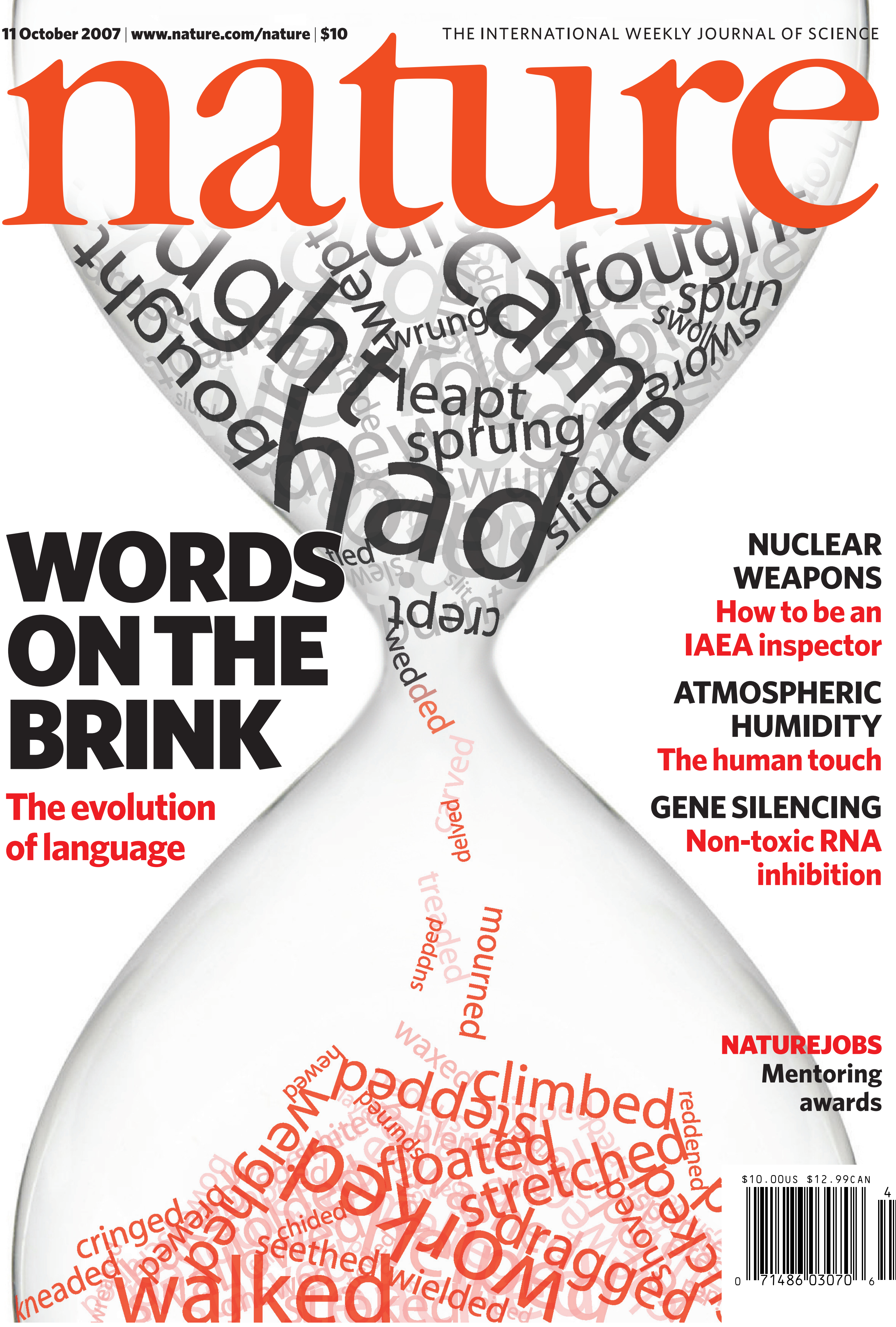
11 October 2007 | www.nature.com/nature | $10

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature



WORDS ON THE BRINK

The evolution of language

NUCLEAR WEAPONS
How to be an IAEA inspector

ATMOSPHERIC HUMIDITY
The human touch

GENE SILENCING
Non-toxic RNA inhibition

NATUREJOBS
Mentoring awards

$10.00US $12.99CAN

41

0 71486 03070 6

Vol 449 | 11 October 2007 | doi:10.1038/nature06137

Quantifying the evolutionary dynamics of language

Erez Lieberman[1,2,3]*, Jean-Baptiste Michel[1,4]*, Joe Jackson[1], Tina Tang[1] & Martin A. Nowak[1]

Human language is based on grammatical rules[1–4]. Cultural evolution allows these rules to change over time[5]. Rules compete with each other: as new rules rise to prominence, old ones die away. To quantify the dynamics of language evolution, we studied the regularization of English verbs over the past 1,200 years. Although an elaborate system of productive conjugations existed in English's proto-Germanic ancestor, Modern English uses the dental suffix, '-ed', to signify past tense[6]. Here we describe the emergence of this linguistic rule amidst the evolutionary decay of its exceptions, known to us as irregular verbs. We have generated a data set of verbs whose conjugations have been evolving for more than a millennium, tracking inflectional changes to 177 Old-English irregular verbs. Of these irregular verbs, 145 remained irregular in Middle English and 98 are still irregular today. We study how the rate of regularization depends on the frequency of word usage. The half-life of an irregular verb scales as the square root of its usage frequency: a verb that is 100 times less frequent regularizes 10 times as fast. Our study provides a quantitative analysis of the regularization process by which ancestral forms gradually yield to an emerging linguistic rule.

