Evolutionary forces in language change

Christopher Ahern[1,*], Mitchell Newberry[2,*], Robin Clark[1,†], Joshua B. Plotkin[2,†]

Departments of [1]Linguistics and [2]Biology, University of Pennsylvania, 19104

[*]Joint first authors, [†]Co-corresponding authors rclark@sas.upenn.edu, jplotkin@sas.upenn.edu

Languages and genes are both transmitted from generation to generation, with opportunity for differential reproduction and survivorship of forms. Here we apply a rigorous inference framework, drawn from population genetics, to distinguish between two broad mechanisms of language change: drift and selection. Drift is change that results from stochasticity in transmission and it may occur in the absence of any intrinsic difference between linguistic forms; whereas selection is truly an evolutionary force arising from intrinsic differences – for example, when one form is preferred by members of the population. Using large corpora of parsed texts spanning from 12th century to the 21st century, we analyze three examples of grammatical changes in English: the regularization of past-tense verbs, the rise of the periphrastic 'do', and syntactic variation in verbal negation. We show that we can reject stochastic drift in favor of a selective force driving some of these language changes, but not others. The strength of drift depends on a word's frequency, and so drift provides an alternative explanation for why some words are more prone to change than others. Our results suggest an important role for stochasticity in language change, and they provide a null model against which selective theories of language evolution must be compared.

There is a rich history of exchange between linguistics and evolutionary biology, dating to the works of Darwin and Haeckel[1–3]. While the mechanisms underlying organismal evolution have been explored extensively, the forces responsible for language evolution remain unclear. Quantitative methods to infer evolutionary forces developed in population genetics have not been widely applied in linguistics, despite the recent availability of massive digital corpora[4–7].

Language change can be viewed as competition between linguistic forms, whether they are sounds, morphemes, or syntactic structures. The field of linguistics has largely assumed that any substantial change in the frequencies of alternative forms is due to selective forces acting in the language community. Linguists have identified many sources of selection that could drive language change, including language internal forces, cognitive forces, and social forces[8–20]. It is unclear, however, whether these are indeed the causative forces responsible for the changes observed in languages over centuries. To infer an evolutionary force, we must first consider whether the observed changes are due to stochasticity in transmission alone – that is, drift. Unlike selective forces, which bias a language learner towards adopting forms that are intrinsically easier to learn or more effective for communication, drift arises purely by chance: the learner chooses randomly among the sample of forms that she happens to encounter. Although drift is recognized as an important null hypothesis in population genetics[21] and cultural evolution[22,23], it has not yet been systematically tested in the context of language change.

Here we analyze three well-known grammatical changes in English: the development of the morphological past tense in contemporary American English[34,35] (spilt → spilled); the rise of the periphrastic 'do' in Early Modern English[36] (Mary ate not John's pizza → Mary did not eat John's pizza); and Jespersen's Cycle of sentential negation in Middle English[37] (Ic ne secge → I ne seye not → I say not). Our analyses are based on parsed English texts ranging from the Norman conquest of England, during the 12th century, to the early 21st century. In each case, we rigorously test whether observed linguistic changes are consistent with neutral drift, or can be attributed to selective forces.

We compare time-series of frequencies of alternative linguistic variants to a null model of drift: the neutral Wright-Fisher model from population genetics[24]. The Wright-Fisher model forms the basis for discriminating between stochastic and selective forces on genetic

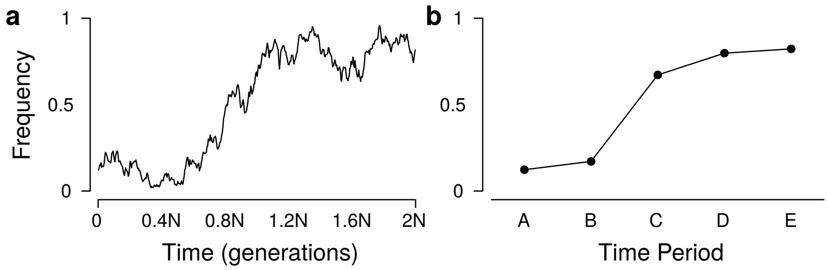

Figure 1: **A null model of language change.** Stochasticity in transmission can significantly change the frequencies of alternative linguistic forms over time, even without any intrinsic differences between forms. We use the neutral Wright-Fisher diffusion from population genetics[24], which has also been derived as a model of language learning[25], as a null model of frequency variation due to stochastic drift. Panel **a** illustrates an example time-series of frequency variation produced by this neutral null model. Although the complete time-series evidently shows random fluctuations, linguistic time-series require binning texts by time period. When this time-series is binned (panel **b**), it produces a characteristic S-shaped curve that is often accepted as evidence of a selective force favoring one linguistic variant over others[23,26–32]. This simple example illustrates the need to test hypotheses against a null model to infer the presence of evolutionary forces in language change[23].

