



Faculty Publications

2015-09-05

Paradigms and (semi)predictability: implications for measuring complexity typologically

Jeffrey R. Parker

Brigham Young University - Provo, jeff_parker@byu.edu

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Linguistics Commons](#)

BYU ScholarsArchive Citation

Parker, Jeffrey R., "Paradigms and (semi)predictability: implications for measuring complexity typologically" (2015). *Faculty Publications*. 6264.

<https://scholarsarchive.byu.edu/facpub/6264>

This Presentation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

48th Annual Meeting of the Societas Linguistica Europaea
Leiden University Centre for Linguistics (LUCL), Leiden
September 2-5, 2015

Paradigms and (semi)predictability:
implications for measuring
complexity typologically



Jeff Parker
The Ohio State University
parker.642@osu.edu

Collaboration...(esp. Greek data)

Andrea D. Sims

The Ohio State University

sims.120@osu.edu

Some big(gish) questions for today

- What aspects of (complex) inflectional systems make them usable for speakers?
 - How do our analytic assumptions about these systems shape our assessment of their complexity and its implications for speakers?
 - To what extent are different aspects of the system more/less useful for speakers?
- Two notions of complexity (Ackerman and Malouf 2013):
- E(numerative) complexity
 - how many inflection classes, distinct sets of exponence, morphosyntactic distinctions, etc.
- I(ntegrative) complexity
 - how difficult is the system for speakers, e.g., how much uncertainty is there in predicting unknown forms

The problem

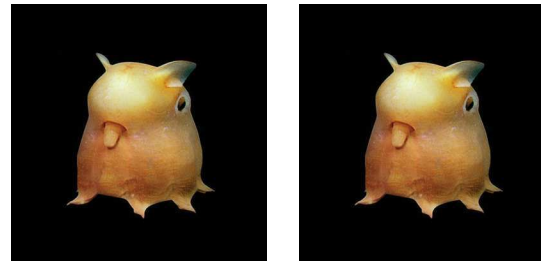
- What aspects of inflectional systems allow speakers to make motivated inferences about unknown inflected forms?

- Paradigm Cell Filling Problem, PCFP (Ackerman, Blevins, and Malouf 2009)

- This is a '*lankus*'



What are these?



'*lanki*'?

'*lankuses*'?

'*lankora*'?

- Typological question: To what extent is the PCFP more important for some languages than others?

E-complexity vs. I-complexity

- Low Entropy Conjecture

- “...enumerative morphological complexity is effectively unrestricted, as long as the average conditional entropy [the uncertainty of predicting one inflected form from another], a measure of integrative complexity, is low” (Ackerman and Malouf 2013:436)

$$H(A|B) = \sum_{b \in B, a \in A} p(b, a) \log_2 \frac{p(b)}{p(b, a)}$$

Singular (B)	virus	syllabus	corpus
Plural (A)	viruses	syllabi	corpora

E-complexity and analytic assumptions

- The description of the system can strongly influence analysis of system's complexity (Bonami 2013)
- An assumption that some/many 'irregular'/semi-regular lexemes fall outside of the morphological system risks underestimating the actual complexity speakers deal with
- How do assumptions about regularity affect our assessment of the E-complexity of Russian nouns?

Traditional description of Russian nouns

- The standard description of Russian nouns has four classes (Corbett 1982)
 - Explicit notion of economy; less storage is better
 - This ignores several (less common) traits of inflectional exponence, e.g., stress patterns, stem changes, etc.

	Class I	Class II	Class III	Class IV
	'law'	'map'	'bone'	'bog'
Nom	<i>zakon</i>	<i>karta</i>	<i>kost'</i>	<i>boloto</i>
Acc	<i>zakon</i>	<i>karta</i>	<i>kost'</i>	<i>boloto</i>
Gen	<i>zakona</i>	<i>karty</i>	<i>kosti</i>	<i>bolota</i>
Loc	<i>zakone</i>	<i>karte</i>	<i>kosti</i>	<i>bolote</i>
Dat	<i>zakonu</i>	<i>karte</i>	<i>kosti</i>	<i>bolotu</i>
Inst	<i>zakonom</i>	<i>kartoj</i>	<i>kost'ju</i>	<i>bolotom</i>

A more 'complex' description

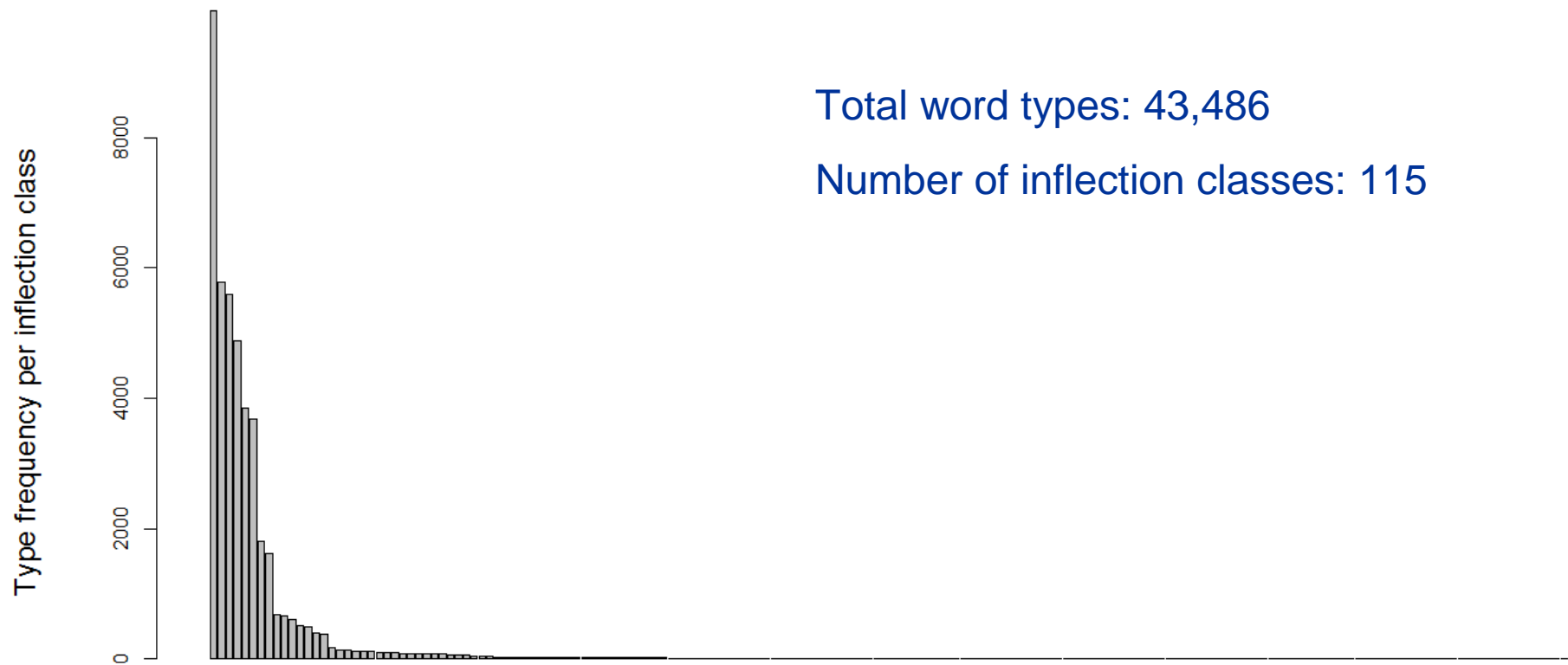
- *Grammatičeskij slovar' russkogo jazyka*
(Zaliznjak 1977)
 - 43,486 morphological nouns
 - morphological info
 - type frequencies based on inflectional traits
 - How many classes do we need to describe all aspects of Russian nouns?
 - How does accepting more semi/ir-regularity into the system affect assessment of its E-complexity
-

