Human language reveals a universal positivity bias

Human language reveals a universal positivity bias

Peter Sheridan Dodds[a,b,1], Eric M. Clark[a,b], Suma Desu[c], Morgan R. Frank[c], Andrew J. Reagan[a,b], Jake Ryland Williams[a,b], Lewis Mitchell[d], Kameron Decker Harris[e], Isabel M. Kloumann[f], James P. Bagrow[a,b], Karine Megerdoomian[g], Matthew T. McMahon[g], Brian F. Tivnan[b,g,1], and Christopher M. Danforth[a,b,1]

[a]Computational Story Lab, Vermont Advanced Computing Core, and the Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05401; [b]Vermont Complex Systems Center, University of Vermont, Burlington, VT 05401; [c]Center for Computational Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [d]School of Mathematical Sciences, The University of Adelaide, SA 5005, Australia; [e]Applied Mathematics, University of Washington, Seattle, WA 98195; [f]Center for Applied Mathematics, Cornell University, Ithaca, NY 14853; and [g]The MITRE Corporation, McLean, VA 22102

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved January 9, 2015 (received for review June 23, 2014)

Using human evaluation of 100,000 words spread across 24 corpora in 10 languages diverse in origin and culture, we present evidence of a deep imprint of human sociality in language, observing that (*i*) the words of natural human language possess a universal positivity bias, (*ii*) the estimated emotional content of words is consistent between languages under translation, and (*iii*) this positivity bias is strongly independent of frequency of word use. Alongside these general regularities, we describe interlanguage variations in the emotional spectrum of languages that allow us to rank corpora. We also show how our word evaluations can be used to construct physical-like instruments for both real-time and offline measurement of the emotional content of large-scale texts.

language | social psychology | happiness | positivity

Human language, our great social technology, reflects that which it describes through the stories it allows to be told and us, the tellers of those stories. Although language's shaping effect on thinking has long been controversial (1–3), we know that a rich array of metaphor encodes our conceptualizations (4), word choice reflects our internal motives and immediate social roles (5–7), and the way a language represents the present and future may condition economic choices (8).

In 1969, Boucher and Osgood (9) framed the Pollyanna hypothesis: a hypothetical, universal positivity bias in human communication. From a selection of small-scale, cross-cultural studies, they marshaled evidence that positive words are likely more prevalent, more meaningful, more diversely used, and more readily learned. However, in being far from an exhaustive, data-driven analysis of language, which is the approach we take here, their findings could only be regarded as suggestive. Indeed, studies of the positivity of isolated words and word stems have produced conflicting results, some pointing toward a positivity bias (10) and others toward the opposite (11, 12), although attempts to adjust for use frequency tend to recover a positivity signal (13).

Materials and Methods

To explore the positivity of human language deeply, we constructed 24 corpora spread across 10 languages. Our global coverage of linguistically and culturally diverse languages includes English, Spanish, French, German, Brazilian Portuguese, Korean, Chinese (Simplified), Russian, Indonesian, and Arabic. The sources of our corpora are similarly broad, spanning books (14), news outlets, social media, the web (15), television and movie subtitles, and music lyrics (16). Our work here greatly expands upon our earlier study of English alone, where we found strong evidence for a use-invariant positivity bias (17). In *SI Appendix*, we provide full details of our corpora (*SI Appendix*, Table S1), survey, and participants (*SI Appendix*, Table S2).

We address the social nature of language in two important ways: (*i*) We focus on the words people most commonly use, and (*ii*) we measure how those same words are received by individuals. We take word use frequency as the primary organizing measure of a word's importance. Such a data-driven approach is crucial for both understanding the structure of language and creating linguistic instruments for principled measurements (18, 19). By contrast, earlier studies focusing on meaning and emotion have used "expert" generated word lists, and these word lists fail statistically to match frequency distributions of natural language (10–12, 20), confounding attempts to make claims about language in general. For each of our corpora, we selected between 5,000 and 10,000 of the most frequently used words, choosing the exact numbers so that we obtained ∼10,000 words for each language.

Of our 24 corpora, we received 17 already parsed into words by the source: the Google Books Project (six corpora), the Google Web Crawl (eight corpora), and movie and television subtitles (three corpora). For the other seven corpora (five Twitter corpora, the *New York Times*, and music lyrics), we extracted words by standard white space separation. Twitter was easily the most variable and complex of our text sources, and required additional treatment. In parsing Twitter, we required strings to contain at least one Unicode character and no invisible control characters, and we excluded strings representing web links, bearing a leading @, ampersand (&), or other punctuation (e.g., Twitter IDs) but kept hashtags. Finally, for all corpora, we converted words to lowercase. We observed that common English words appeared in the Twitter corpora of other languages, and we have chosen simply to acknowledge this reality of language and allow these commonly used borrowed words to be evaluated.