variants in a replicating population of size N; and the same model has been derived for linguistic change under Bayesian learning[25], where the inverse of the population size parameter N governs the amount of stochasticity in transmission. Importantly, even in the neutral case the Wright-Fisher model can produce large frequency changes that may appear, prima facie, to be the result of selection. It can even produce the characteristic, logistic curve of one variant replacing another in binned time-series (Figure 1) that has typically been accepted as evidence of selective forces in language change[23,26–32].

The population size parameter N is unknown to us. And so to infer the action of selection we must show that observed linguistic changes are inconsistent with neutral drift, regardless of N. A stringent statistical test to reject this composite null hypothesis ($s = 0$, N arbitrary) has recently been developed, called the Frequency Increment Test (FIT)[33]. The Frequency Increment Test compares the frequency changes observed between sampled time points to the expectations under drift, with $s = 0$ and arbitrary N. The test is valid for a large class of neutral null models: all those with the same diffusion limit as the Wright-Fisher model. For each linguistic time-series we can also estimate the most likely population size, N, and the most likely selection coefficient, s, favoring one linguistic variant over another[33].

We began by analyzing past-tense verb conjugation in contemporary American English. One view contends that irregular past-tense forms should regularize over time[34,38,39], for reasons of economy, phonological analogy, or cognitive ease[15,34,40]. In this view an irregular past-tense form, such as "wove," should be selectively replaced by the regular form, "weaved", produced by adding the voiced alveolar suffix "-ed" to the verb. Although there is substantial support for past-tense regularization, especially for rare words over long timescales[4,41], most studies have simply reported trends in usage frequencies over time, and several apparent exceptions have been noted within Modern English[4,35].

We collected all past tense verb tokens from the Corpus of Historical American English[7], comprising over four million words from $> 100,000$ texts of American English, between the years 1810 and 2009. The corpus was parsed for part of speech using the CLAWS tagger[42]. Among all tokens assigned the simple past tense as the most likely part of speech, we retained only those lemmas with two variants each occurring at least 50 times in the corpus[43]. This produced 704,081 tokens in total which provide frequency trajectories of regular versus irregular past-tense variants for 36 verbs (Figure 2)[44].

We used these linguistic time-series to determine whether selection or drift is driving