Number of inflection classes for Russian nouns

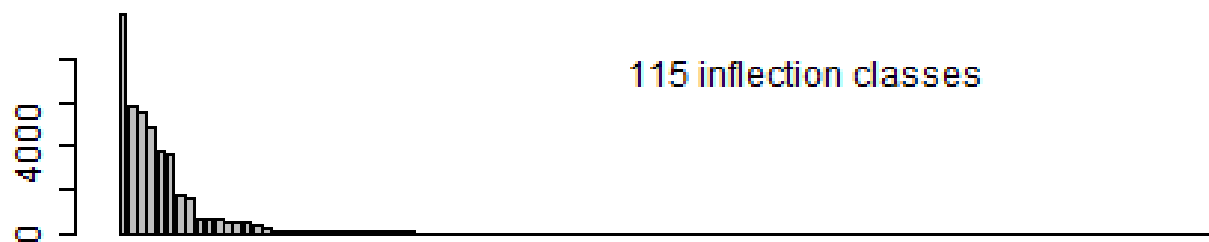
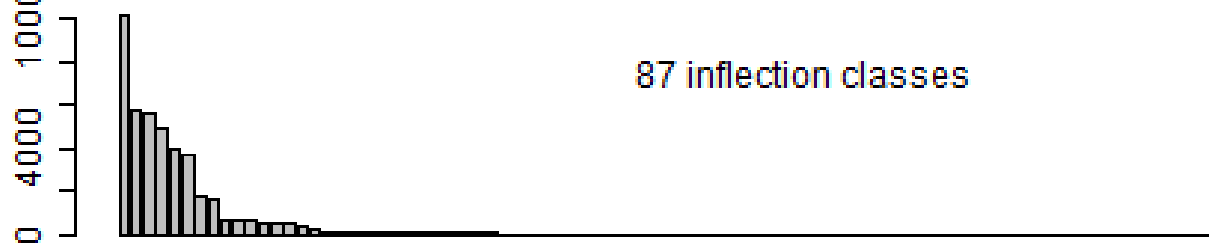
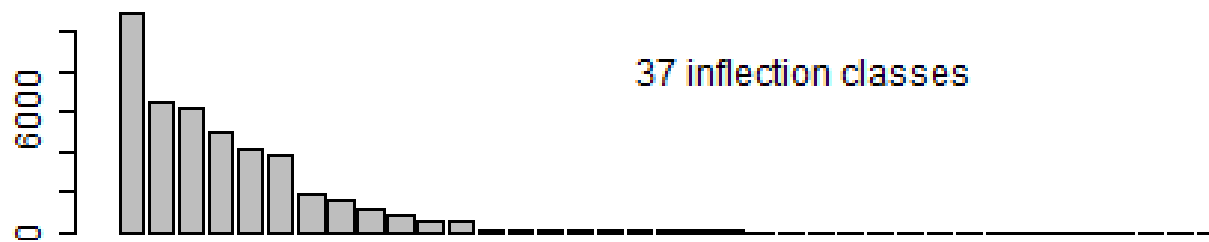
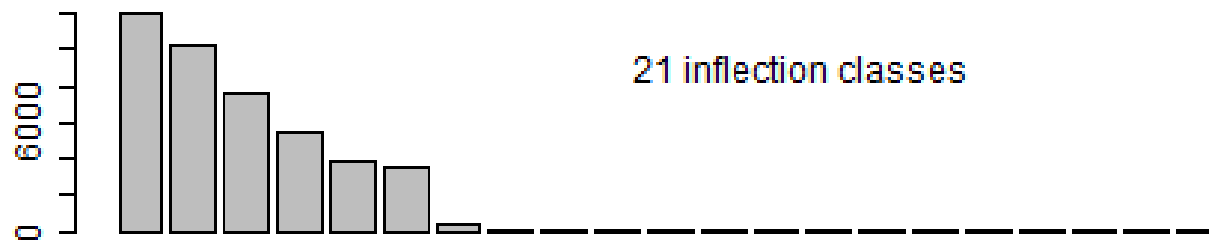
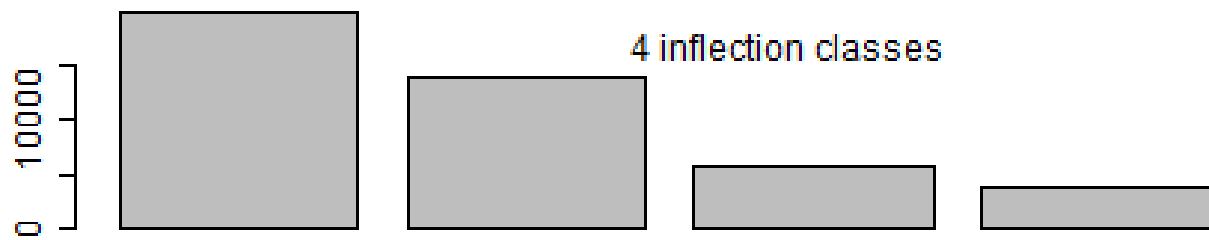
- Including everything: 795
- excluding idiosyncratic lexemes: 656
- ... and ignoring alternate traits: 214
- ... and ignoring alternate forms: 115
- ... and ignoring defectiveness: 87
- ... and ignoring stress: 37
- ... and ignoring stem changes: 21
- ... the traditional classes: 4

A closer look
at these ...

Number of word types per inflection classes



-
- Where does 'regular' end and 'irregular' begin?