Although there are many complications with inflections and variable orthography, we have found merit for our broad analysis in not collapsing related words. For example, we have observed that allowing different conjugations of verbs to stand in our corpora is valuable because human evaluations of such have proved to be distinguishable [e.g., present vs. past tense (18)]. As should be expected, a more nuanced treatment going beyond the present paper's bounds by involving stemming and word type, for example, may lead to minor corrections (21), although our central observations will remain robust and will in no way change the behavior of the instruments we generate.

There is no single, principled way to merge corpora to create an ordered list of words for a given language. For example, it is impossible to weight the most commonly used words in the *New York Times* against the most commonly used words in Twitter. Nevertheless, we are obliged to choose some method for doing so to facilitate comparisons across languages and for the purposes of building adaptable linguistic instruments. For each language where we had more than one corpus, we created a single quasi-ranked word list by finding the smallest integer r such that the union of all words with a rank $\leq r$ in at least one corpus formed a set of at least 10,000 words.

Significance

The most commonly used words of 24 corpora across 10 diverse human languages exhibit a clear positive bias, a big data confirmation of the Pollyanna hypothesis. The study's findings are based on 5 million individual human scores and pave the way for the development of powerful language-based tools for measuring emotion.

Author contributions: P.S.D., B.F.T., and C.M.D. designed research; P.S.D., E.M.C., S.D., M.R.F., A.J.R., J.R.W., L.M., K.D.H., I.M.K., J.P.B., K.M., M.T.M., and C.M.D. performed research; P.S.D., E.M.C., A.J.R., K.M., and M.T.M. contributed new reagents/analytic tools; P.S.D., E.M.C., S.D., M.R.F., A.J.R., J.R.W., J.P.B., and C.M.D. analyzed data; and P.S.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Data are available in Dataset S1 and at www.uvm.edu/storylab/share/papers/dodds2014a/index.html.

[1]To whom correspondence may be addressed. Email: peter.dodds@uvm.edu, btivnan@mitre.org, or chris.danforth@uvm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1411678112/-/DCSupplemental.

We then paid native speakers to rate how they felt in response to individual words on a nine-point scale, with 1 corresponding to most negative or saddest, 5 to neutral, and 9 to most positive or happiest (10, 18) (*SI Appendix*). This happy–sad semantic differential (20) functions as a coupling of two standard five-point Likert scales. Participants were restricted to certain regions or countries (e.g., Portuguese was rated by residents of Brazil). Overall, we collected 50 ratings per word for a total of around 5 million individual human assessments. We provide all datasets as part of *SI Appendix*.

Results and Discussion

In Fig. 1, we show distributions of the average happiness scores for all 24 corpora, leading to our most general observation of a clear positivity bias in natural language. We indicate the above-neutral part of each distribution with yellow and the below-neutral part with blue, and we order the distributions moving upward by increasing median (vertical red line). For all corpora, the median clearly exceeds the neutral score of 5. The background gray lines connect deciles for each distribution. In *SI Appendix*, Fig. S1, we provide the same distributions ordered instead by increasing variance.

As is evident from the ordering in Fig. 1 and *SI Appendix*, Fig. S1, although a positivity bias is the universal rule, there are minor differences between the happiness distributions of languages. For example, Latin American-evaluated corpora (Mexican Spanish and Brazilian Portuguese) exhibit relatively high medians and, to a lesser degree, higher variances. For other languages, we see that those languages with multiple corpora have more variable medians, and specific corpora are not ordered by median in the same way across languages (e.g., Google Books has a lower median than Twitter for Russian, but the reverse is true for German and English). In terms of emotional variance, all four English corpora are among the highest, whereas Chinese and Russian Google Books seem especially constrained.

We now examine how individual words themselves vary in their average happiness score between languages. Owing to the scale of our corpora, we were compelled to use an online service, choosing Google Translate. For each of the 45 language pairs, we translated isolated words from one language to the other and then back. We then found all word pairs that (*i*) were translationally stable, meaning the forward and back translation returns the original word, and (*ii*) appeared in our corpora for each language.

We provide the resulting comparison between languages at the level of individual words in Fig. 2. We use the mean of each language's word happiness distribution derived from its merged corpora to generate a rough overall ordering, acknowledging that frequency of use is no longer meaningful, and moreover is not relevant, because we are now investigating the properties of individual words. Each cell shows a heat map comparison with word density increasing as shading moves from gray to white. The background colors reflect the ordering of each pair of languages: yellow if the row language had a higher average happiness than the column language and blue for the reverse. Also, in each cell, we display the number of translation-stable words between language pairs, N, along with the difference in average word happiness, Δ, where each word is equally weighted.