Natural languages comprise elaborate systems of rules that enable one speaker to communicate with another[7]. These rules serve to simplify the production of language and enable an infinite array of comprehensible formulations[8–10]. However, each rule has exceptions, and even the rules themselves wax and wane over centuries and millennia[11,12]. Verbs that obey standard rules of conjugation in their native language are called regular verbs[13]. In the Modern English language, regular verbs are conjugated into the simple past and past-participle forms by appending the dental suffix '-ed' to the root (for instance, infinitive/simple past/past participle: talk/talked/talked). Irregular verbs obey antiquated rules (sing/sang/sung) or, in some cases, no rule at all (go/went)[14,15]. New verbs entering English universally obey the regular conjugation (google/googled/googled), and many irregular verbs eventually regularize. It is much rarer for regular verbs to become irregular: for every 'sneak' that 'snuck' in[16], there are many more 'flews' that 'flied' out.

Although less than 3% of modern verbs are irregular, the ten most common verbs are all irregular (be, have, do, go, say, can, will, see, take, get). The irregular verbs are heavily biased towards high frequencies of occurrence[17,18]. Linguists have suggested an evolutionary hypothesis underlying the frequency distribution of irregular verbs: uncommon irregular verbs tend to disappear more rapidly because they are less readily learned and more rapidly forgotten[19,20].

To study this phenomenon quantitatively, we studied verb inflection beginning with Old English (the language of *Beowulf*, spoken around AD 800), continuing through Middle English (the language of Chaucer's *Canterbury Tales*, spoken around AD 1200), and ending with Modern English, the language as it is spoken today. The modern '-ed' rule descends from Old English 'weak' conjugation, which applied to three-quarters of all Old English verbs[21]. The exceptions—ancestors of the modern irregular verbs—were mostly members of the so-called 'strong' verbs. There are seven different classes of strong verbs with exemplars among the Modern English irregular verbs, each with distinguishing markers that often include characteristic vowel shifts. Although stable coexistence of multiple rules is one possible outcome of rule dynamics, this is not what occurred in English verb inflection[22]. We therefore define regularity with respect to the modern '-ed' rule, and call all these exceptional forms 'irregular'.

We consulted a large collection of grammar textbooks describing verb inflection in these earlier epochs, and hand-annotated every irregular verb they described (see Supplementary Information). This provided us with a list of irregular verbs from ancestral forms of English. By eliminating verbs that were no longer part of Modern English, we compiled a list of 177 Old English irregular verbs that remain part of the language to this day. Of these 177 Old English irregulars, 145 remained irregular in Middle English and 98 are still irregular in Modern English. Verbs such as 'help', 'grip' and 'laugh', which were once irregular, have become regular with the passing of time.

Next we obtained frequency data for all verbs by using the CELEX corpus, which contains 17.9 million words from a wide variety of textual sources[23]. For each of our 177 verbs, we calculated the frequency of occurrence among all verbs. We subdivided the frequency spectrum into six logarithmically spaced bins from 10^{-6} to 1. Figure 1a shows the number of irregular verbs in each frequency bin. There are only two verbs, 'be' and 'have', in the highest frequency bin, with mean frequency $>10^{-1}$. Both remain irregular to the present day. There are 11 irregular verbs in the second bin, with mean frequency between 10^{-2} and 10^{-1}. These 11 verbs have all remained irregular from Old English to Modern English. In the third bin, with a mean frequency between 10^{-3} and 10^{-2}, we find that 37 irregular verbs of Old English all remained irregular in Middle English, but only 33 of them are irregular in Modern English. Four verbs in this frequency range, 'help', 'reach', 'walk' and 'work', underwent regularization. In the fourth frequency bin, 10^{-4} to 10^{-3}, 65 irregular verbs of Old English have left 57 in Middle and 37 in Modern English. In the fifth frequency bin, 10^{-5} to 10^{-4}, 50 irregulars of Old English have left 29 in Middle and 14 in Modern English. In the sixth frequency bin, 10^{-6} to 10^{-5}, 12 irregulars of Old English decline to 9 in Middle and only 1 in Modern English: 'slink', a verb that aptly describes this quiet process of disappearance (Table 1).

Plotting the number of irregular verbs against their frequency generates a unimodal distribution with a peak between 10^{-4} and 10^{-3}. This unimodal distribution again demonstrates that irregular verbs are not an arbitrary subset of all verbs, because a random subset of verbs (such as all verbs that contain the letter 'm') would follow Zipf's law, a power law with a slope of -0.75 (refs 24,25).

