifically, $L = \{\langle i,o \rangle \mid i \in I, o \in O\}$ is a set of observations, where *I* is the set of possible inputs (i.e., morphosyntactic words) and *O* is the set of possible outputs (i.e., realizations). If $u = \langle i,o \rangle$, then $I(u) = i$ and $O(u) = o$

To complete a partial observation $u = \langle i,? \rangle$, first find the set of triples that define an analogy with $I(u)$

$$T(u) = \{\langle x,y,z \rangle \in L^3 \mid I(x) : I(y) :: I(z) : I(u) \}$$

Then, solve the corresponding surface analogies

$$\hat{O}(u) = \{o \in O \mid \exists \langle x,y,z \rangle \in T(u) \text{ s.t. } O(x) : O(y) :: O(z) : o \}$$

Analogical inference

Applications to unsupervised morphological learning (Lepage 1998; Stroppa & Yvon 2005; Lavallée & Langlais 2009) and translation (Langlais & Yvon 2008; Langlais, *et al.* 2009)

Morphological models assume the usefulness of segmentation into morphemes, but translation models don't

beeta-agonistit ↔ *adrenergic beta-agonists*

beetasalpaajat ↔ *adrenergic beta-antagonists*

alfa-agonistit ↔ *adrenergic alpha-agonists*

alfasalpaajat ↔ ?

We propose to use a similar analogical inference procedure for solving the Paradigm Cell Filling Problem

Nom Sg	Gen Sg	Gen Pl	Iness Sg	Iness Pl	
ovi	oven	ovien	ovessa	ovissa	'door'
kieli	kielen	kielien	kielessa	kielissä	'language'
vesi	veden	vesien	vedessa	vesissä	'water'
lasi	**lasin**	lasien	lasissa	laseissa	'glass'
nalle	nallen	nallein	nallessa	nalleissa	'teddy'
kirje	kirjeen	kirjeitten	kirjeessa	kirjeissä	'letter'
pantti	**?**	panttien	panttissa	pantteissa	'pants'

NOM.SG. 'glass'	:	GEN.SG. 'glass'	::	NOM.SG. 'pants'	:	GEN.SG. 'pants'
lasi	:	lasin	::	pantti	:	panttin

paradigmatic inference

Nom Sg	Gen Sg	Gen Pl	Iness Sg	Iness Pl	
ovi	*oven*	ovien	ovessa	ovissa	'door'
kieli	kielen	kielien	kielessa	kielissä	'language'
vesi	veden	vesien	vedessa	vesissä	'water'
lasi	lasin	lasien	lasissa	laseissa	'glass'
nalle	nallen	nallein	nallessa	nalleissa	'teddy'
kirje	kirjeen	kirjeitten	kirjeessa	kirjeissä	'letter'
pantti	**?**	panttien	panttissa	pantteissa	'pants'

NOM.SG. 'door'	:	GEN.SG. 'door'	::	NOM.SG. 'pants'	:	GEN.SG. 'pants'
ovi	:	*oven*	::	*pantti*	:	*pantten*

paradigmatic inference

Nom Sg	Gen Sg	Gen Pl	Iness Sg	Iness Pl	
ovi	*oven*	*ovien*	*ovessa*	*ovissa*	'door'
kieli	*kielen*	*kielien*	*kielessa*	*kielissä*	'language'
vesi	**veden**	*vesien*	*vedessa*	**vesissä**	'water'
lasi	*lasin*	*lasien*	*lasissa*	*laseissa*	'glass'
nalle	*nallen*	*nallein*	*nallessa*	*nalleissa*	'teddy'
kirje	*kirjeen*	*kirjeitten*	*kirjeessa*	*kirjeissä*	'letter'
pantti	**?**	*panttien*	*panttissa*	**pantteissa**	'pants'