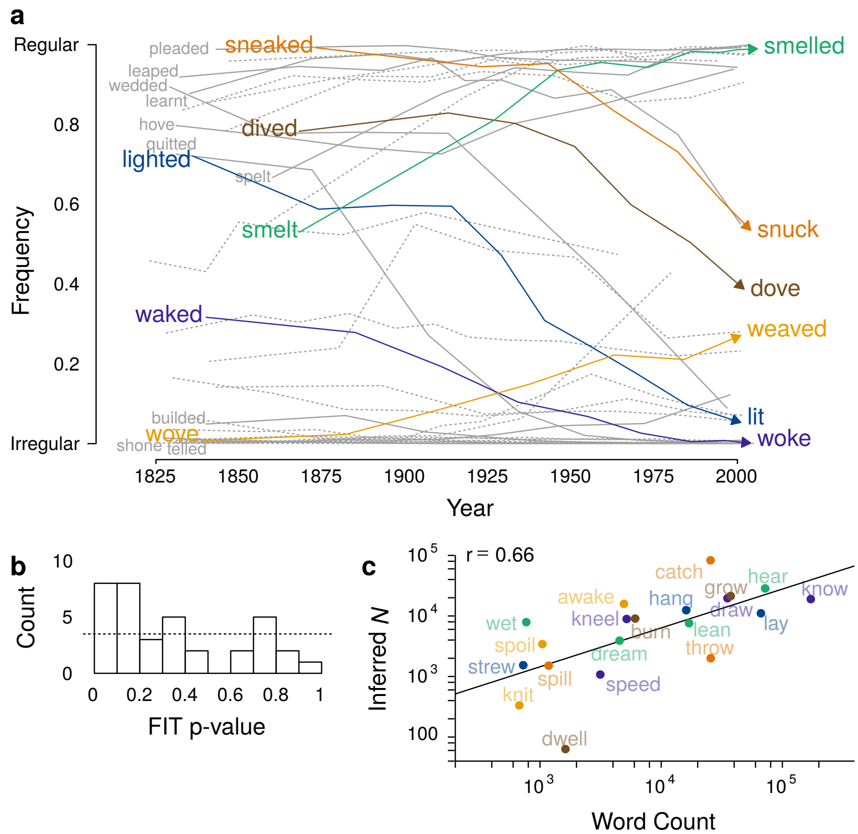

Figure 2: **Verb regularization and irregularization.** We analyzed 36 verbs with multiple past-tense forms appearing in the Corpus of Historical American English[7]. Six of these verbs experience selection for either regularization or irregularization, each with $p < 0.05$ by the Frequency Increment Test of selection[33] (**a**, colored lines). The regular form is favored in two of these cases, and the irregular form in the remaining four cases. Ten more verbs, of which 4 are regularizing (**a**, solid gray lines), are significant at specificity $1 - \alpha = 0.8$, with a false discovery rate of 45%. The distribution of nominal FIT p-values (**b**) is non-uniform (Kolmogorov-Smirnov $p = 0.002$), which confirms that some verbs experience selection. Among the remaining 20 verbs, for which we fail to reject neutrality (**a**, dashed gray lines), the log inferred population size correlates with log token count in the corpora (**c**, $r = 0.66, p = 0.002$).

changes in past-tense conjugation in Modern English (Figure 2). We computed a two-sided p-value by the Frequency Increment Test for each of the 36 verbs with irregular variants. For six of these verbs we can reject neutral drift for all population sizes N, with nominal $p < 0.05$. Contrary to the standard linguistic expectation, in four of these cases we infer selection towards the *irregular* variant (dived→dove, waked→woke, lighted→lit, sneaked→snuck), whereas only two cases exhibit regularization (wove→weaved, smelt→smelled). Moreover, among the 16 verbs we identify as possibly under selection, at specificity $1 - \alpha = 0.8$ with a false discovery rate of 45%, the majority exhibit irregularization (Figure 2). Examples of irregularization have been noted previously, based on trends in usage[4,35], whereas here we have definitively inferred an evolutionary force operating on Modern English verb conjugation (Kolmogorov-Smirnov p= 0.002, Figure 2b)[45].

Our analysis of irregular verbs illustrates the value of a null model for language change. Notably, for some verbs previously described as undergoing regularization[4,35], such as spilt→spilled and burnt→burned, we cannot reject neutral drift, even with sample sizes

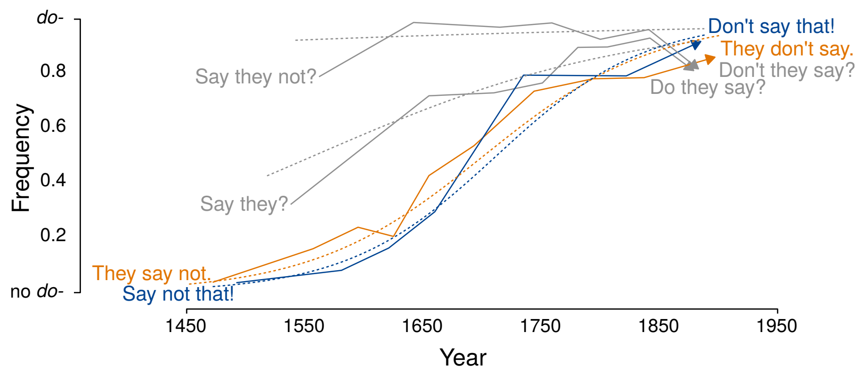

Figure 3: **The rise of the periphrastic 'do' in British English** The use of 'do' as an auxiliary verb first arose in the context of interrogative sentences (gray). However, we cannot reject drift for either affirmative interrogatives (4,401 cases in parsed corpora, FIT $p = 0.23$) or negative interrogatives (513 cases, FIT $p = 0.77$). Subsequently, the frequency of *do*-support rose rapidly in negative declarative (11,286 examples) and negative imperative (953 examples) sentences, where we detect selection (FIT $p = 0.01$ and $p = 0.02$, respectively). Dotted lines plot the logistic curve with slope determined by the maximum-likelihood selection coefficient inferred in each context. Thus, *do*-support may have arisen by chance in interrogative statements, setting the stage for selection to drive the evolution of *do*-support in other grammatical contexts.