[1]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, [2]Department of Applied Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work

©2007 Nature Publishing Group

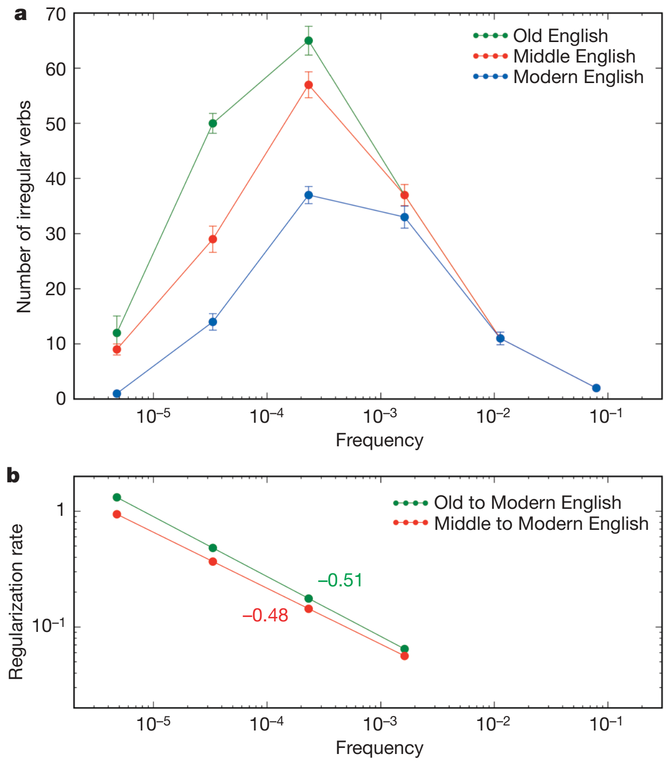


Figure 1 | Irregular verbs regularize at a rate that is inversely proportional to the square root of their usage frequency. a, The evolution of 177 verbs from Old English (green) over time, through Middle English (red) and Modern English (blue). The fraction remaining irregular in each bin decreases as the frequency decreases. The frequency shown is that of the modern descendant, and was computed using the CELEX corpus. Error bars indicate standard deviation and were calculated using the bootstrap method. **b,** The regularization rate of irregular verbs as a function of frequency. The relative regularization rates obtained by comparing Old versus Modern English (green) and Middle versus Modern English (red) scale linearly on a log–log plot with a downward slope of nearly one-half. The regularization rate and the half-life scale with the square root of the frequency.

Four of our six frequency bins, those between 10^{-6} and 10^{-2}, allow us to estimate the relative regularization rates of irregular verbs. Calculating the relative regularization rates of verbs of different frequencies is independent of time, which makes the dating of Old and Middle English irrelevant for this calculation. We can plot regularization rate versus frequency and fit a straight line in a log–log plot (Fig. 1b). By comparing Old and Modern English we obtain a slope of about −0.51. Therefore, an irregular verb that is 100 times less frequent is regularized 10 times as fast. In other words, the half-life of irregular verbs is proportional to the square root of their frequency. By comparing Middle and Modern English we find a slope of about −0.48, consistent with the previous result. Both comparisons show that low-frequency irregular verbs are selectively forgotten.

Figure 2a shows the exponential decay of the irregular verbs in the four frequency bins between 10^{-6} and 10^{-2} as a function of time. From these data, which depend on the dating of Old and Middle English, we can estimate actual half-lives of the irregular verbs in different frequency bins. Irregular verbs that occur with a frequency between 10^{-6} and 10^{-5} have a half-life of about 300 years, whereas those with a frequency between 10^{-4} and 10^{-3} have a half-life of 2,000 years. If we fit half-life versus frequency with a straight line in a log–log plot, we obtain a slope of 0.50, which again suggests that the half-life of irregular verbs is proportional to approximately the square root of their frequency (Fig. 2b). It is noteworthy that various methods of fitting the data give the same results.

We cannot directly determine the regularization rate for frequency bins above 10^{-2}, because regularization is so slow that no event was observed in the time span of our data; however, we can extrapolate. For instance, the half-life of verbs with frequencies between 10^{-2} and 10^{-1} should be 14,400 years. For these bins, the population is so small and the half-life so long that we may not see a regularization event in the lifetime of the English language.

To test whether the dynamics within individual competing rules were captured by our global analysis, we studied the decay of individual classes of strong verbs (for example, hit/hit/hit, hurt/hurt/hurt; draw/drew/drawn, grow/grew/grown)[26]. Although our resolution is limited by the small sample size, exponential decay is once again observed, with similar exponents (see Supplementary Fig. 1). Like a Cheshire cat, dying rules vanish one instance at a time, leaving behind a unimodal frown.

Because adequate corpora of Old and Middle English do not exist, we have estimated the frequency of an irregular verb of Old and Middle English by the frequency of the corresponding (regular or irregular) verb of Modern English[27]. A large fraction of verbs would have had to change frequency by several orders of magnitude to

Table 1 | The 177 irregular 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, 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	43	2,000
10^{-5}–10^{-4}	bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe	72	700
10^{-6}–10^{-5}	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

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.