INESS.PL. 'water' : GEN.SG. 'water' :: INESS.PL. 'pants' : GEN.SG. 'pants'

vesissä : *veden* :: *pantteissa* :

failed inference

Nom Sg	Gen Sg	Gen Pl	Iness Sg	Iness Pl	
ovi	*oven*	*ovien*	*ovessa*	*ovissa*	'door'
kieli	*kielen*	*kielien*	*kielessa*	*kielissä*	'language'
vesi	*veden*	*vesien*	*vedessa*	*vesissä*	'water'
lasi	*lasin*	*lasien*	*lasissa*	*laseissa*	'glass'
nalle	*nallen*	*nallein*	*nallessa*	*nalleissa*	'teddy'
kirje	*kirjeen*	*kirjeitten*	*kirjeessa*	*kirjeissä*	'letter'
pantti	**?**	***panttien***	***panttissa***	***pantteissa***	'pants'

INESS.PL. 'pants' : INESS.SG. 'pants' :: GEN.PL. 'pants' : GEN.SG. 'pants'

pantteissa : *panttissa* :: *panttien* : *panttin*

syntagmatic inference

Analogical inference

Given this collection of known forms, there are 35,904 possible proportional analogies, of which 74 are valid

None	24
panttin	23
pantten	13
pantteen	4
panttien	4
epantten	2
inpantt	2
npantti	1
enpantt	1

Simulation

"...we can gain some insight into the forces shaping languages by considering them as products of a historical spiral involving both acquisition and production, learning and speaking, and occasionally innovating, over many generations" (Hurford 2009:51)

Linguistic 'life cycle' (Bybee 2006, Hurford 2009)

Competence

↓

Performance

↓

Competence
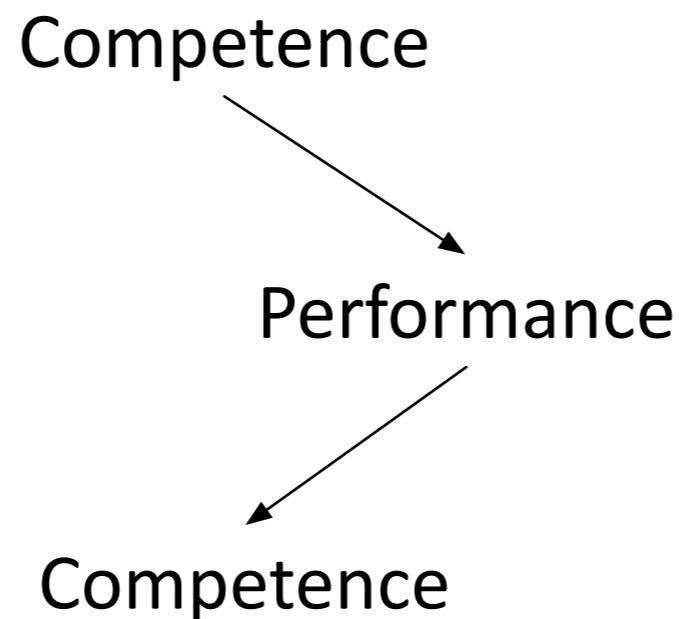
Learning bottlenecks in linguistic evolution (Hare & Elman 1995, Kirby 2001, de Boer 2001)