sufficient to reject drift[46]. Conversely, we identify selection towards irregularization in some cases, such as wedded→wed, that were previously predicted to be regularizing based on long-term trends[41]. We even identify incipient grammatical changes, such as wove→weaved, in which the selected variant is in the minority at present, but predicted by our analysis to eventually replace the ancestral form.

Many studies have found that common words are more robust to change over time than rare words[29,31,41,47]. Prevailing explanations for this phenomenon are based on selection – for example, purifying selection against novel variants is assumed to be stronger for common words[48]. We propose an alternative and complementary theory based on drift: more common words, whether under selection to change or not, experience less stochasticity in transmission. This theory would predict less variability over time among alternative variants of common words, even in the absence of selection towards one form or another. Indeed, we find that for those past-tense verbs consistent with neutral drift the most-likely inferred population size is correlated with the word's frequency in the corpus ($r = 0.66, p = 0.002$, Figure 2c). Thus, the tendency for common words to resist frequency variation[29,31,41,47] extends even to cases where we detect no selective pressure for grammatical change. The relationship between word frequency and the strength of drift also predicts that different linguistic substitutions will occur by different mechanisms: for rare words substitutions are more likely to occur by random chance, whereas for common words substitutions are more likely to be caused by selective forces.

Turning next to the rise of the periphrastic 'do' in Early Modern English, we collected tokens of *do*-support from the York-Helsinki Parsed Corpus of Early English Correspondence (1400-1700), the Penn-Helsinki Parsed Corpus of Early Modern English (1500-1700), and the Penn Parsed Corpus of Modern British English (1710-1910), which include roughly seven million parsed words from 1,220 texts of British English. We extracted 16,072 tokens[49] of *do*-support in the context of affirmative questions, negative questions, negative declaratives, and negative imperatives. Over the course of these four centuries, for example, we see "You asked not." become "You did not ask." and we see "Asked you a question?" become "Did you ask a question?".

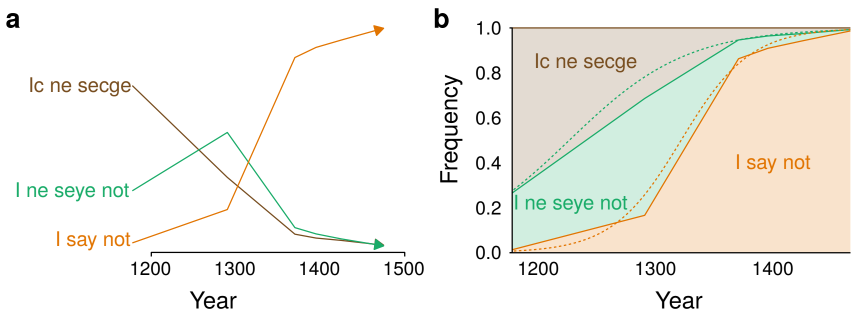

Figure 4: **Evolution of sentential negation.** In English and French, pre-verbal negation (e.g. Old English "Ic ne secge") gave way to embracing bipartite negation (Middle English "I ne seye not") and then to post-verbal negation (Early Modern English "I say not"), in a pattern known as Jespersen's Cycle. We show the frequencies these forms among 5,475 instances of negation from 52 written works in the Penn-Helsinki Parsed Corpus of Middle English (**a**). We infer selection for bipartite and post-verbal negation in the background of pre-verbal forms (FIT $p = 0.03$, **b**, green lines) and selection for post-verbal negation in a mixed population of pre-verbal and bipartite forms (FIT $p = 0.04$, **b**, orange lines). Dotted lines indicate logistic curves corresponding to maximum-likelihood selection coefficients.

