Language

The structure and evolution of language

Principles of Complex Systems CSYS/MATH 300, Fall, 2011

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 Word lifespans Meanings References





200 1 of 16

Outline

Irregular verbs

Word lifespans

Meanings

References

Language

Irregular verbs Word lifespans Meanings References





Cleaning up English:

"Quantifying the evolutionary dynamics of language"^[1] Lieberman et al., Nature, Vol 449, 713-716, 2007.



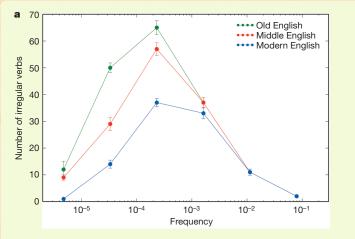
- Exploration of how verbs with irregular conjugation gradually become regular over time.
- Comparison of verb behavior in Old, Middle, and Modern English.

Irregular verbs Word lifespans Meanings References





200 3 of 16



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

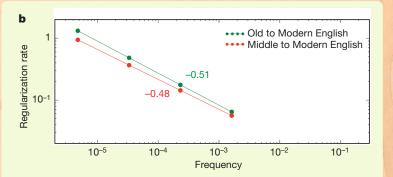
Language

Irregular verbs Word lifespans Meanings References



Dac 4 of 16

Language



Rates are relative.

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

Irregular verbs Word lifespans Meanings Beferences





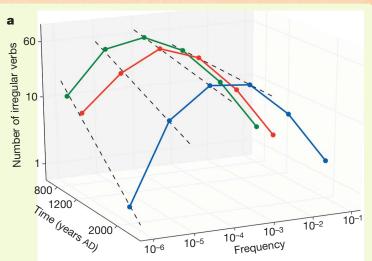
Table 1	The 177 irre	gular verbs studied
---------	--------------	---------------------

Frequency	Verbs	Regularization (%)	Half-life (yr)
10-1-1	be, have	0	38,800
10-2-10-1	come, do, find, get, give, go, know, say, see, take, think	0	14,400
10-3-10-2	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	10	5,400
10 ⁻⁴ -10 ⁻³	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, silde, silp, smoke, spin, spring, starve, stael, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, weigh, wind, kiel, vield	43	2,000
10 ⁻⁵ -10 ⁻⁴	bark, bellow, bid, bland, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, reddan, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wind, writhe	72	700
10 ⁻⁶ -10 ⁻⁵	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

177 OI Brajksh 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 frequencydependent regularization of irregular verbs becomes immediately apprent.

Red = regularized

Estimates of half-life for regularization.



Irregular verbs Word lifespans Meanings References

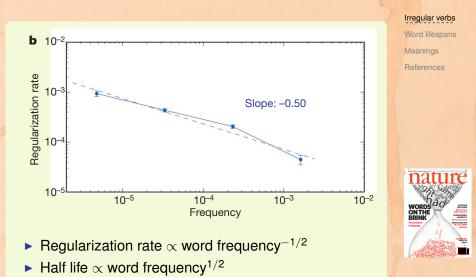


- 'Wed' is next to go.
- -ed is the winning rule...

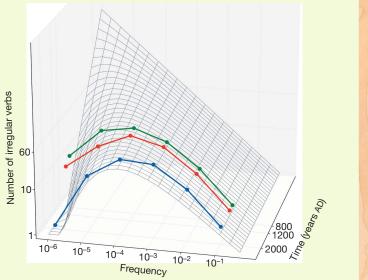
Language

VIVERSITY

DQ @ 8 of 16



Irregular verbs Word lifespans Meanings



Projecting back in time...



Irregular verbs Word lifespans

Language

Meanings References

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.

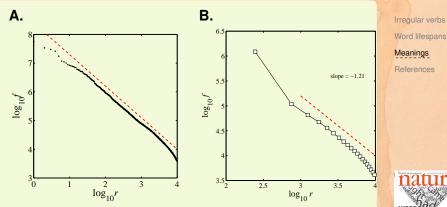




Language

DQ @ 11 of 16

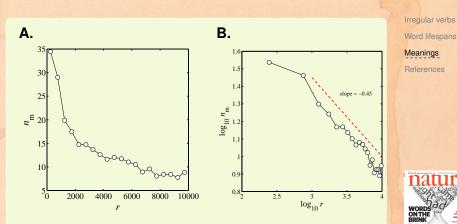
Word meanings



A. Word frequency versus rank, slope $\alpha \sim -1.2$ corresponds to 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^[4].

Language

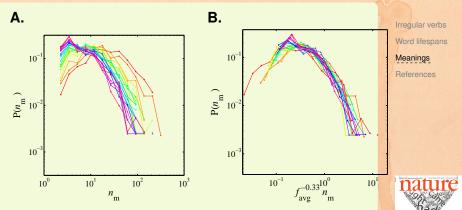
Vermont 8



- Meaning number as a function of word rank.
- ► The three exponents combine within error: $1.2 \times 1/3 = 0.4 \simeq 0.45$.

Language

DQ @ 13 of 16



- Scaling collapse for meaning number distribution
- Each curve corresponds to approximately 500 words group according to rank (1–500, 501–1000, ...).
- With normalization

$$P(n_m) = f^{-1/3}G(f^{-1/3}n_m).$$

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^[3]
- 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.

Language

Irregular verbs Word lifespans

Meanings References





References I

 [1] E. Lieberman, J.-B. Michel, J. Jackson, T. Tang, and M. A. Nowak.
Quantifying the evolutionary dynamics of language. Nature, 449:713–716, 2007. pdf (⊞)

[2] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, T. G. B. Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. A. Lieberman.

Quantitative analysis of culture using millions of digitized books.

Science Magazine, 2010. pdf (⊞)

[3] H. A. Simon.
On a class of skew distribution functions.
Biometrika, 42:425–440, 1955. pdf (⊞)

Language

Irregular verbs Word lifespans Meanings References





References II

Language

Irregular verbs Word lifespans Meanings References

[4] G. K. Zipf. Human Behaviour and the Principle of Least-Effort. Addison-Wesley, Cambridge, MA, 1949.





990 16 of 16