Interim conclusions: E-complexity and regularity

- There is no distributional break between ‘regular’ and ‘irregular’ classes
 - Assumptions about regularity drastically affect how many classes we posit for Russian nouns (one aspect of E-complexity)
 - To what extent do different granularities of inflectional class information affect our assessment of the systems’ I-complexity (in terms of the PCFP)?
-

More 'complex' descriptions

- Russian (43,486 nouns):
 - 6 cases x 2 numbers = 12 paradigm cells
 - morphological class info and type frequencies from *Grammatičeskij slovar' russkogo jazyka* (Zaliznjak 1977)
- Greek (27,270 nouns):
 - 3 cases x 2 numbers = 6 paradigm cells
 - morphological class info from *Lexikó tīs koinīs neoellīnikīs* (Triantafillidis Institute 1998)
 - type frequencies from Hellenic National Corpus (hnc.ilsp.gr/en/)

Granularity of inflection class info

Russian nouns						
Number of classes		Traditional	All suffixes	Stem changes	Stress	Defectiveness
4		+				
21		+	+			
37		+	+	+		
87		+	+	+	+	
115		+	+	+	+	+

Greek nouns						
Number of classes		Suffixes	Inflectional stress	Stem changes	Lexical stress	Defectiveness
7		+				
19		+	+			
36		+	+	+		
49		+	+	+	+	
59		+	+	+	+	+

Sources of information and I-complexity

- Implicative paradigmatic structure
 - inflected forms vary in how much they are predictive of and/or predictable from other inflected forms
- Inflectional class type frequency
 - inflection classes differ in the number of lexemes they represent

(Wurzel 1989; Ackerman and Malouf 2013; Baerman and Corbett 2012; Sims 2015; Stump and Finkel 2013)

Lowering uncertainty: Implicative work

- A (re)definition of *work*: the reduction in the entropy of a system due to a given information source
- Implicative work - difference between entropy and conditional entropy

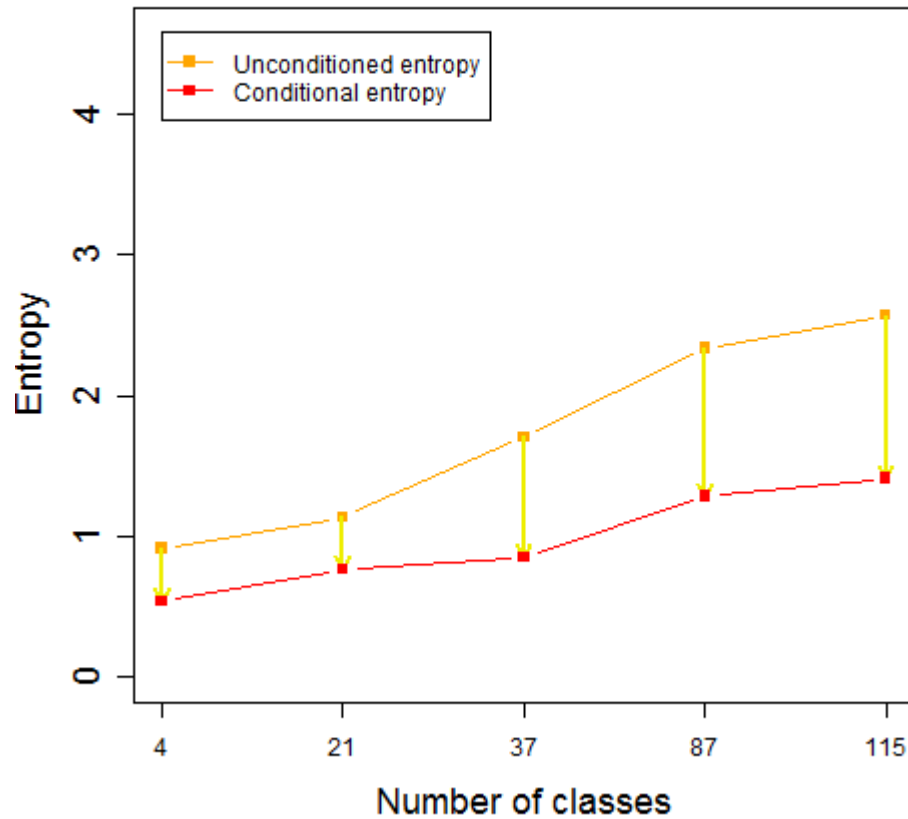
- (Unconditioned) Entropy: $H(A) = - \sum_{a \in A} p(a) \log_2 p(a)$

- Conditional entropy: $H(A|B) = \sum_{b \in B, a \in A} p(b, a) \log_2 \frac{p(b)}{p(b, a)}$

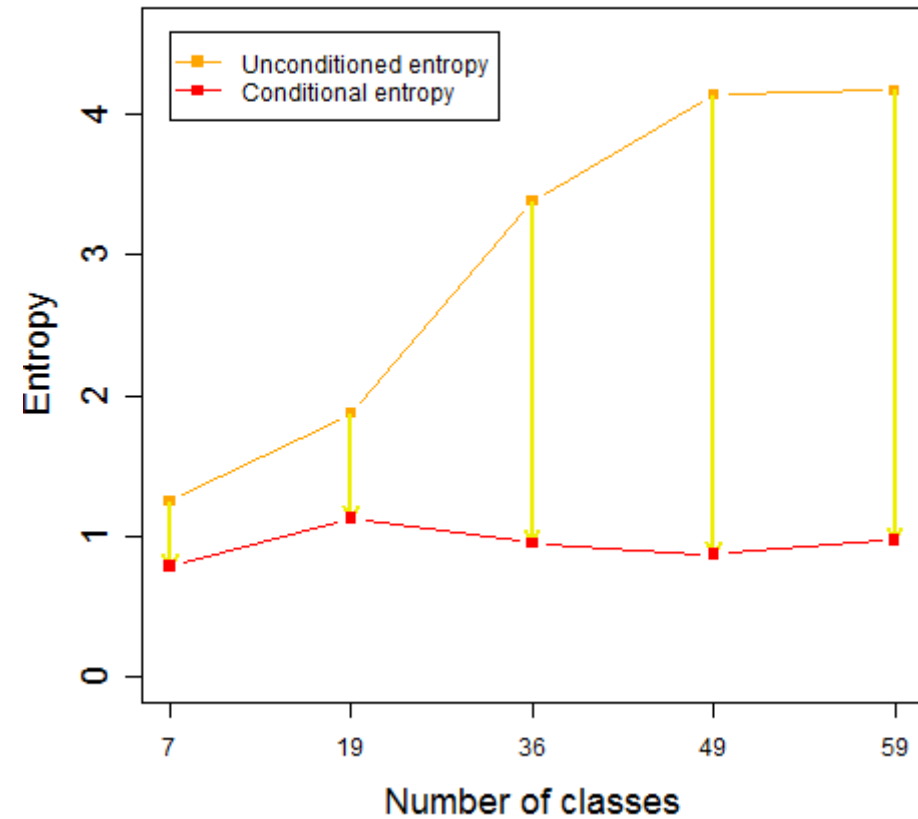
- Implicative work:
(Mutual information) $I(A;B) = H(A) - H(A|B)$

