On the interaction of implicative structure and type frequency in inflectional systems

Jeff Parker and Andrea D. Sims
The Ohio State University
{parker.642, sims.120}@osu.edu
Some big(gish) questions for today

- How do sources of information minimize the uncertainty associated with predicting unknown inflected forms?
  - Paradigm Cell Filling Problem, PCFP (Ackerman, Blevins, and Malouf 2009)
  - This is a ‘lankus’
  - What are these? ‘lanki’?

- Typological question: To what extent are languages similar in how sources of information interact?
PCFP and implicative structure

Low Entropy Conjecture

“…enumerative morphological complexity is effectively unrestricted, as long as the average conditional entropy, a measure of integrative complexity, is low” (Ackerman and Malouf 2013:436)

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

<table>
<thead>
<tr>
<th>SINGULAR (A)</th>
<th>virus</th>
<th>syllabus</th>
<th>corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLURAL (B)</td>
<td>viruses</td>
<td>syllabi</td>
<td>corpora</td>
</tr>
</tbody>
</table>
Implicative paradigmatic structure is …

- Not the only kind of information that can do work towards solving the PCFP
- Not necessarily independent of other info
- Low entropy can exist in the absence of implicative structure doing any work
Sources of information

- 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)
Starting point: describing the systems

- The description of the system can strongly influence analysis of system’s complexity (Bonami 2013)

- An assumption that some/many ‘irregular’ lexemes fall outside of the morphological system risks underestimating the actual complexity speakers deal with

- Rather than assume a particular analysis of the system, we empirically explore the extent to which granularity of inflectional information affects the results
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 tîs koinîs neoellînikîs* (Triantafillidis Institute 1998)
  - type frequencies from Hellenic National Corpus (hnc.ilsp.gr/en/)
# Granularity of inflection class info

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Suffixes</th>
<th>Animacy</th>
<th>Stem changes</th>
<th>Stress</th>
<th>Defectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>115</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Suffixes</th>
<th>Inflectional stress</th>
<th>Stem changes</th>
<th>Lexical stress</th>
<th>Defectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>59</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
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
Implicative work

- Our (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
  - 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)
    \[ I(A;B) = H(A) - H(A|B) \]
Overall, implicative work 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

- 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

Unconditioned entropy
Conditional entropy

Type frequency work (before implicative structure)

Type frequency work (after implicative structure)

Entropy

nonweighted
weighted

Russian; 87 classes
Calculating work

Implicative work (after type frequency)

Implicative work (before type frequency)

Entropy

Unconditioned entropy

Conditional entropy

nonweighted

weighted

Russian; 87 classes
Calculating work

Implicative work
(before type frequency)

Overlap in work

Implicative work
(after type frequency)

Russian; 87 classes
Proportion of work done in Russian and Greek

**Russian**

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Work done (entropy reduction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td>21</td>
<td>0.5</td>
</tr>
<tr>
<td>37</td>
<td>1.0</td>
</tr>
<tr>
<td>87</td>
<td>1.5</td>
</tr>
<tr>
<td>115</td>
<td>2.0</td>
</tr>
</tbody>
</table>

**Greek**

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Work done (entropy reduction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.0</td>
</tr>
<tr>
<td>19</td>
<td>0.5</td>
</tr>
<tr>
<td>36</td>
<td>1.0</td>
</tr>
<tr>
<td>49</td>
<td>1.5</td>
</tr>
<tr>
<td>59</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Conclusions

- 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…

- Expanding this type of analysis to more languages

- Testing of the cognitive reality of implicative structure for speakers
This work was supported by…

- Presidential Fellowship, OSU (Parker)
- Research start-up funds, OSU (Sims)
- Andrew W. Mellon Foundation, through the John E. Sawyer Seminar on the Comparative Study of Cultures program (grant 21200638, Co-PI Sims)
- Kenneth E. Naylor Professorship in South Slavic Linguistics, OSU
References