The rise of the periphrastic 'do' in British English was more rapid in negative declarative and imperative statements, where we reject the neutral null model ($p < 0.02$), than in interrogative statements, where we fail to reject drift (Figure 3). It may seem natural that selection for an auxiliary verb should operate in all grammatical contexts equally[50], and yet the extensive parsed corpora available do not support this hypothesis. Our analysis suggests an alternative scenario: the periphrastic 'do' first drifted by chance to high frequency in interrogative statements, which then set the stage for subsequent selection in declarative and imperative statements, for reasons of grammatical consistency or cognitive ease.

Finally, we studied the evolution of sentential negation from the 12th to the 16th century, based on 5,475 negative declaratives extracted from the Penn Parsed Corpus of Middle English. We observe pre-verbal negation (e.g. "Ic ne secge") giving way to embracing bipartite negation ("I ne seye not") and then finally to post-verbal negation ("I say not"), in a pattern known as Jespersen's Cycle[51]. For both the transitions that form this cycle we can definitively reject neutral drift in favor of a selective force changing the formation of English negation (Figure 4). This quantitative analysis supports longstanding linguistic hypotheses about forces driving verbal negation, such as a tendency for speakers to use more emphatic forms of negation[52–55] which then becomes normalized through frequent use by "pooling action"[52–59].

The field of comparative linguistics has long benefited from quantitative techniques drawn from phylogenetics, producing a detailed and nuanced characterization of the relationships between different languages[11,48,60,61]. By contrast, theories of how a given language changes over short timescales have not been been subjected to quantitative inference techniques. And yet, changes within a language must be the origin of divergences and differentiation between languages. Now, the combination of massive digital corpora along with time-series techniques from population genetics allow us to rigorously distinguish hypotheses about the causes of language change from stochastic drift, laying a foundation for empirically testable theories of language evolution.

—

1. August Schleicher. Die ersten spaltungen des indogermanischen urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur*, 1853:786–787, 1853.
2. Charles Darwin. *The descent of man, and selection in relation to sex.* Murray, 1888.