Implicative work

Russian



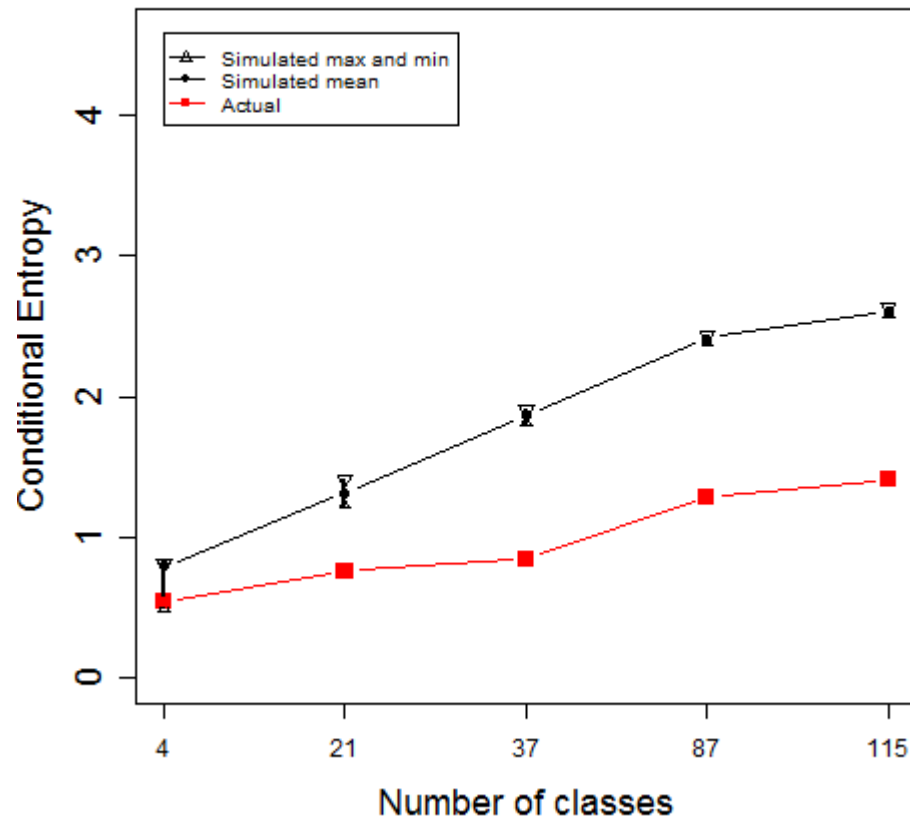
Greek



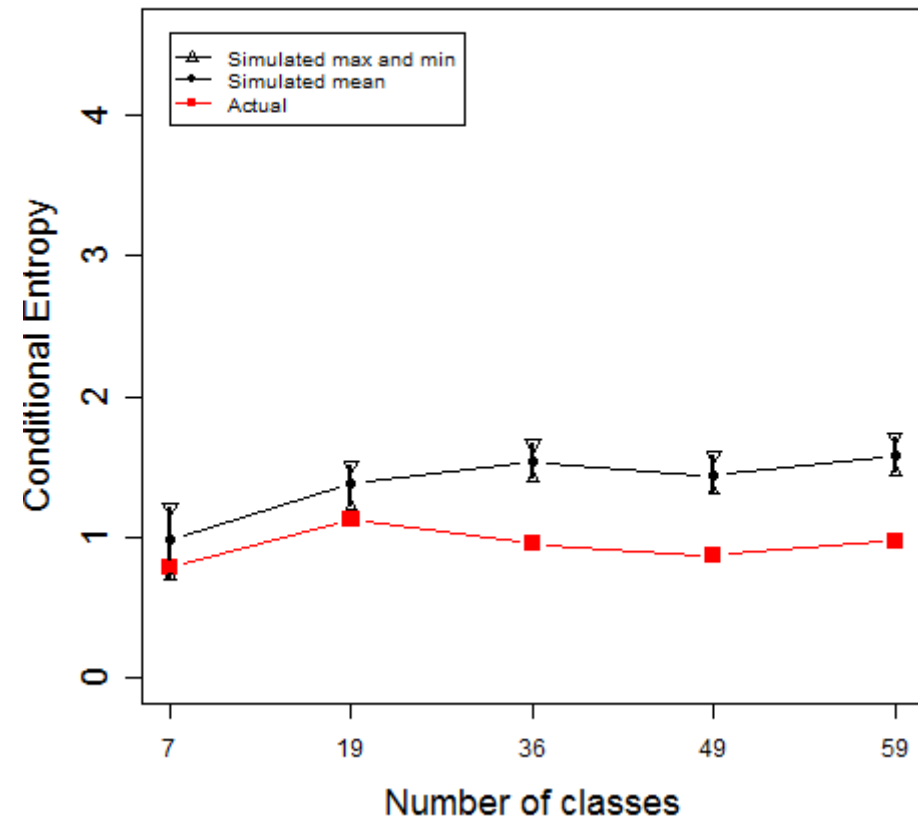
- Overall, implicative work increases as granularity increases