A linear relationship is clear for each language–language comparison, and is supported by Pearson's correlation coefficient r being in the range 0.73–0.89 ($P < 10^{-118}$ across all pairs; Fig. 2 and *SI Appendix*, Tables S3–S5). Overall, this strong agreement between languages suggests that approximate estimates of word happiness for unscored languages could be generated with no expense from our existing dataset. Some words will, of course, translate unsatisfactorily, with the dominant meaning changing between languages. For example "lying" in English, most readily interpreted as speaking falsehoods by our participants, translates to "acostado" in Spanish, meaning recumbent. Nevertheless, happiness scores obtained by translation will be serviceable for purposes where the effects of many different words are incorporated.

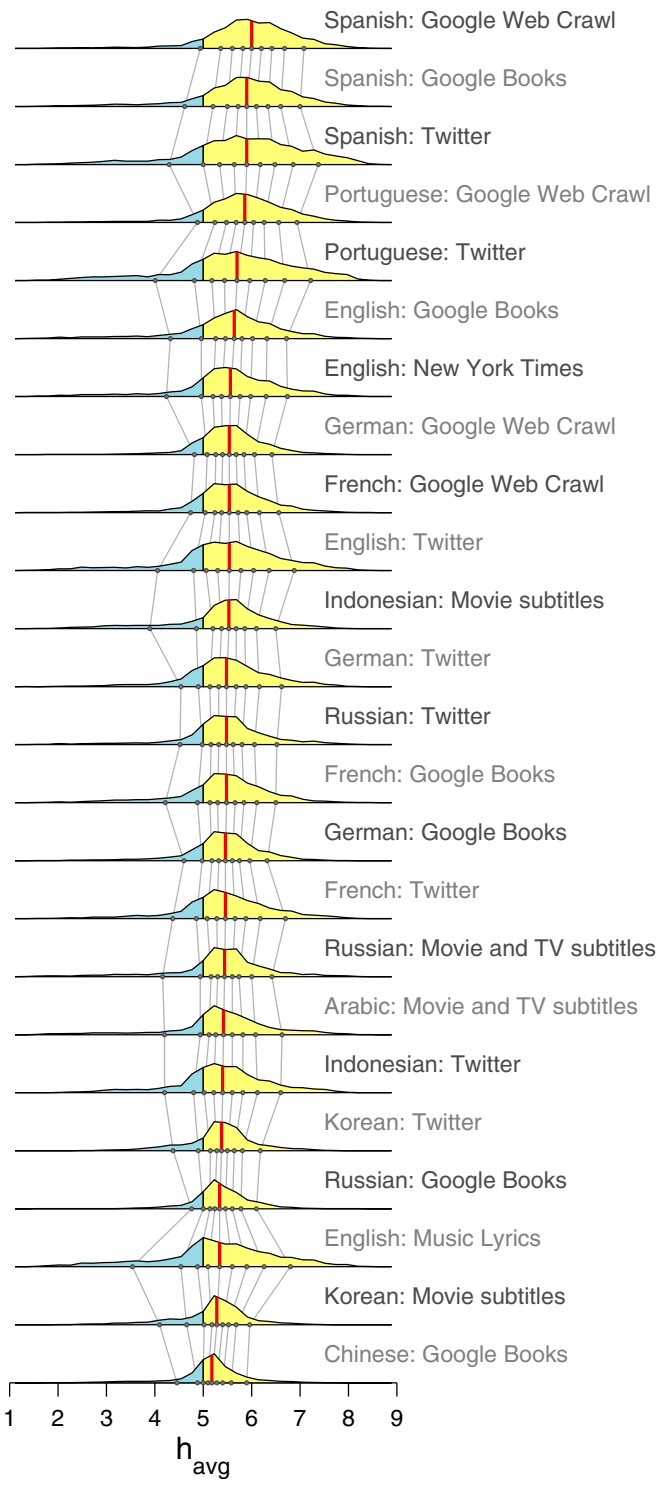

Fig. 1. Distributions of perceived average word happiness h_{avg} for 24 corpora in 10 languages. The histograms represent the 5,000 most commonly used words in each corpora (details are provided in *SI Appendix*), and native speakers scored words on a 1–9 double-Likert scale, with 1 being extremely negative, 5 being neutral, and 9 being extremely positive. Yellow indicates positivity ($h_{avg} > 5$), and blue indicates negativity ($h_{avg} < 5$). Distributions are ordered by increasing median (red vertical line). The background gray lines connect deciles of adjacent distributions. The same distributions arranged according to increasing variance are shown in *SI Appendix*, Fig. S1.