©2007 Nature Publishing Group

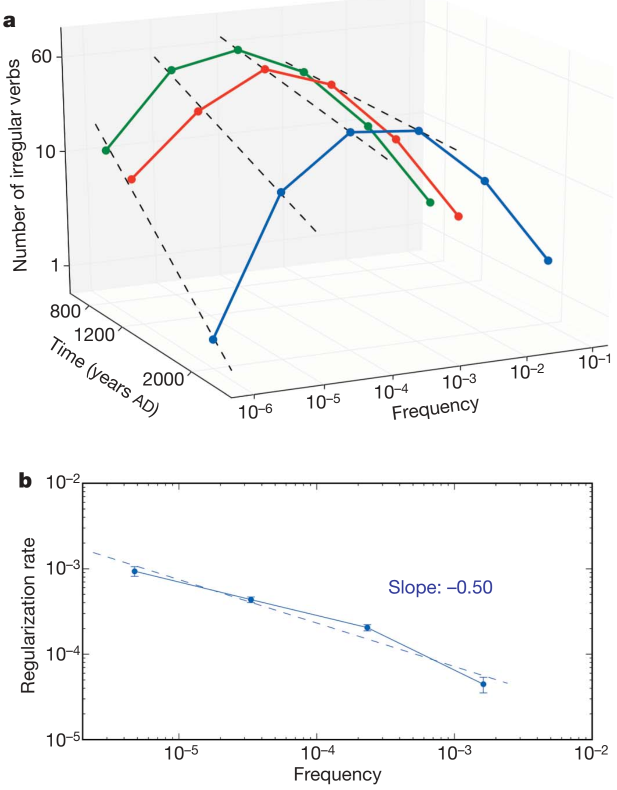


Figure 2 | Irregular verbs decay exponentially over time. a, Specifying approximate dates of Old and Middle English allows computation of absolute regularization rates. Regularization rates increase as frequencies decrease, but are otherwise constant over time. **b**, Absolute rates of regularization are shown as a function of frequency. Error bars indicate standard deviation and were calculated using the bootstrap method. The square-root scaling is obtained again.

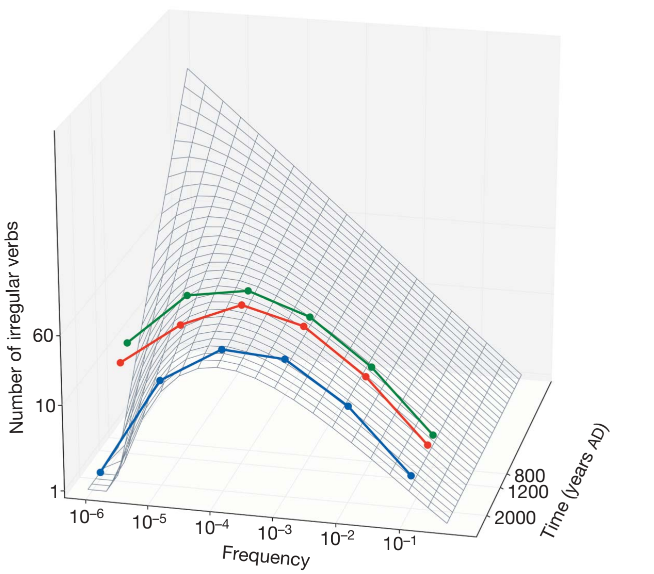


Figure 3 | Extrapolating forward and backward in time using the observation that regularization rate scales as the square root of frequency. The differential system is exactly solvable and the solution fits all three observed distributions. As we move backward in time, the distribution of irregular verbs approaches the zipfian distribution characteristic of random sets of words. The distribution for exceptions to the '-ed' rule became non-random because of frequency-dependent regularization due to selective pressure from the emerging rule.

interfere with the effects observed. To verify that large changes in frequency are rare, we compared frequency data from CELEX with frequencies drawn from the largest available corpus of Middle English texts[28]. Of 50 verbs, only 5 had frequency changes greater than a factor of 10 (Supplementary Fig. 2).

Our analysis covers a vast period, spanning the Norman invasion and the invention of the printing press, but these events did not upset the dynamics of English regularization. Therefore, it is possible to retrospectively trace the evolution of the irregular verbs, moving backwards in time from the observed Modern English distribution and up through Middle and Old English. Going still further back in time allows us to explore the effects of completely undoing the frequency-dependent selective process that the irregular verbs have undergone. Eventually, the shape of the curve changes from unimodal to a power law decline, with slope of nearly −0.75 (Fig. 3). This finding is consistent with the fact that random subsets of verbs (and of all types of words) exhibit such a zipfian distribution. The observed irregular verb distribution is the result of selective pressure on a random collection of ancestral verbs.

We can also make predictions about the future of the past tense. By the time one verb from the set 'begin, break, bring, buy, choose, draw, drink, drive, eat, fall' will regularize, five verbs from the set 'bid, dive, heave, shear, shed, slay, slit, sow, sting, stink' will be regularized. If the current trends continue, only 83 of the 177 verbs studied will be irregular in 2500.

What will be the next irregular verb to regularize? It is likely to be wed/wed/wed. The frequency of 'wed' is only 4.2 uses per million verbs, ranking at the very bottom of the modern irregular verbs. Indeed, it is already being replaced in many contexts by wed/wedded/wedded. Now is your last chance to be a 'newly wed'. The married couples of the future can only hope for 'wedded' bliss.

In previous millennia, many rules vied for control of English language conjugation, and fossils of those rules remain to this day. Yet, from this primordial soup of conjugations, the dental suffix '-ed' emerged triumphant. The competing rules are long dead, and unfamiliar even to well-educated native speakers. These rules disappeared because of the gradual erosion of their instances by a process that we call regularization. But regularity is not the default state of a language—a rule is the tombstone of a thousand exceptions.