Significance of Conditional entropy

Russian



Greek



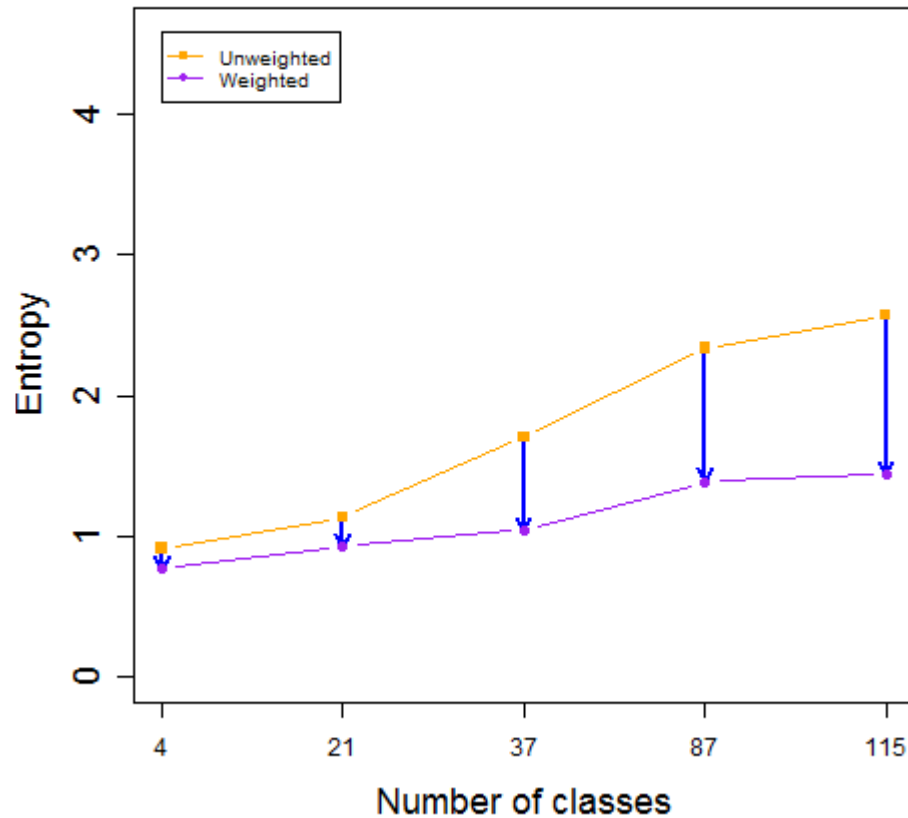
- In both languages, average uncertainty is less than by chance at all granularities; consistent with Low Entropy Conjecture
- Mostly, difference from chance increases as granularity increases

Type frequency work

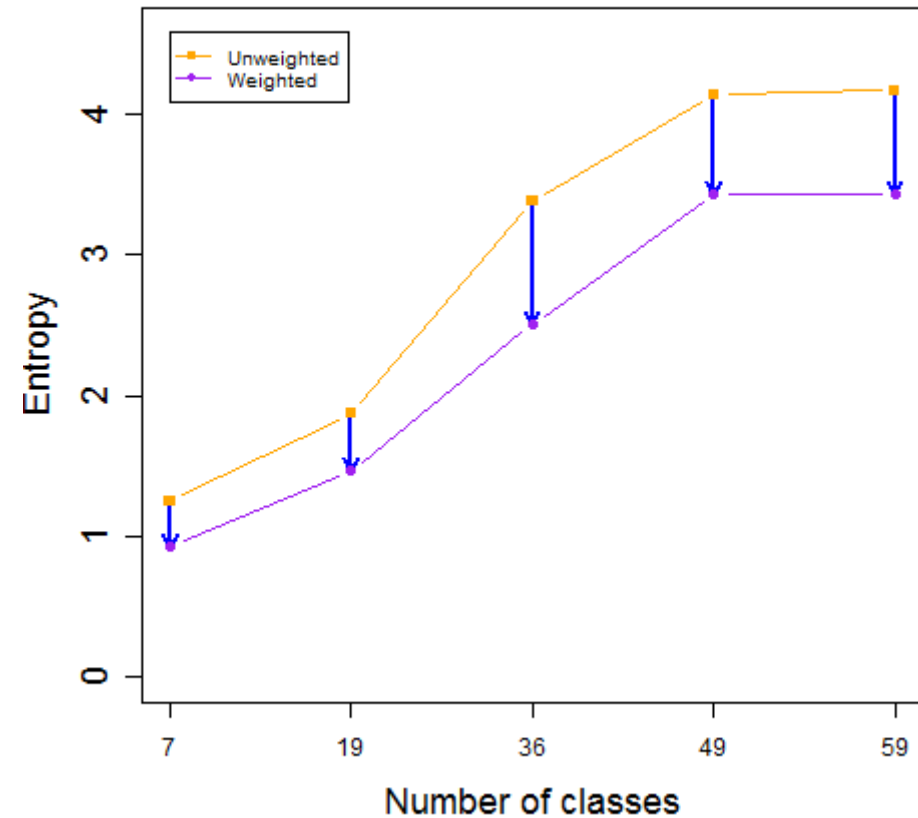
- Some classes contain thousands of lexemes, others have only one
- Type frequency work: Difference between entropy when calculated based on evenly weighted (U) and type frequency weighted (W) data structures
 - Type frequency work = $H(A)_U - H(A)_W$