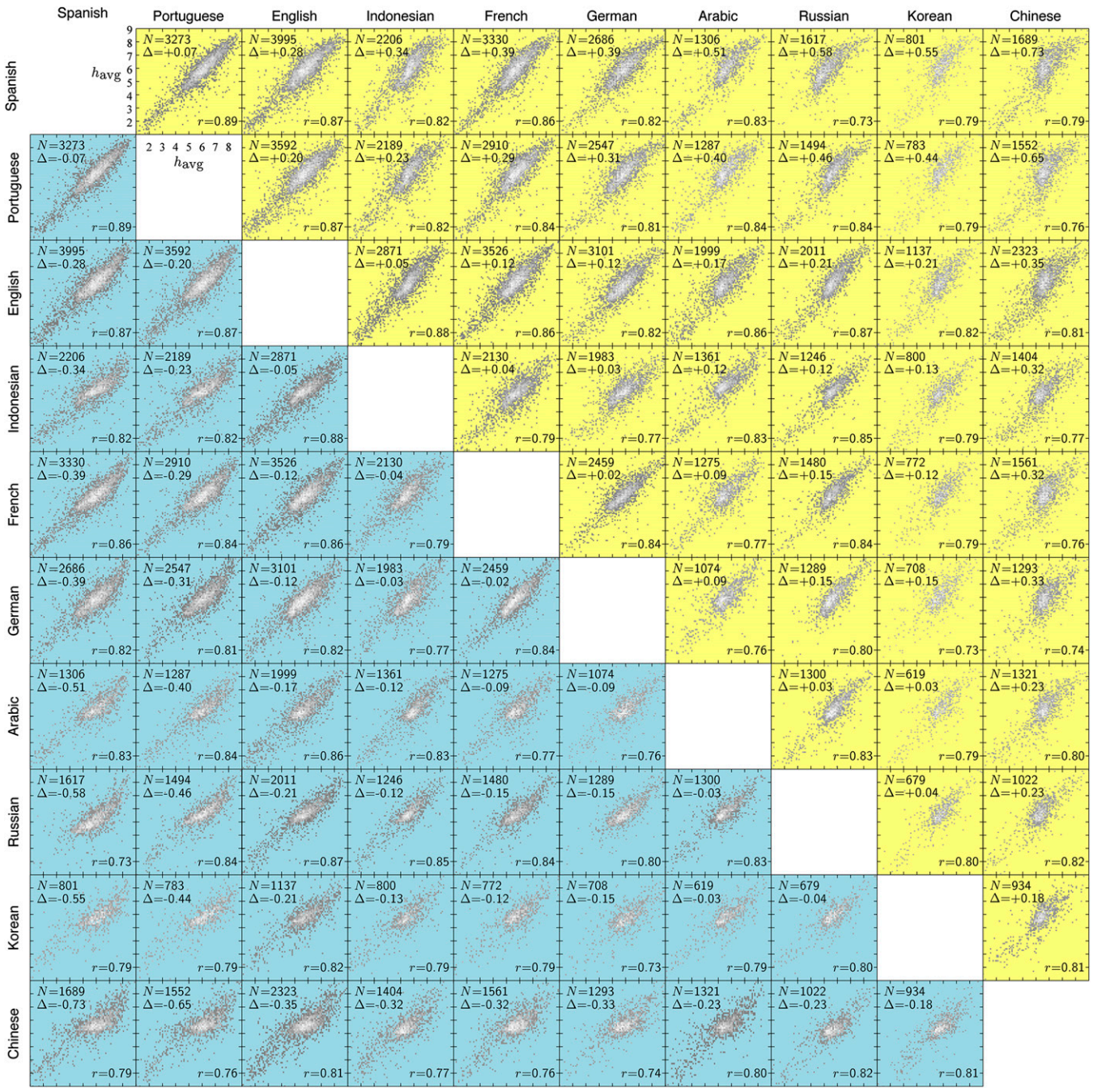

Fig. 2. Scatter plots of average happiness for words measured in different languages. We order languages from relatively most positive (Spanish) to relatively least positive (Chinese); a yellow background indicates the row language is more positive than the column language, and a blue background indicates the converse. The overall plot matrix is symmetrical about the leading diagonal, with the redundancy allowing for easier comparison between languages. In each scatter plot, the key gives the number of translation-stable words for each language pair, *N*; the average difference in translation-stable word happiness between the row language and the column language, Δ; and the Pearson correlation coefficient for the regression, *r*. All *P* values are less than 10^{-118} for the Pearson correlation coefficient and less than 10^{-82} for the Spearman correlation coefficient.