METHODS SUMMARY

We searched 11 reference works on Old and Middle English, compiling a list of every irregular verb that we found. We determined whether each verb is still present in Modern English. For all Old English verbs whose descendants remained in the English language, we checked whether they were still irregular using a complete listing of the modern irregular verbs. If they had regularized, we determined when regularization had occurred on the basis of the last time period in which we found a positive annotation listing the verb as irregular. A list of sources used and the entire resulting annotation are provided in the Supplementary Information.

We determined usage frequencies for all the verbs using the CELEX database. We then binned the Old English irregular verbs using a standard logarithmic binning algorithm in Python. We used the resulting binning to determine regularization rates for verbs of differing frequencies. Regularization rates (Fig. 1b) for each bin were computed directly. The fits to exponential decay (Fig. 2) and to the solution of the irregular equation (Fig. 3 and Supplementary Information) were produced using the method of least squares. The Python source code for producing the figures and the table is available at http://www.languagedata.org.

Received 20 March; accepted 27 July 2007.

1. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, Cambridge, 1965).
2. Lightfoot, D. *The Development of Language: Acquisition, Change and Evolution* (Blackwell, Oxford, 1999).
3. Clark, R. & Roberts, I. A computational model of language learnability and language change. *Linguist. Inq.* 24, 299–345 (1993).
4. Abrams, D. & Strogatz, S. Modelling the dynamics of language death. *Nature* 424, 900 (2003).
5. Nowak, M. A., Komarova, N. L. & Niyogi, P. Computational and evolutionary aspects of language. *Nature* 417, 611–617 (2002).

©2007 Nature Publishing Group

6. Hooper, J. in *Current Progress in Historical Linguistics* (ed. Christie, W.) 95–105 (North-Holland, Amsterdam, 1976).
7. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298,** 1569–1579 (2002).
8. Chomsky, N. & Lasnik, H. in *Syntax: An International Handbook of Contemporary Research* (ed. Jacobs, J.) 506–569 (de Gruyte, Berlin, 1993).
9. Dougherty, R. C. *Natural Language Computing* (Lawrence Erlbaum, Hillsdale, 1994).
10. Stabler, E. P. & Keenan, E. L. Structural similarity within and among languages. *Theor. Comput. Sci.* **293,** 345–363 (2003).
11. Niyogi, P. *The Computational Nature of Language Learning and Evolution* (MIT Press, Cambridge, 2006).
12. Labov, W. Transmission and diffusion. *Language* **83,** 344–387 (2007).
13. Pinker, S. *Words and Rules: The Ingredients of Language* (Basic Books, New York, 1999).
14. Kroch, A. Reflexes of grammar in patterns of language change. *Lang. Var. Change* **1,** 199–244 (1989).
15. Kroch, A. in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory* (eds Beals, K. *et al.*) 180–201 (CLS, Chicago, 1994).
16. Pinker, S. The irregular verbs. *Landfall* 83–85 (Autumn issue, 2000).
17. Bybee, J. *Morphology: a Study of Relation Between Meaning and Form* (Benjamins, Amsterdam, 1985).
18. Greenberg, J. in *Current Trends in Linguistics III* (eds Sebeok, T. A. *et al.*) 61–112 (Mouton, The Hague, 1966).
19. Bybee, J. From usage to grammar: the mind's response to repetition. *Language* **82,** 711–733 (2006).
20. Corbett, G., Hippisley, A., Brown, D. & Marriott, P. in *Frequency and the Emergence of Linguistic Structure* (eds Bybee, J. & Hopper, P.) 201–226 (Benjamins, Amsterdam, 2001).
21. Hare, M. & Elman, J. Learning and morphological change. *Cognition* **56,** 61–98 (1995).
22. Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. German inflection: the exception that proves the rule. *Cognit. Psychol.* **29,** 189–256 (1995).
23. Van der Wouden, T. in *Papers from the 3rd International EURALEX Congress* (eds Magay, T. & Zigány, J.) 363–373 (Akadémiai Kiadó, Budapest, 1988).
24. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, 1949).
25. Miller, G. A. Some effects of intermittent silence. *Am. J. Psychol.* **70,** 311–314 (1957).
26. Yang, C. *Knowledge and Learning in Natural Language* (Oxford Univ. Press, New York, 2002).
27. Glushko, M. Towards the quantitative approach to studying evolution of English verb paradigm. *Proc. 19th Scand. Conf. Ling.* **31,** 30–45 (2003).
28. Kroch, A. & Taylor, A. *Penn-Helsinki Parsed Corpus of Middle English* [CD-ROM] 2nd edn (2000) 〈http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-2/〉.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the John Templeton Foundation and by a grant from the NSF-NIH joint programme in mathematical biology. The Program for Evolutionary Dynamics is sponsored by J. Epstein. E.L. was supported by the National Defense Science and Engineering Graduate Fellowship and the National Science Foundation Graduate Fellowship. We thank S. Pinker, J. Rau, D. Donoghue and A. Presser for discussions, and J. Saragosti for help with visualization.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.L. (erez@erez.com).

©2007 Nature Publishing Group

Sea-ice decline due to more than warming alone

SIR — The dramatic loss of sea-ice cover over the Arctic this summer was widely reported, for example in your News story 'Arctic melt opens Northwest passage' (*Nature* **449,** 267; 2007), and frequently attributed to global warming. Although the gradual decline in sea-ice extent during the past four decades is in line with that expected from global warming, it is very unlikely that the loss of sea-ice cover this year is explicable solely in terms of temperature change.