3. Ernst Haeckel. *Natürliche Schöpfungsgeschichte*. Gg. Reimer, 1868.
4. Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, et al. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.
5. Yuri Lin, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman, and Slav Petrov. Syntactic annotations for the google books ngram corpus. In *Proceedings of the ACL 2012 system demonstrations*, pages 169–174. Association for Computational Linguistics, 2012.
6. Mark Davies. The corpus of historical american english: 400 million words, 1810-2009, 2010.
7. Mark Davies. Expanding horizons in historical linguistics with the 400-million word corpus of historical american english. *Corpora*, 7(2):121–157, 2012.
8. Hermann Paul. *Principles of the History of Language*. Macmillan & co, New York, 1890.
9. Leonard Bloomfield. *Language*. Holt, Rinehart and Winston, 1933.
10. Otto Jespersen. Language: Its nature, origin and development. *New York: Holt*, 1922.
11. W. Tecumseh Fitch. *The Evolution of Language*. Cambridge University Press, Cambridge, U.K., 2010.
12. Hans Heinrich Hock. *Principles of Historical Linguistics*. Mouton de Gruyter, New York, second edition, 1991.
13. Hans Heinrich Hock and Brian D. Joseph. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Mouton de Gruyter, New York, 2009.
14. Henry M Hoenigswald. The annus mirabilis 1876 and posterity. *Transactions of the Philological Society*, 76(1):17–35, 1978.
15. Roman Jakobson. *On Language*. Harvard University Press, Cambridge, MA, 1995.
16. William Labov. *Principles of linguistic change Volume 1: Internal factors*. Blackwell, 1994.
17. William Labov. *Principles of linguistic change Volume 2: Social factors*. Blackwell, 2001.
18. William Labov. *Principles of linguistic change Volume 3: Cognitive and Cultural Factors*. Blackwell, 2010.
19. William S.Y Wang. Competing changes as a cause of residue. *Language*, pages 9–25, 1969.
20. Uriel Weinreich, William Labov, and Marvin I. Herzog. Empirical foundations for a theory of language change. In Winfred P. Lehman and Yakov Malkiel, editors, *Directions for Historical Linguistics: A Symposium*, pages 95–188. University of Texas Press, Austin, TX, 1968.
21. Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931.
22. R. Alexander Bentley, Matthew W. Hahn, and Stephen J. Shennan. Random drift and culture change. *Proceedings of the Royal Society B*, 271:1443–1450, 2004. DOI 10.1098/rspb.2004.2746.
23. Richard A. Blythe. Neutral evolution: a null model for language dynamics. *Advances in Complex Systems*, 15(03n04):1150015 (20 pages), 2012. doi:10.1142/S0219525911003414.
24. James F Crow, Motoo Kimura, et al. *An introduction to population genetics theory.* Harper & Row, 1970.
25. Florencia Reali and Thomas L Griffiths. Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1680):429–436, 2010.
26. Gareth J. Baxter, Richard A. Blythe, William Croft, and Alan J. McKane. Modeling language change: an evalution of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(2):257–296, 2009.
27. Richard A Blythe and William Croft. S-curves and the mechanisms of propagation in language change. *Language*, 88(2):269–304, 2012.
28. L.L. Cavalli-Sforza and M.W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, Princeton, NJ, 1980.
29. W. Tecumseh Fitch. Linguistics: An invisible hand. *Nature*, 449:665–667, 2007.
30. Henri Kauhanen. Neutral change. *Journal of Linguistics*, pages 1–32, 2016.
31. FLorencia Reali and Thomas L. Griffiths. Evolution of frequency distributions: Relating regularization to inductive biases through iterate learning. *Cognition*, 111:317–328, 2009.
32. Peter Trudgill. *New-Dialect Formation: The Inevitability of Colonial Englishes*. University of Edinburgh Press, Edinburgh, Scotland, 2004.
33. Alison F Feder, Sergey Kryazhimskiy, and Joshua B Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.
34. Steven Pinker. Rules of language. *Science*, 253(5019):530–535, 1991.
35. Christine F Cuskley, Martina Pugliese, Claudio Castellano, Francesca Colaiori, Vittorio Loreto, and Francesca Tria. Internal and external dynamics in language: evidence from verb regularity in a historical corpus of english. *PloS one*, 9(8):e102882, 2014.
36. Alvar Ellegård. The auxiliary do: The establishment and regulation of its use in english. 1953.
37. Otto Jespersen. *Negation in English and other Languages*. Host, 1917.
38. Joan Bybee. From usage to grammar: the mind's response to repetition. *Language*, 82(4):711–733, 2006.
39. Greville Corbett, Andrew Hippisley, Dunstan Brown, and Paul Marriott. Frequency, regularity and the paradigm: A perspective from russian on a complex relation. In Joan Bybee and Paul Hopper, editors, *Frequency and the emergence of linguistic structure*, pages 201–226. Amsterdam: John Benjamins, 2001.
40. George K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Cambridge, MA, 1949.

41. Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang, and Martin A. Nowak. Quantifying the evolutionary dynamics of language. *Nature*, 449:713–716, October 2007. coi:10.1038/nature06137.
42. Roger Garside. The claws word-tagging system. 1987.
43. We also excluded lemmas with temporal variation caused by spelling conventions (e.g cancelled versus canceled), lemmas with semantic ambiguities (e.g. bear versus bore, wind), and lemmas with multiple irregular variants (e.g. begin, bid, drink, ring).
44. For each verb, we binned its tokens into date ranges of variable lengths to ensure roughly the same number of tokens per bin, setting the number of bins equal to the logarithm of the number of tokens, rounded up. We applied Laplace (add-one) smoothing to counts with only one of the two variants present, in order to remove apparent absorption events[33].
45. Results of the Frequency Increment Test for selection on these 36 verbs are not driven by power or sample size. There is no significant difference in the mean number of tokens among the 16 verbs with FIT $p < 0.2$ compared to the remaining verbs (Mann-Whitney test, $p = 0.29$).
46. The number of tokens for spill (1,178) exceeds that of four verbs with FIT $p < 0.2$. Likewise the number of tokens for burn (6,097) exceeds that of eight verbs with FIT $p < 0.2$. See Table S1.
47. Joan Bybee and Paul Hopper. Introduction to frequency and the emergence of linguistic structure. In *Frequency and the emergence of linguistic structure*, pages 1–24. John Benjamins, Amsterdam, 2001.
48. Mark Pagel, Quentin D Atkinson, and Andrew Meade. Frequency of word-use predicts rates of lexical evolution throughout indo-european history. *Nature*, 449(7163):717–720, 2007.
49. 2009. URL `http://copussearch.sourceforge.net/develops.html`.
50. Anthony S Kroch. Reflexes of grammar in patterns of language change. *Language variation and change*, 1(03):199–244, 1989.
51. Otto Jespersen. Negation in english and other languages. 1917.
52. Christopher Ahern. *Cycles and Stability in Linguistic Signaling*. PhD thesis, University of Pennsylvania, 2015.
53. Phillip Wallage. Functional differentiation and grammatical competition in the english jespersen cycle. *Journal of Historical Syntax*, 2(1), 2013.
54. Osten Dahl. Typology of sentence negation. *Linguistics*, 17(1-2):79–106, 1979.
55. Osten Dahl. *Inflationary effects in language and elsewhere*. John Benjamins Publishing Company, Philadelphia, 2001.
56. Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.
57. Paul Kiparsky and Cleo Condoravdi. Tracking Jespersen's cycle. In Mark Janse, editor, *Proceedings of the 2nd International Conference of Modern Greek Dialects and Linguistic Theory*. University of Patras, 2006.
58. Scott A. Schwenter. Fine-tuning Jespersen's cycle. In Betty T. Birner and Gregory Ward, editors, *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honour of Laurence R. Horn*, pages 327–344. John Benjamins, Philadelphia, 2006.
59. Phillip Wallage. Jespersen's cycle in middle english: parametric variation and grammatical competition. *Lingua*, 118(5):643–674, 2008.
60. Mark Pagel. Linguistics and the evolution of human languages. In J. Losos, editor, *The Princeton Guide to Evolution*, chapter Chapter VIII.9, pages 786–794. Princeton University Press, Princeton, 2013.
61. Mark Pagel. Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, 10(6): 405–415, 2009.