Stepping back from examining interlanguage robustness, we return to a more detailed exploration of the rich structure of each corpus's happiness distribution. In Fig. 3, we show how average word happiness h_{avg} is largely independent of word use frequency for four example corpora. We first plot use frequency rank *r* of the 5,000 most frequently used words as a function of their average happiness score, h_{avg} (background dots), along with some example evenly spaced words. We note that words at the extremes of the happiness scale are ones evaluators agreed upon strongly, whereas words near neutral range from being clearly neutral [e.g., h_{avg} ("the") = 4.98] to contentious with high SD (17). We then compute deciles for contiguous sets of 500 words, sliding this window through rank *r*. These deciles form the vertical strands. We overlay randomly chosen, equally spaced example words to give a sense of each corpus's emotional texture.

We chose the four example corpora shown in Fig. 3 to be disparate in nature, covering diverse languages (French, Egyptian Arabic, Brazilian Portuguese, and Chinese), regions of the

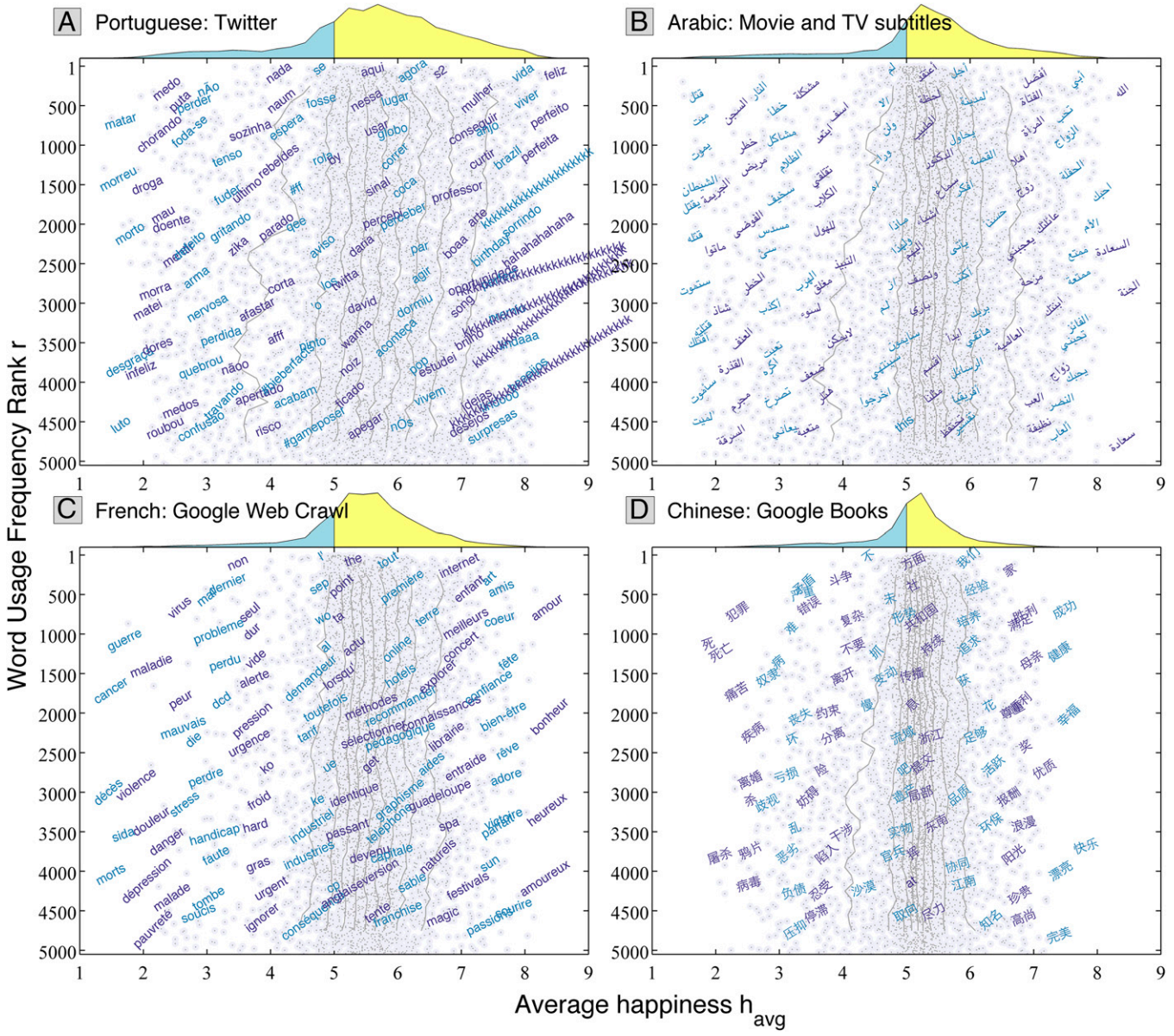

Fig. 3. Examples of how word happiness varies little with use frequency. (*A–D*) Above each plot is a histogram of average happiness h_{avg} for the 5,000 most frequently used words in the given corpus, matching Fig. 1. Each point locates a word by its usage rank r and average happiness h_{avg}, and we show some regularly spaced example words. The descending gray curves of these jellyfish plots indicate deciles for windows of 500 words of contiguous use rank, showing that the overall histogram's form is roughly maintained at all scales. The "kkkkkk…" words represent laughter in Brazilian Portuguese, in the manner of "hahaha…," and an English translation is provided in *SI Appendix*, Fig. S2.

world (Europe, the Middle East, South America, and Asia), and texts [Twitter, movies and television, the web (15), and books (14)]. The remaining 20 corpora yield similar plots (*SI Appendix*), and all corpora also exhibit an approximate self-similarity in SD for word happiness.