Changing wind patterns are an important influence on the distribution of sea ice. Throughout summer 2007, exceptional pressure and wind patterns persisted over the Arctic Ocean. The observed migration of ice cover, from the Siberian and Beaufort seas northwards and eastwards into the Arctic Basin, was in line with the expected response to the anomalous winds. These Arctic wind anomalies were part of a global-scale pattern of highly unusual circulation this summer, the causes of which are as yet unclear.

The growing La Niña in the East Pacific undoubtedly had a major influence globally, and there is some evidence from past events that La Niña predisposes the circulation towards the type of exceptional patterns seen this summer.

Julia Slingo, Rowan Sutton
National Centre for Atmospheric Science, Walker Institute for Climate System Research, University of Reading, Earley Gate, Reading RG6 6BB, UK

Ethics reviews can be centralized without delays

SIR — The problems inherent in continuing the current system of local ethics reviews for multicentre trials are appropriately highlighted in your News Feature 'Trial and error' (*Nature* **448,** 530–532; 2007). However, you state that the Central Institutional Review Board of the US National Cancer Institute (NCI) adds bureaucracy by its efforts to centralize review. We disagree.

A unique feature of our model is that both central and local boards can be involved; the extent of the local review remains the prerogative of the local board, and their task is facilitated by access to the central board's review. This model has had widespread acceptance: 66% of NCI-designated cancer centres have now joined. The institute set up a central board for adult and paediatric cancer treatment trials because of the redundancy that occurs when each site's local institutional review board performs a separate ethics review. For example, 50 to 200 sites typically participate in NCI-sponsored phase III trials.

We do not agree that this centralization delays the time it takes to activate a study. The institute's Cancer Therapy Evaluation Program reveals that, for more than two-thirds of the protocols reviewed by the central board, activation is not delayed, as numerous other activities continue in parallel, such as creation of case report forms, finalization of contracts with pharmaceutical partners and drug shipment plans.

We believe that centralized ethics review, with local context supplied by the local board, is the preferred way to conduct review of multisite protocols. The NCI is currently supporting a formal cost analysis by an independent third party. In view of the high local costs cited in your News story, such data should compel administrators of local boards to reconsider their position on centralized review.

Jeffrey Abrams, Jacquelyn Goldberg
Central Institutional Review Board Initiative, National Cancer Institute, 6130 Executive Boulevard, Bethesda, Maryland 20892, USA

Missing the mark on biomedical research

SIR — Brian Martinson in his Commentary 'Universities and the money fix' (*Nature* **449,** 141–142; 2007) suggests that short-sighted greed for federal grants motivates academic leadership to sacrifice the future interests of biomedical research. On the contrary, such grants do not cover the full costs of the research projects they support.

The annual shortfall of billions of dollars (see C. A. Goldman and T. Williams *Paying for University Research Facilities and Administration* Rand, Santa Monica; 2000) must be offset from the limited sources of discretionary funds available to universities: tuition fees, state funding, philanthropy, technology transfer or revenues derived from the clinical practice of the faculty.

Of these, philanthropy is serendipitous and generally attracts investment in new initiatives rather than in sustaining existing programmes. The pressures on tuition costs and clinical-practice margins have been widely publicized. Technology transfer only rarely brings significant, sustainable revenues.

If academic institutions merely sought to maximize short-term revenues, they would not make the substantial long-term commitments of capital to physical plant, state-of-the-art technologies and skilled personnel that they do.

We believe that the expansion of research capacity is being driven instead by the explosive pace of advances in the biomedical sciences and in our understanding of the processes of health and disease, as well as by mounting dependence on multidisciplinary research teams and increasingly sophisticated new technologies.

These investments are inherently risky in their anticipation of future returns from federal and other sponsored research. Accordingly, the decision-making behind them is deliberate and calculated: university governing boards, bound by their fiduciary obligations, are inherently conservative.

The distress caused by National Institutes of Health budgets that have steadily declined in purchasing power since 2003 is being widely and acutely felt. But increases in these budgets are unlikely; simply sustaining the enterprise to keep pace with inflation — which Martinson terms "prudent" — has been difficult.

The biomedical research enterprise does need to adapt to fiscal realities, but restructuring the historic 'business model' for academic research, including the composition of its workforce, will be very difficult. At minimum, such restructuring must be guided by sensitivity to the institutions' missions, recognition of the intense sociopolitical expectations of them and accurate perception of what is necessary for their academic and economic survival.

David Korn, Stephen J. Heinig
Association of American Medical Colleges, 2450 N Street, Washington DC 200037, USA

Words of wisdom worth reading aloud

SIR — I congratulate Erez Lieberman and colleagues on their entertaining, enthralling and, above all, well-written Letter 'Quantifying the evolutionary dynamics of language' (*Nature* **449,** 713–716; 2007). *Nature* articles outside my sphere of interest are of interest to me as far as the abstract; after that I am rapidly lost in a forest of jargon, technicalities and poorly articulated background information.

I had little prior interest in linguistic evolution, yet Lieberman and colleagues captured my attention to the end of their Letter. It was so well written that I was compelled to read out the cleverest snippets to my office-mates.

This is a scientific paper that has truly exemplified the aim of *Nature* to provide a forum for work with interdisciplinary appeal. I congratulate and thank both the authors and the selecting editors.