Type frequency work

Russian



Greek

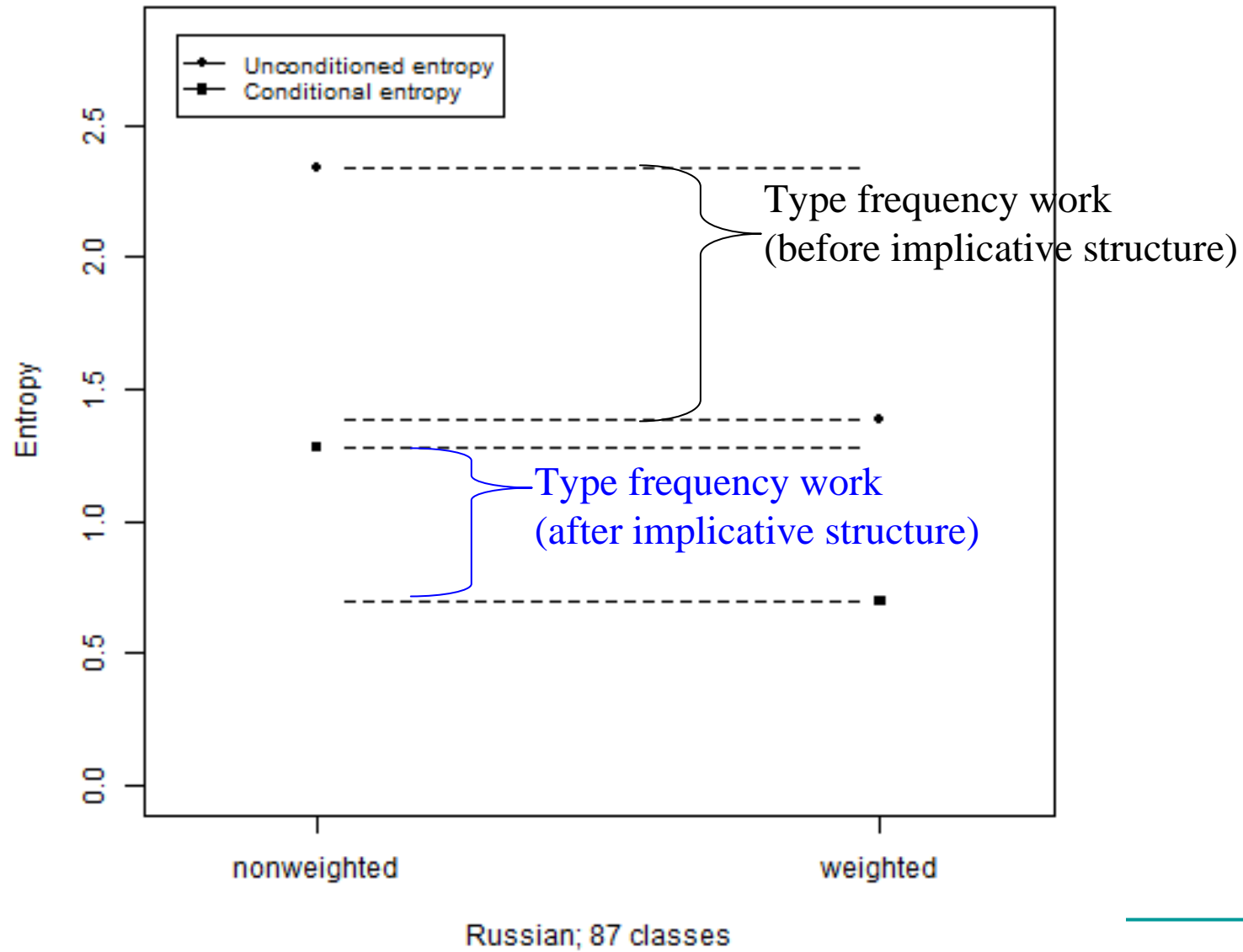


- Weighting by type frequency lowers entropy, more so in finer granularities

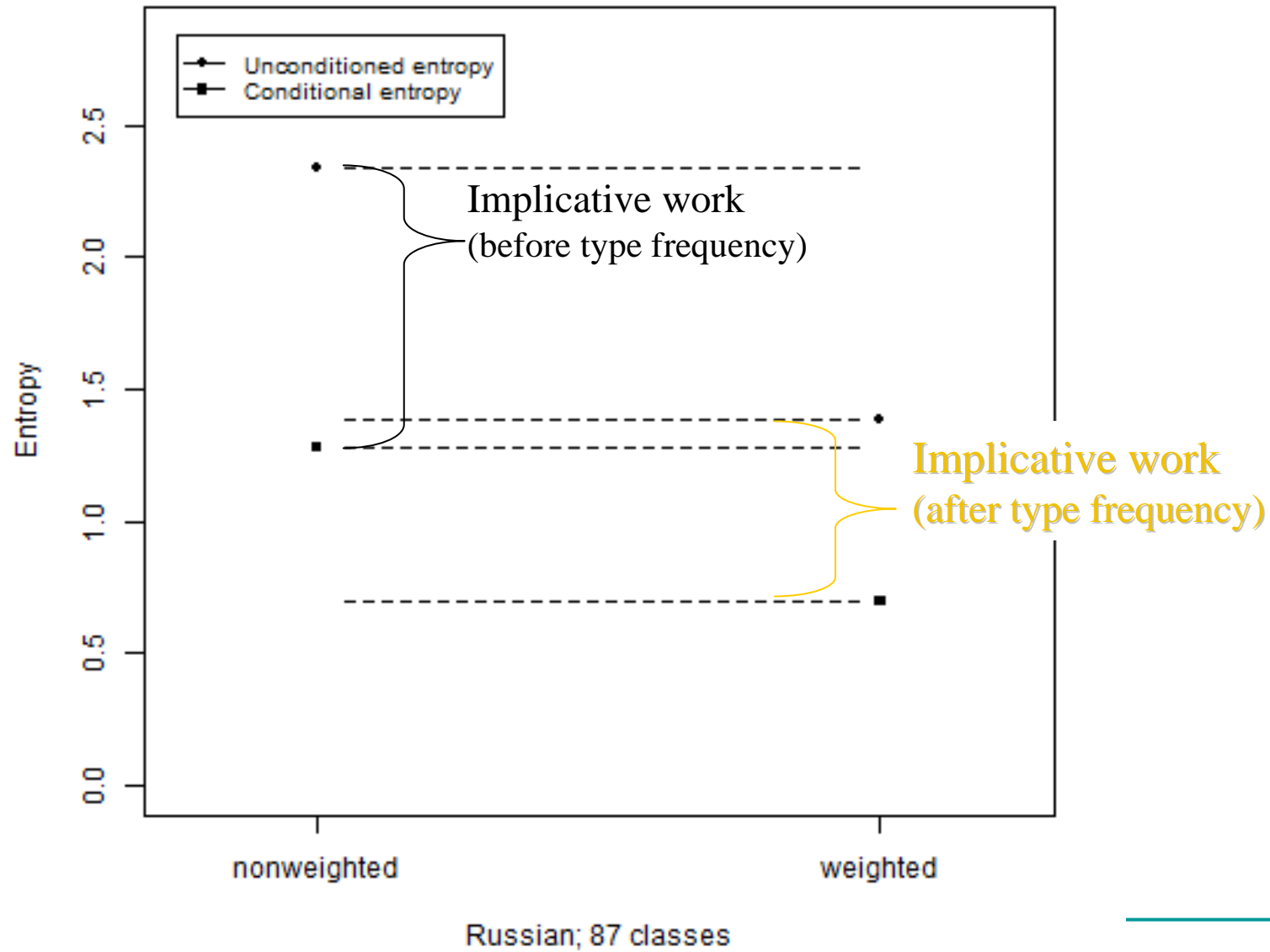
What's really doing the work?

- Both implicational structure and type frequency have the capacity to do work by lowering the entropy of the system (and do so in Russian and Greek)
- To what extent are their contributions independent and/or overlapping?