Across all corpora, we observe visually that the deciles tend to stay fixed or move slightly toward the negative, with some expected fragility at the 10% and 90% levels (due to the distributions' tails), indicating that the overall happiness distribution of each corpus approximately holds independent of word use. In Fig. 3, for example, we see that both the Brazilian Portuguese and French examples show a small shift to the negative for increasingly rare words, whereas there is no visually clear trend for the Arabic and Chinese cases. Fitting $h_{avg} = \alpha r + \beta$ typically returns α on the order of -1×10^{-5} suggesting h_{avg} decreases 0.1 per 10,000 words. For SDs of happiness scores, we find a similarly weak drift toward higher values for increasingly rare words (correlations and linear fits for h_{avg} and h_{std} as a function of word rank r for all corpora are provided in *SI Appendix*, Tables S6 and S7). We thus find that, to first order, not just the positivity bias but the happiness distribution itself applies for common words and rare words alike, revealing an unexpected addition to the many well-known scalings found in natural language, as famously exemplified by Zipf's law (22).

In constructing language-based instruments for measuring expressed happiness, such as our hedonometer (18), this frequency independence allows for a way to "increase the gain" in a fashion resembling standard physical instruments. Moreover, we have earlier demonstrated the robustness of our hedonometer for the

English language, showing, for example, that measurements derived from Twitter correlate strongly with Gallup well-being polls and related indices at the state and city level for the United States (19).

Here, we provide an illustrative use of our hedonometer in the realm of literature, inspired by Vonnegut's shapes of stories (23, 24). In Fig. 4, we show "happiness time series" for three famous works of literature, evaluated in their original languages of English, Russian, and French, respectively: Melville's *Moby Dick* (www.gutenberg.org), Dostoyevsky's *Crime and Punishment* (25), and Dumas' *The Count of Monte Cristo* (www.gutenberg.org). We slide a 10,000-word window through each work, computing the average happiness using a "lens" for the hedonometer in the following manner. We capitalize on our instrument's tunability to obtain a strong signal by excluding all words for which $3 < h_{\text{avg}} < 7$ (i.e., we keep words residing in the tails of each distribution) (18). Denoting a given lens by its corresponding set of allowed words L, we estimate the happiness score of any text T as $h_{\text{avg}}(T) = \sum_{w \in L} f_w h_{\text{avg}}(w) / \sum_{w \in L} f_w$, where f_w is the frequency of word w in T (16).

The three resulting happiness time series provide interesting detailed views of each work's narrative trajectory revealing