Vanessa S. Solomon
Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria, Australia

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 300 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

©2007 Nature Publishing Group

THIS YEAR IN NATURE

***Nature*'s manuscript editors made a selection of 'favourites' from the papers we published in 2007.**

STEM CELLS

Now in primates

Producing primate embryonic stem cells by somatic cell nuclear transfer, J A Byrne *et al. Nature* **450**, 497–502 (22 November).

This important Article brought somatic cell nuclear transfer to primates for the first time. A group led by Shoukhrat Mitalipov generated two embryonic stem cell lines from 314 oocytes taken from 14 rhesus monkeys. The hope is that this approach could work in humans for generating patient-derived embryonic stem cells. [doi: 10.1038/nature06357]

GENOMICS

Disease risks

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, The Wellcome Trust Case Control Consortium *Nature* **447**, 661–678 (7 June).

An extensive genome-wide association study in the British population for bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, and type I and type II diabetes. The analysis confirms previously identified loci and provides strong evidence for many novel disease susceptibility loci. [doi: 10.1038/nature05911]

QUANTUM PHYSICS

Travelling light

Coherent control of optical information with matter wave dynamics, N S Ginsberg, SR Garner & L V Hau *Nature* **445**, 623–626 (8 February).

This development in quantum control may find application in quantum information processing. The cover illustration represents a unique experiment: a light pulse stopped and extinguished in one 'box' is revived from a completely different box in a separate location and sent back on its way. The 'boxes' in the actual experiment were Bose–Einstein condensates, 160 μm apart. Information was transferred by converting the optical pulse into a travelling matter wave more amenable to manipulation than light. [doi: 10.1038/nature05493]

EXTRASOLAR PLANETS

Caught in transit

A spectrum of an extrasolar planet, L J Richardson, D Deming, K Horning, S Seager & J Harrington *Nature* **445**, 892–895 (22 February).

Fourteen of the 200 plus known extrasolar planets exhibit transits in front of their parent stars as seen from Earth. In theory, subtracting the spectrum seen with the planet 'behind' its star from that seen when it is 'in front', should leave the actual spectrum of an extrasolar planet. This year practice caught up with theory, and an infrared spectrum was obtained for the transiting 'hot Jupiter' HD 209458b. [doi: 10.1038/nature05636]

LINGUISTICS

Lost words

Quantifying the evolutionary dynamics of language, E Lieberman, J-B Michel, J Jackson, T Tang & M A Nowak *Nature* **449**, 713–716 (11 October).

An unusual paper for *Nature* perhaps. A calculation of the rate at which a language grows more regular, based on 1,200 years of English usage. The trend follows a simple rule: a verb's half-life scales as the square root of its frequency. Irregular verbs that are 100 times as rare regularize 10 times faster. Exceptional forms are gradually lost. Next to go, and next to tumble in the cover 'hour-glass', is the word 'wed'. [doi: 10.1038/nature06137]

MATERIALS

Good on the flat

The structure of suspended graphene sheets, J C Meyer *et al. Nature* **446**, 60–63 (1 March 2007).

Graphene — a one-atom-thick layered form of graphite — is a hot topic in materials science and also in condensed matter physics, where it is a popular model system for investigation. This experiment suspended individual graphene sheets over a microscale scaffold so that transmission electron microscopy and diffraction could be used in structure determination. The 'wavy' structure that emerged is a step towards an answer to the question of why such a 'two-dimensional' structure can exist at all. [doi: 10.1038/nature05545]

CANCER GENOMICS

A feat of sequencing

Patterns of somatic mutation in human cancer genomes, C Greenman *et al. Nature* **446**, 153–158 (8 March 2007)

A series of outstanding papers on large-scale cancer genomics have appeared this year. In this example a sample of 518 kinases associated with more than 200 different cancers were chosen for a major sequencing effort. More than 1,000 previously unknown mutations linked to tumour formation — some as 'passengers' that don't contribute to cancer formation, but over 100 of them as 'driver' mutations that do contribute to disease development. [doi: 10.1038/nature05610]

MECHANOCHEMISTRY

At a push

Biasing reaction pathways with mechanical force, C R Hickenboth *et al. Nature* **446**, 423–427 (22 March).

Mechanical force joins heat, light, pressure and electrical potential as a means of kick-starting a chemical reaction by pushing reactants over an energy barrier. In specially designed polymers exposed to ultrasound, rearrangement reactions are accelerated and reaction pathways biased to yield products not obtainable from heat- or light-induced reactions. This work offers new ways of controlling chemical reactions and may lead to mechanically adaptable polymers that can signal impending damage, undergo structure modification to slow the rate of damage, or even self-repair. [doi: 10.1038/nature05681]

NEUROSCIENCE

Leading light

Multimodal fast optical interrogation of neural circuitry, F Zhang *et al. Nature* **446**, 633–639 (5 April).

A powerful new technique for controlling neural circuits that is likely to have an impact in all fields of neurobiology. It uses a light-activated chloride pump derived from *Archea* to label neurons, which can then be inhibited by exposure to light. This complements an existing tool

©2007 Nature Publishing Group

for activating neurons via a photoactivatable algal channel. As the channels are sensitive to different wavelengths of light, fast bidirectional control over neural activity can be achieved in a single neural circuit. [doi: 10.1038/nature05744]

ATTOPHYSICS

Solid progress

Attosecond spectroscopy in condensed matter, A L Cavalieri *et al. Nature* **449**, 1029–1032 (25 October).