lemma	regular	irregular	count	FIT p-value	Inferred N	Inferred s
light	lighted	lit	8,869	0.003	770	-0.024
wake	waked	woke	7,186	0.006	4,005	-0.024
sneak	sneaked	snuck	898	0.02	29	-0.039
weave	weaved	wove	907	0.02	126	0.013
dive	dived	dove	1,036	0.05	710	-0.018
smell	smelled	smelt	4,555	0.05	1,708	0.038
wed	wedded	wed	211	0.08	154	-0.026
quit	quitted	quit	2,734	0.10	430	-0.048
spell	spelled	spelt	962	0.10	2,055	0.025
shine	shined	shone	8,424	0.10	989	0.019
tell	telled	told	129,041	0.12	314,100	-0.050
leap	leaped	leapt	8,336	0.13	346	-0.019
build	builded	built	9,109	0.14	8,602	-0.063
plead	pleaded	pled	3,810	0.14	3,756	-0.006
learn	learned	learnt	18,851	0.16	6,918	0.027
heave	heaved	hove	2,392	0.20	3,663	0.008
dwell	dwelled	dwelt	1,621	0.21	63	0.033
wet	wetted	wet	770	0.24	7,903	-0.009
burn	burned	burnt	6,097	0.24	9,045	0.014
kneel	kneeled	knelt	5,185	0.30	8,912	-0.006
spoil	spoiled	spoilt	1,045	0.30	3,431	0.018
awake	awaked	awoke	4,926	0.32	15,860	-0.036
know	knowed	knew	171,518	0.34	18,980	-0.007
speed	speeded	sped	3,142	0.39	1,077	-0.002
lay	laid	lay	66,436	0.45	11,070	-0.002
spill	spilled	spilt	1,178	0.47	1,509	0.031
throw	throwed	threw	25,612	0.61	2,001	-0.011
strew	strewed	strew	727	0.66	1,537	0.000
hear	heared	heard	72,052	0.72	28,530	-0.033
knit	knitted	knit	675	0.76	333	-0.010
hang	hanged	hung	16,079	0.77	12,450	-0.012
dream	dreamed	dreamt	4,530	0.77	3,907	0.005
catch	catched	caught	25,529	0.78	83,060	-0.021
lean	leaned	leant	16,981	0.85	7,594	0.013
draw	drawed	drew	35,213	0.87	19,620	-0.026
grow	growed	grew	37,444	0.93	21,340	-0.013

Table S1: We analyzed 36 verbs with multiple past-tense forms appearing in the Corpus of Historical American English[7]. The table shows the number times each lemma occurs in the corpus, the FIT p-value for rejecting the neutral null hypothesis, and the most likely inferred population size N and selection coefficient s.