The structure and evolution of language

Principles of Complex Systems
CSYS/MATH 300, Fall, 2010

Prof. Peter Dodds

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.
Irregular verbs

Cleaning up English:

“Quantifying the evolutionary dynamics of language”[1]

- Exploration of how verbs with irregular conjugation gradually become regular over time.
- Comparison of verb behavior in Old, Middle, and Modern English.
Irregular verbs

![Graph showing the number of irregular verbs across different frequency bins for Old English, Middle English, and Modern English.]

- Universal tendency towards regular conjugation
- Rare verbs tend to be regular in the first place
Irregular verbs

Rates are relative.

The more common a verb is, the more resilient it is to change.
Irregular verbs

### Table 1 | The 177 irregular verbs studied

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Verbs</th>
<th>Regularization (%)</th>
<th>Half-life (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-1}$–$10^{0}$</td>
<td>come, do, find, get, give, go, know, say, see, take, think</td>
<td>0</td>
<td>38,800</td>
</tr>
<tr>
<td>$10^{-2}$–$10^{-1}$</td>
<td>be, have</td>
<td>0</td>
<td>14,400</td>
</tr>
<tr>
<td>$10^{-3}$–$10^{-2}$</td>
<td>begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie, lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, work, write</td>
<td>10</td>
<td>5,400</td>
</tr>
<tr>
<td>$10^{-4}$–$10^{-3}$</td>
<td>arise, bake, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee, float, flow, fly, fold, freeze, grind, leap, lend, lock, melt, reckon, ride, rush, shape, shine, shoot, shrink, sigh, sing, sink, slide, slip, smoke, spin, spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, weigh, wind, yell yield bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, khead, low, milk, morn, mow, prescribe, redden, reel, row, scrape, seethe, shear, shed, showe, slay, slit, smite, sow, span, spurn, sting, stink, straw, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe</td>
<td>43</td>
<td>2,000</td>
</tr>
<tr>
<td>$10^{-5}$–$10^{-4}$</td>
<td>bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, khead, low, milk, morn, mow, prescribe, redden, reel, row, scrape, seethe, shear, shed, showe, slay, slit, smite, sow, span, spurn, sting, stink, straw, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe</td>
<td>72</td>
<td>700</td>
</tr>
<tr>
<td>$10^{-6}$–$10^{-5}$</td>
<td>bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak</td>
<td>91</td>
<td>300</td>
</tr>
</tbody>
</table>

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

- Red = regularized
- Estimates of half-life for regularization.

References
‘Wed’ is next to go.
- ed is the winning rule...
Irregular verbs

- Regularization rate $\propto$ word frequency $^{-1/2}$
- Half life $\propto$ word frequency $^{1/2}$
Irregular verbs

Projecting back in time...
Word meanings

Preliminary findings on word frequency and number of meanings

- Corpus: 10,000 most frequent words from Project Gutenberg
- # meanings for each word estimated using dictionary.com
- Friends: perl, regular expressions, wget.
A. Word frequency versus rank, slope $\alpha \sim -1.2$ corresponds to a frequency distribution with $\gamma \sim 1.8$.

B. Relationship between average number of meanings and average frequency (bins are by rank, with each circle representing 500 words). Slope of 1/3 lower than Zipf’s 1/2 [3].
Word meanings

A. Meaning number as a function of word rank.

B. The three exponents combine within error:
\[ 1.2 \times \frac{1}{3} = 0.4 \approx 0.45. \]
Scaling collapse for meaning number distribution

Each curve corresponds to approximately 500 words group according to rank (1–500, 501–1000, ...).

With normalization

\[ P(n_m) = f^{-1/3} G \left( f^{-1/3} n_m \right). \]
Further work:

- Check these scalings again
- Explore alternate data sources
- Think about why meaning number might scale with frequency.
- May be an information theoretic story.
- If we add context, we may be able to use a modified version of Simon’s approach [2]
- The city story here would be that there may be many cities and towns with the same name (e.g., Springfield) with an uneven distribution in populations.
References I

Quantifying the evolutionary dynamics of language.  

On a class of skew distribution functions.  

_Human Behaviour and the Principle of Least-Effort._  
Addison-Wesley, Cambridge, MA, 1949.