Electron dynamics on the attosecond timescale (an attosecond is a billionth of a billionth of a second) used to be directly measured only in atomic gases. This paper reported the first attosecond spectroscopic measurements in a solid. The ability to time electrons moving in solids over merely a few interatomic distances makes it possible to probe the solid-state electronic processes occurring at the theoretical ultimate speed limit for electronics. [doi: 10.1038/nature06229]

APPLIED PHYSICS

Weighing in

Weighing of biomolecules, single cells and single nanoparticles in fluid, T P Burg *et al. Nature* **446**, 1066–1069 (26 April).

Tiny particles — molecules included — can be weighed with remarkably high resolution using nanoscale mechanical resonators. But not in the presence of fluids, which dampen the vibrations that make the system work. This rules out practical applications such as medical diagnostics or environmental monitoring. But here is an ingenious way around the problem: 'hide' the fluid inside the resonator. A vacuum-packaged resonator holds the solution with particles of interest in microfluidic channels, and weighs single nanoparticles and bacteria at subfemtogram resolution. [doi: 10.1038/nature05741]

MOLECULAR BIOLOGY

Polymerase structures

Structural basis for transcription elongation by bacterial RNA polymerase, D G Vassylyev *et al. Nature* **448**, 157–162 ; Structural basis for substrate loading in bacterial RNA polymerase, D G Vassylyev *et al. Nature* **448**, 163–168 (12 July).

Two papers in the classic mould of structural biology. The first establishes the crystal structure of bacterial RNA polymerase bound to the DNA template and RNA product, revealing a detailed view of the transcription elongation complex. And the second determines the structures of bacterial RNA polymerase elongation complexes bound to NTP substrate analogues with an antibiotic, revealing the mechanism of substrate loading and antibiotic inhibition. [doi: 10.1038/nature05932; doi: 10.1038/nature05931]

MOLECULAR BIOLOGY

A no-nonsense drug

PTC124 targets genetic disorders caused by nonsense mutations, E M Welch *et al. Nature* **447**, 87–91 (3 May)

Many inherited diseases result from premature termination during translation of an mRNA into protein. Muscular dystrophy is one such disease. This work shows that a small molecule, PTC124, enables the translation machinery to bypass sites that cause premature termination, but still terminate normally at the end of the mRNA. PTC124 can restore normal translation of the gene that is mutated in muscular dystrophy, and it also restores muscle function in a mouse model of the disease. This drug offers hope that a wide variety of diseases with similar translation defects might be amenable to treatment that will restore protein function. [doi: 10.1038/nature05756]

CELL BIOLOGY

Architectural prize

Determining the architectures of macromolecular assemblies, F Alber *et al. Nature* **450**, 683–694; The molecular architecture of the nuclear pore complex, F Alber *et al. Nature* **450**, 695–701 (29 November).

The proteomics based technology described in the first of these two papers will allow cell biologists to look at the detailed structure of all manner of macrocellular machines. The second paper shows the power of the technique by tackling the architecture of nuclear pore complexes, the macromolecular assemblies that selectively transport cargo across the nuclear envelope. [doi: 10.1038/nature06404; doi: 10.1038/nature06405]

THE SOLAR SYSTEM

A touch of Venus

Venus Express package, *Nature* **450**, 629–662 (29 November)

The 29 November issue included eight research papers presenting results from ESA's Venus Express mission — which has been in orbit since April 2006. Subjects covered included the atmosphere, polar features, interactions with the solar wind and the controversial matter of venusian lightning. [doi: 10.1038/nature06432]

STRUCTURAL BIOLOGY

Structural work

Crystal structure of the sodium–potassium pump, J P Morth *et al. Nature* **450**, 1043–1049 (13 December).

One of a package of three papers on the structure of P-type ATPases, this Article reports the long-awaited crystal structure of the Na^+,K^+-pump at a resolution of 3.5 Å. P-type ATPases are cation pumps of fundamental importance for all eukaryotes and many prokaryotes. [doi: 10.1038/nature06419]

BIOFUELS

Steady on the alcohol

Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates, Y Román-Leshkov, C J Barrett, Z Y Liu & J A Dumesic *Nature* **447**, 982–985 (21 June).

Ethanol has its limitions as a biofuel: it is highly volatile, absorbs water and has a low energy density. A potentially better liquid biofuel on those three counts is 2,5-dimethylfuran (DMF). This two-step catalytic process, still in the development stage, can produce DMF from fructose, which can be made either directly from biomass or from glucose. [doi: 10.1038/nature05923]

CLIMATE

The human factor

Detection of human influence on twentieth-century precipitation trends, X Zhang *et al. Nature* **448**, 461–465 (26 July)

Climate models suggested that human activity has caused changes in precipitation on a global scale, but no evidence had been found to support the prediction. This paper produced that evidence. A comparison of observed changes in precipitation over land during the twentieth century with climate simulations points to a detectable influence on the latitudinal patterns of precipitation. Anthropogenic factors contributed to moistening in Northern Hemisphere mid-latitudes, but elsewhere, for instance in the Northern Hemisphere tropics, the effect was drying. [doi: 10.1038/nature06025]

©2007 Nature Publishing Group