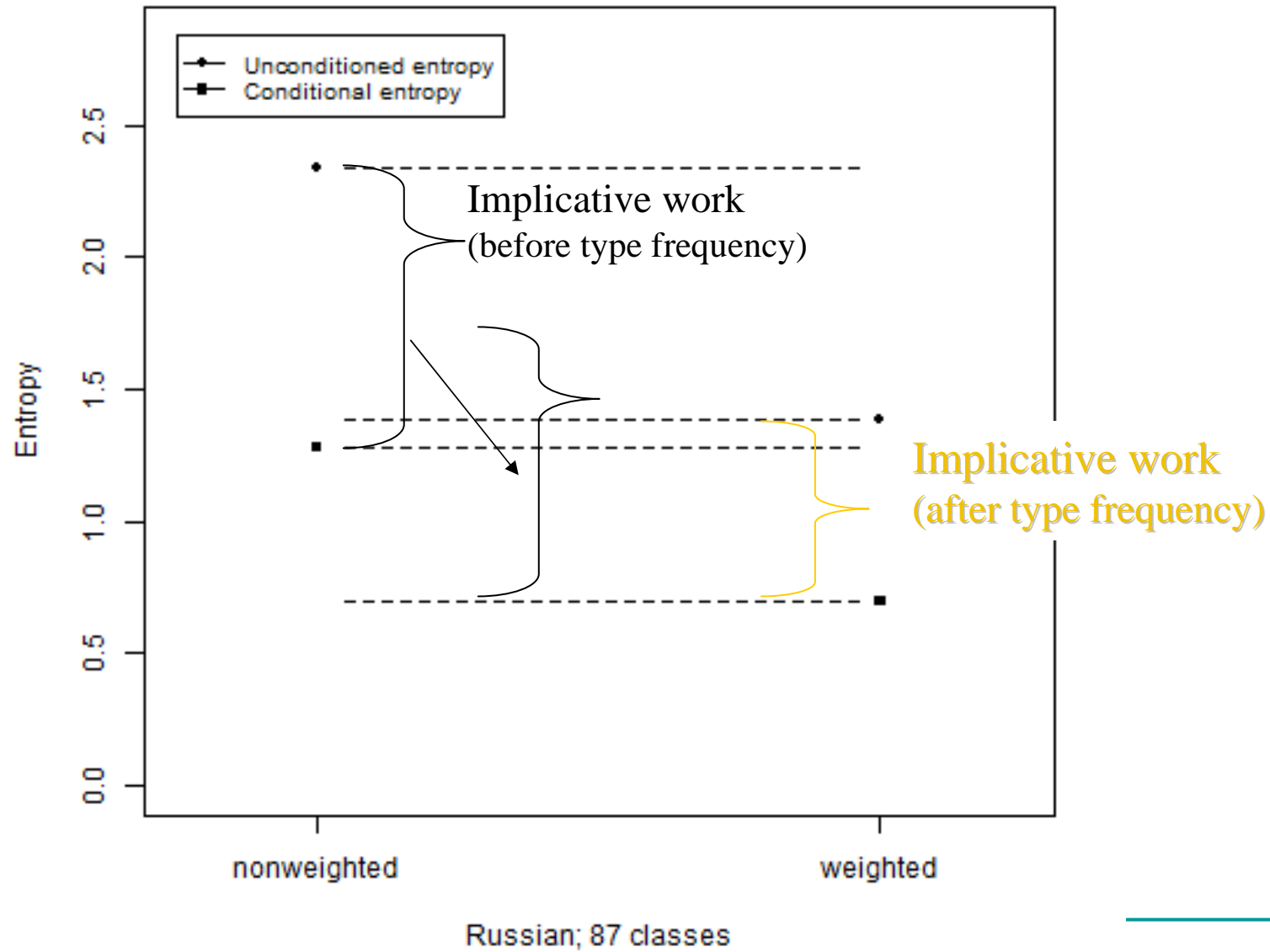
Calculating work



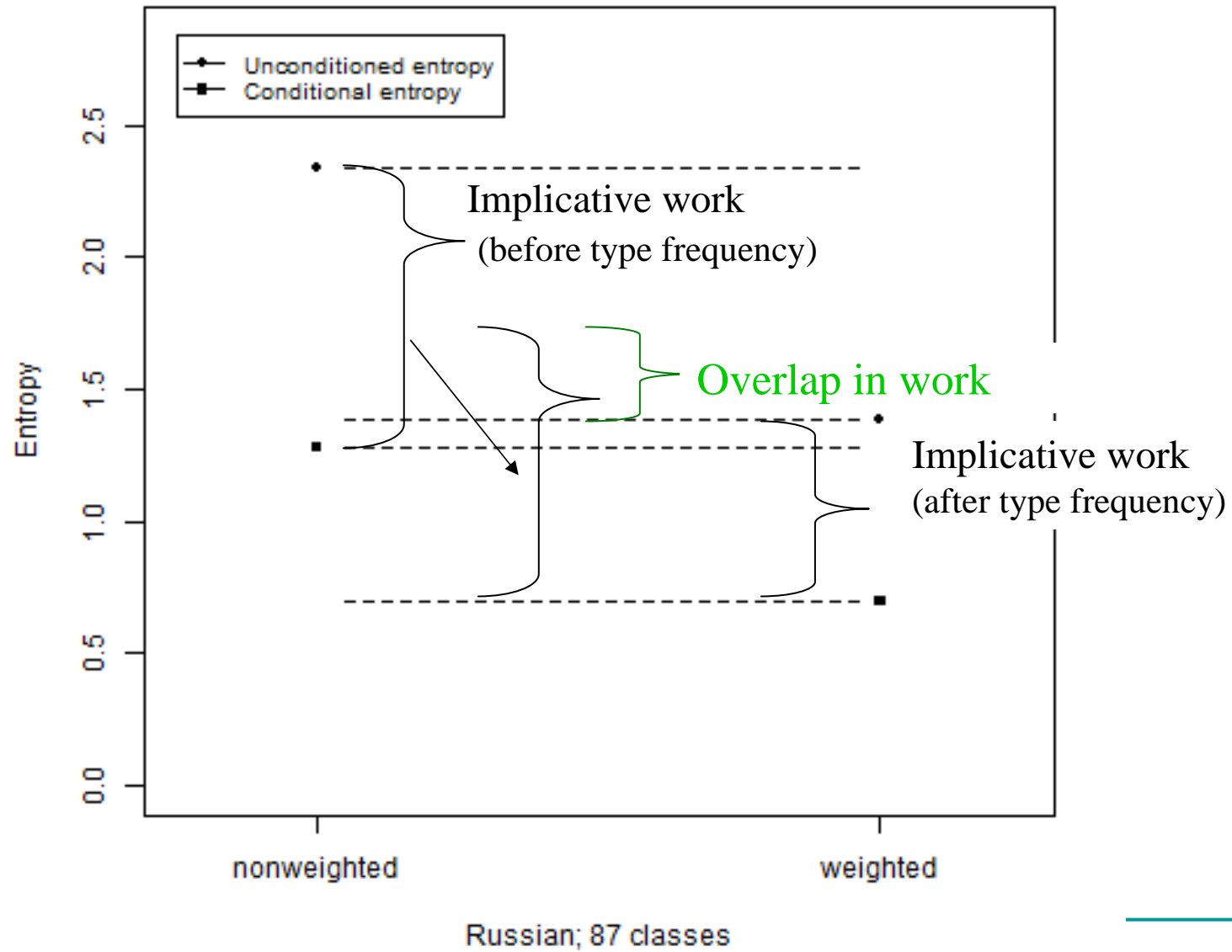
Calculating work



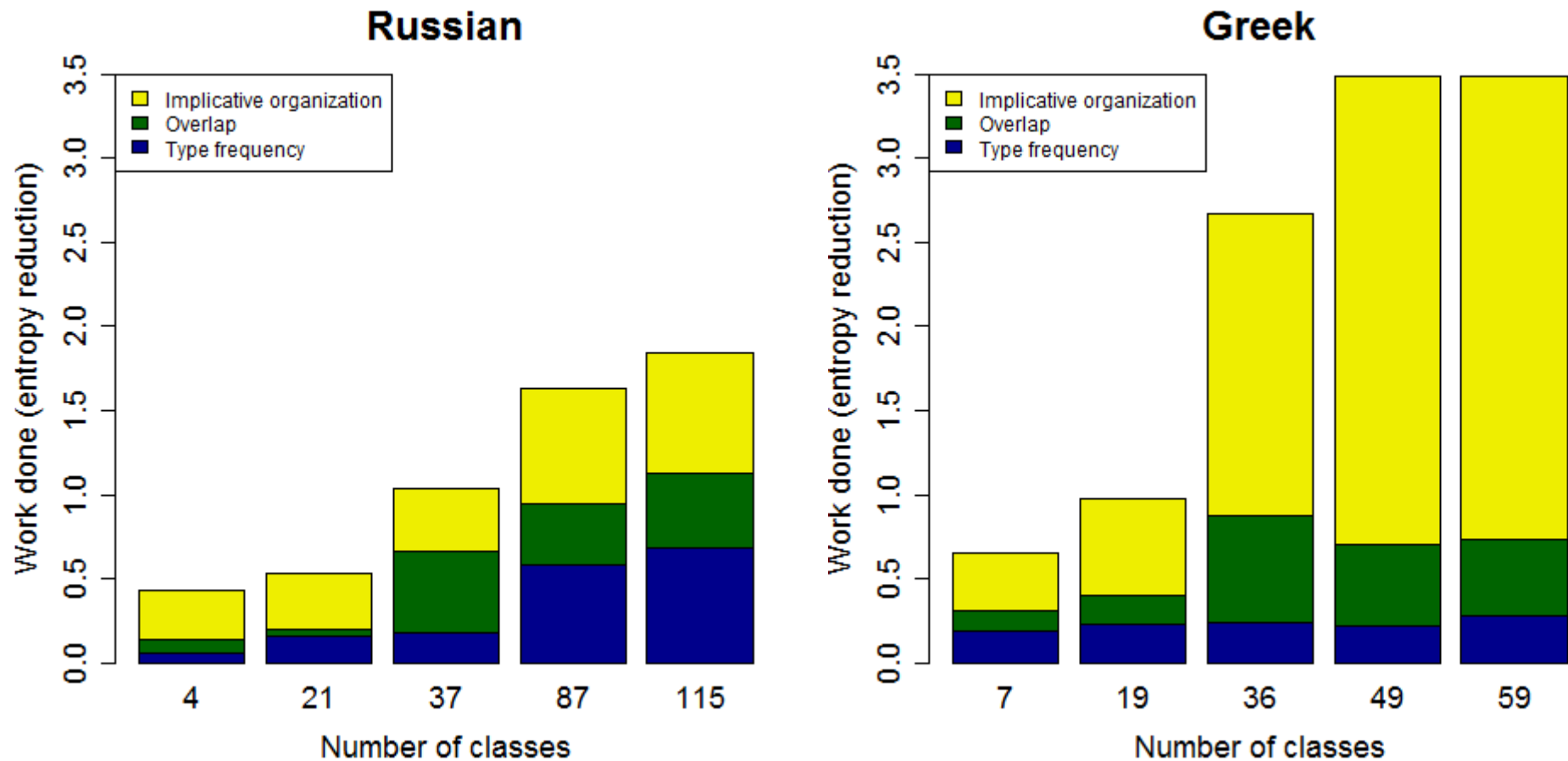
Calculating work



Calculating work



Proportion of work done in Russian and Greek



Conclusions (1)

- Assumptions about (semi)regularity affect assessments of both E-complexity and I-complexity
 - more semi-regularity → more inflection classes
 - more inflection classes → higher conditional entropy
 - Including structural and distributional sources of information (implicative structure and inflection class type frequency) lessens the differences that arise from assumptions about (ir)regularity
-

Conclusions (2)

- Both Russian and Greek exhibit lower conditional entropy than expected from chance, regardless of inflection class granularity and type frequency weighting
 - consistent with Low Entropy Conjecture
- However, the extent to which type frequency and implicative structure do work differs
 - Implicative structure plays a greater role in Greek (regardless of granularity), despite Greek having fewer paradigm cells
 - The extent to which implicative structure and type frequency are redundant sources of information differs

Ongoing work...

- Testing of the cognitive reality of implicative structure for speakers
 - Looking at the impact of inflection class granularity in more languages and how it affects the work done by different sources of information
 - Thank you!
-