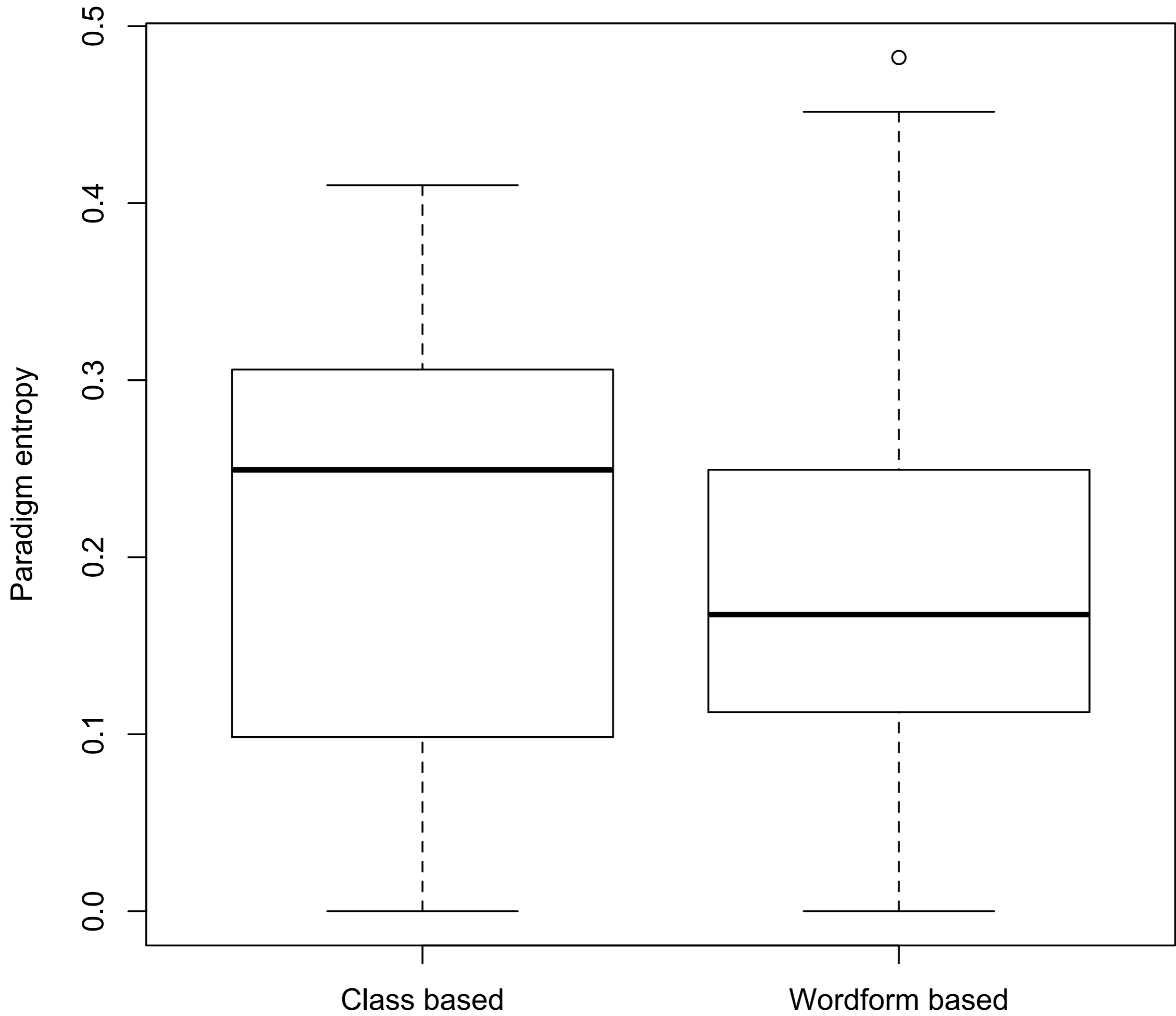# Simulation

Iterated Learning model (Kirby & Hurford 2002)

We start with a maximally complex language: given f morphosyntactic features with v possible values with r realizations each, there are r^{v^f} inflection classes