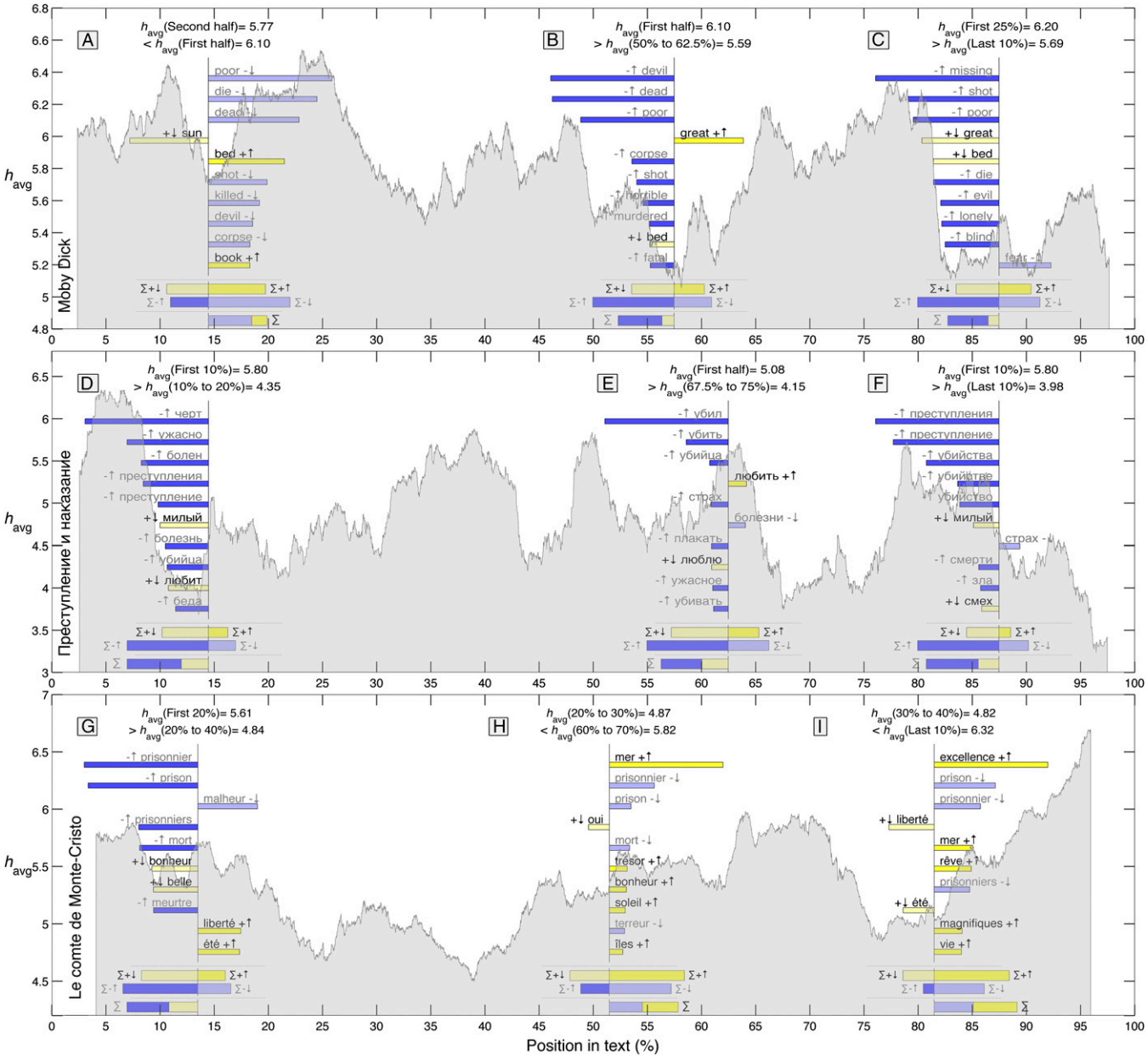

Fig. 4. Emotional time series for three famous 19th century works of literature: Melville's *Moby Dick* (*Top*), Dostoyevsky's *Crime and Punishment* (*Middle*), and Dumas' *The Count of Monte Cristo* (*Bottom*). Each point represents the language-specific happiness score for a window of 10,000 words (converted to lowercase), with the window translated throughout the work. The overlaid word shifts (A–I) show example comparisons between different sections of each work. Word shifts indicate which words contribute the most toward and against the change in average happiness between two texts (*SI Appendix*, pp. S9–S10). Although a robust instrument in general, we acknowledge the hedonometer's potential failure for individual words due to both language evolution and words possessing more than one meaning. For *Moby Dick*, we excluded "cried" and "cry" (to speak loudly rather than weep) and "Coffin" (surname, still common in Nantucket). Such alterations, which can be done on a case-by-case basis, do not noticeably change the overall happiness curves, although leaving the word shifts more informative. We provide online, interactive visualizations of the emotional trajectories of over 10,000 books at hedonometer.org/books.html.

numerous peaks and troughs throughout, at times clearly dropping below neutral. Both *Moby Dick* and *Crime and Punishment* end on low notes, whereas *The Count of Monte Cristo* culminates with a rise in positivity, accurately reflecting the finishing arcs of all three. The "word shifts" overlaying the time series compare two distinct regions of each work, showing how changes in word abundances lead to overall shifts in average happiness. Such word shifts are essential tests of any sentiment measurement, and are made possible by the linear form of our instrument (16, 18) [a full explanation is provided in *SI Appendix*, pp. S9–S10]. As one example, the third word shift for *Moby Dick* shows why the average happiness of the last 10% of the book is well below the average happiness of the first 25%. The major contribution is an increase in relatively negative words, including "missing," "shot," "poor," "die," and "evil."