References (1)

- Ackerman, Farrell, James P. Blevins, and Robert Malouf. 2009. “Parts and Wholes: Implicative Patterns in Inflectional Paradigms.” In *Analogy in Grammar: Form and Acquisition*, edited by James P. Blevins and Juliette Blevins, 54–82. New York: Oxford University Press.
- Ackerman, Farrell, and Robert Malouf. 2013. “Morphological Organization: The Low Entropy Conjecture.” *Language* 89 (3): 429–64.
- Baerman, Matthew, and Greville G. Corbett. 2012. “Stem Alternations and Multiple Exponence.” *Word Structure* 5 (1): 52–68.
- Bonami, Olivier. 2013. “Towards a Robust Assessment of Implicative Relations in Inflectional Systems”. presented at the Workshop on Computational Approaches to Morphological Complexity, Paris.

References (2)

- Corbett, Greville G. 1982. “Gender in Russian: An Account of Gender Specification and Its Relationship to Declension.” *Russian Linguistics* 6 (2): 197–232.
- Lexikó tīs koinís neoellīnikís. 1998. Thessaloniki: Triantafillidis Institute, Aristoteleio Panepistimio Thessalonikis
- Sims, Andrea D. 2015. *Inflectional Defectiveness*. Cambridge: Cambridge University Press.
- Stump, Gregory T., and Raphael A. Finkel. 2013. *Morphological Typology: From Word to Paradigm*. Cambridge University Press.
- Wurzel, Wolfgang. 1989. *Inflectional Morphology and Naturalness*. *Studies in Natural Language and Linguistic Theory*.
- Zaliznjak, Andrei A. 1977. *Grammatičeskij Slovar’ Russkogo Jazyka*. Moskva: Russkij jazyk.