On each iteration:
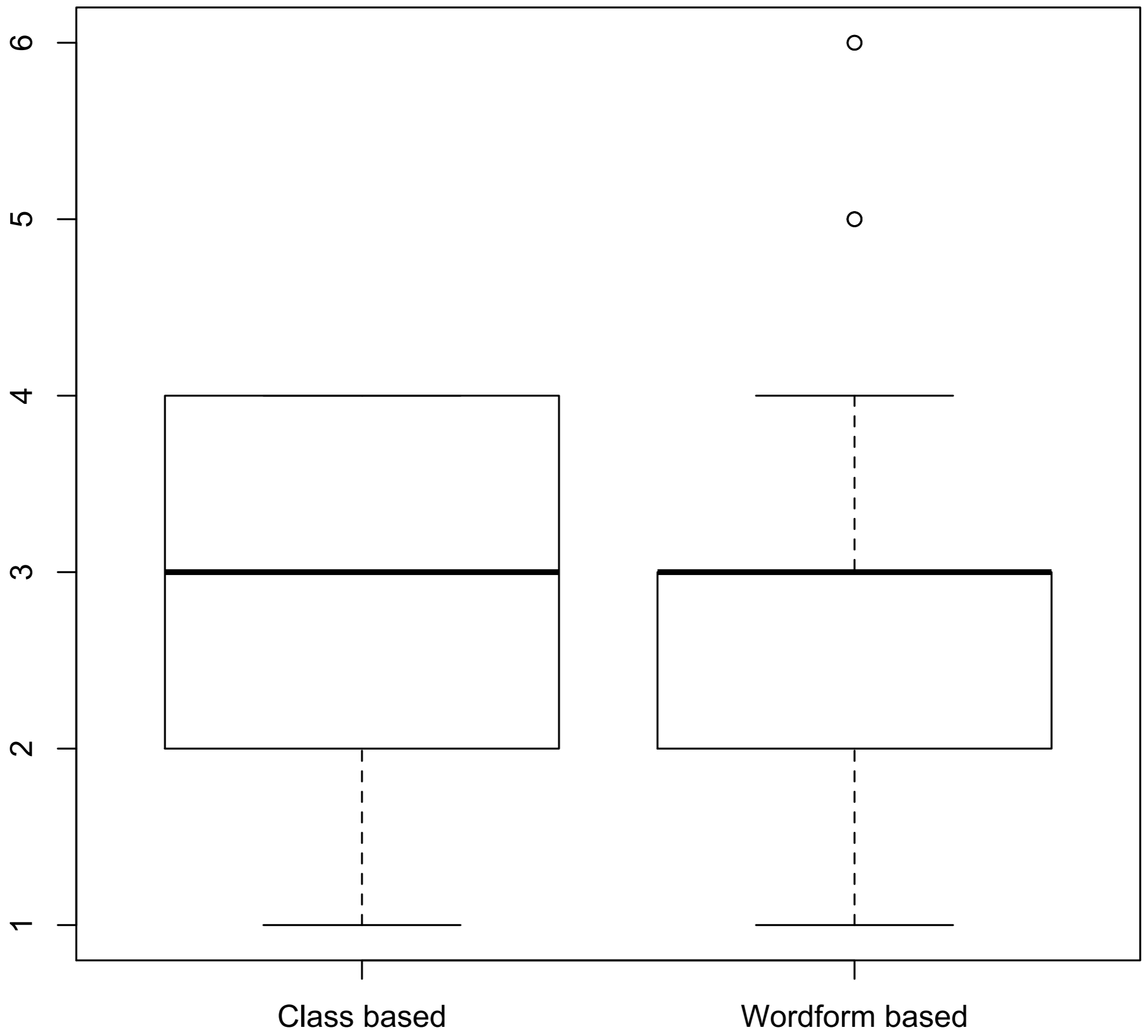
Generate a partial random sample of wordforms from the previous two iterations

Guess the realization of unobserved wordforms by analogy from the observed wordforms (occasionally, a new form is introduced)

Continue for 20 iterations, and examine the system that has emerged

Simulation

What about the agglutinative ideal?

Some of the classes that emerge in these simulated languages are agglutinative:

	$f_2 = v_1$	$f_2 = v_2$
$f_1 = v_1$	*jckh*	*jcyc*
$f_1 = v_2$	*rzkh*	*rzyc*

And some aren't:

	$f_2 = v_1$	$f_2 = v_2$
$f_1 = v_1$	*mmmw*	*gsom*
$f_1 = v_2$	*immw*	*jumw*

Simulation prospects

These simulations suggest that class-based inferences aren't necessary for developing low class-based paradigm entropy

Future directions

Address implementation issues

Apply to real languages

Pure wordform-based entropy measure?

Concluding remarks

Focus on agglutinative structures does nothing to explain the results concerning the organization of morphological systems in terms of low entropy.

But, focus on entropy measures (either using an analogical model with atomic words or segmentable words) helps discover the pervasive presence of low entropy in paradigmatics and syntagmatics and associated consequences for solving the PCFP.