By adjusting the lens, many other related time series can be formed, such as those produced by focusing on only positive or negative words. Emotional variance as a function of text position can also be readily extracted. At hedonometer.org/books.html, we provide online, interactive emotional trajectories for over 10,000 works of literature where different lenses and regions of comparisons may be easily explored. Beyond this example tool we have created here for the digital humanities and our hedonometer for measuring population well-being, the datasets we have generated for the present study may be useful in creating a great variety of language-based instruments for assessing emotional expression.

Overall, our major scientific finding is that when experienced in isolation and weighted properly according to use, words, which are the atoms of human language, present an emotional spectrum with a universal, self-similar positive bias. We emphasize that this apparent linguistic encoding of our social nature is a system-level property, and in no way asserts all natural texts will skew positive (as exemplified by certain passages of the three works in Fig. 4) or diminishes the salience of negative states (26). Going forward, our word happiness assessments should be periodically repeated and carried out for new languages, tested on different demographics, and expanded to phrases both for the improvement of hedonometric instruments and to chart the dynamics of our collective social self.

ACKNOWLEDGMENTS. We thank M. Shields, K. Comer, N. Berry, I. Ramiscal, C. Burke, P. Carrigan, M. Koehler, and Z. Henscheid, in part, for their roles in developing hedonometer.org. We also thank F. Henegan, A. Powers, and N. Atkins for conversations. P.S.D. was supported by National Science Foundation CAREER Award 0846668.

1. Whorf BL (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed Carroll JB (MIT Press, Cambridge, MA).
2. Chomsky N (1957) *Syntactic Structures* (Mouton, The Hague).
3. Steven Pinker (1994) *The Language Instinct: How the Mind Creates Language* (William Morrow and Company, New York).
4. Lakoff G, Johnson M (1980) *Metaphors We Live By* (Univ of Chicago Press, Chicago).
5. Campbell RS, Pennebaker JW (2003) The secret life of pronouns: Flexibility in writing style and physical health. *Psychol Sci* 14(1):60–65.
6. Newman MEJ (2003) The structure and function of complex networks. *SIAM Review* 45(2):167–256.
7. Pennebaker JW (2011) *The Secret Life of Pronouns: What Our Words Say About Us* (Bloomsbury Press, New York).
8. Chen MK (2013) The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *Am Econ Rev* 103(2):690–731.
9. Boucher J, Osgood CE (1969) The Pollyanna hypothesis. *J Verbal Learning Verbal Behav* 8(1):1–8.
10. Bradley MM, Lang PJ (1999) *Affective Norms for English Words (Anew): Stimuli, Instruction Manual and Affective Ratings. Technical Report C-1* (University of Florida, Gainesville, FL).
11. Stone PJ, Dunphy DC, Smith MS, Ogilvie DM (1966) *The General Inquirer: A Computer Approach to Content Analysis* (MIT Press, Cambridge, MA).
12. Pennebaker JW, Booth RJ, Francis ME (2007) *Linguistic Inquiry and Word Count: Liwc 2007*. Available at bit.ly/S1Dk2L. Accessed May 15, 2014.
13. Jurafsky D, Chahuneau V, Routledge BR, Smith NA (2014) Narrative framing of consumer sentiment in online restaurant reviews. *First Monday*, 10.5210/fm.v19i4.4944.
14. Michel J-B, et al. (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331:176–182.
15. Brants T, Franz A (2006) Google Web 1T 5-gram (Linguistic Data Consortium, Philadelphia). Version 1.
16. Dodds PS, Danforth CM (2009) Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *J Happiness Stud* 11(4):441–456.
17. Kloumann IM, Danforth CM, Harris KD, Bliss CA, Dodds PS (2012) Positivity of the English language. *PLoS ONE* 7(1):e29484.
18. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE* 6(12):e26752.
19. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM (2013) The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE* 8(5):e64417.
20. Osgood C, Suci G, Tannenbaum P (1957) *The Measurement of Meaning* (Univ of Illinois, Urbana, IL).
21. Warriner AB, Kuperman V (2014) Affective biases in English are bi-dimensional. *Cogn Emotion*, 10.1080/02699931.2014.968098.
22. Zipf GK (1949) *Human Behaviour and the Principle of Least-Effort* (Addison–Wesley, Cambridge, MA).
23. Vonnegut K, Jr (2005) *A Man Without a Country* (Seven Stories Press, New York).
24. Vonnegut K (2010) Kurt Vonnegut on the shapes of stories. Available at www.youtube.com/watch?v=oP3c1h8v2ZQ. Accessed May 15, 2014.
25. Dostoyevsky F (1866) Crime and Punishment. Original Russian text. Available at ilibrary.ru/text/69/p.1/index.html. Accessed December 15, 2013.
26. Forgas JP (2013) Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. *Curr Dir Psychol Sci* 22(3):225–232.