Information theoretic approaches to morphological systems also account for why there is agglutination when it occurs: it is one simple strategy associated with low conditional entropy.

It's easy to learn, but other systems with greater E-complexity are also easy to learn.

Concluding remarks

Chomsky was right:

This [morphemic analysis] 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 ...

Appendix: 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

Paradigm entropy

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	oven	ovea	ovia	ovissa	'door'
kieli	kielen	kieltä	kieliä	kielissä	'language'
vesi	veden	vettä	vesiä	vesissä	'water'
lasi	lasin	lasia	laseja	laseissa	'glass'
nalle	nallen	nallea	nalleja	nalleissa	'teddy'
kirje	kirjeen	kirjettä	kirjeitä	kirjeissä	'letter'

$$H(\text{nom.sg.}) \quad = \quad -\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}\right)$$

$$= \quad 0.918 \text{ bits}$$

Paradigm entropy

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	**oven**	ovea	ovia	ovissa	'door'
kieli	**kielen**	kieltä	kieliä	kielissä	'language'
vesi	**veden**	vettä	vesiä	vesissä	'water'
lasi	**lasin**	lasia	laseja	laseissa	'glass'
nalle	**nallen**	nallea	nalleja	nalleissa	'teddy'
kirje	**kirjeen**	kirjettä	kirjeitä	kirjeissä	'letter'

$$H(\text{gen.sg.}) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{1}{6}\log_2\frac{1}{6}\right)$$

$$= 1.8 \text{ bits}$$

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** is the average uncertainty in guessing the realization for a single cell of a single lexeme's paradigm

The expected entropy for the Finnish noun paradigm fragment is 1.5 bits, equivalent to choosing among 2.8 equally likely declensions

The **declension entropy**, the uncertainty in guessing the declension of a lexeme, is $\log_2 6 = 2.6$ bits

Conditional entropy

Entropy measures the difficulty of problem that speakers never have to solve

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

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

$$= \quad H(X, Y) - H(X)$$

Information is anything that reduces entropy/uncertainty

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

Conditional entropy

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	**oven**	ovea	ovia	ovissa	'door'
kieli	**kielen**	kieltä	kieliä	kielissä	'language'
vesi	**veden**	vettä	vesiä	vesissä	'water'
lasi	**lasin**	lasia	laseja	laseissa	'glass'
nalle	nallen	nallea	nalleja	nalleissa	'teddy'
kirje	kirjeen	kirjettä	kirjeitä	kirjeissä	'letter'

$$H(\text{gen.sg.}|\text{nom.sg.}=\text{-}i) \quad = \quad -\left(\frac{2}{4} \log_2 \frac{2}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right)$$

$$= \quad 1.5 \text{ bits}$$

Conditional entropy

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	oven	ovea	ovia	ovissa	'door'
kieli	kielen	kieltä	kieliä	kielissä	'language'
vesi	veden	vettä	vesiä	vesissä	'water'
lasi	lasin	lasia	laseja	laseissa	'glass'
nalle	**nallen**	nallea	nalleja	nalleissa	'teddy'
kirje	**kirjeen**	kirjettä	kirjeitä	kirjeissä	'letter'

$$H(\text{gen.sg.}|\text{nom.sg.}=\text{-}e) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$= 1 \text{ bit}$$

Conditional entropy

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	**oven**	ovea	ovia	ovissa	'door'
kieli	**kielen**	kieltä	kieliä	kielissä	'language'
vesi	**veden**	vettä	vesiä	vesissä	'water'
lasi	**lasin**	lasia	laseja	laseissa	'glass'
nalle	**nallen**	nallea	nalleja	nalleissa	'teddy'
kirje	**kirjeen**	kirjettä	kirjeitä	kirjeissä	'letter'

$$H(\text{gen.sg.}|\text{nom.sg.}) \;=\; \left(\frac{4}{6} \times 1.5 + \frac{2}{6} \times 1\right)$$

$$=\; 1.3 \text{ bits}$$

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 Finnish fragment, the paradigm entropy is 0.663 bits (equivalent to 1.6 declensions)