



## Short Communication

## Recent evolution of learnability in American English from 1800 to 2000



Thomas T. Hills\*, James S. Adelman

Department of Psychology, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

## ARTICLE INFO

## Article history:

Received 23 February 2015

Revised 11 June 2015

Accepted 19 June 2015

## Keywords:

Semantic bleaching  
 Language evolution  
 Concreteness  
 Historical linguistics  
 Culturomics

## ABSTRACT

Concreteness—the psycholinguistic property of referring to a perceptible entity—enhances processing speed, comprehension, and memory. These represent selective filters for cognition likely to influence language evolution in competitive language environments. Taking a culturomics approach, we use multiple language corpora representing more than 350 billion words combined with concreteness norms for over 40,000 English words and demonstrate a systematic rise in concrete language in American English over the last 200 years, both within and across word classes (nouns, verbs, and prepositions). Comparisons between new and old concreteness norms indicate this is not explained by semantic bleaching, but we find some evidence that the rise is related to changes in population demographics and may be associated with increasing numbers of second language learners or attention economics in response to crowding in the language market. We also examine the influence of gender and literacy. In sum, we demonstrate evolution in the psycholinguistic structure of American English, with a well-established impact on cognitive processing, which is likely to permeate modern language use.

© 2015 Elsevier B.V. All rights reserved.

For more than half a century, the central question of language acquisition has been how is the mind shaped to learn language (Chomsky, 1966; Pinker & Bloom, 1990). Recently this question has been inverted to ask how does language evolve to be learnable (Christiansen & Chater, 2008; Deacon, 1997; Smith & Kirby, 2008). Languages clearly do evolve over time (Labov, 2011; Lieberman, Michel, Jackson, Tang, & Nowak, 2007) and laboratory as well as computational studies have indicated that this evolution should be in the direction of learnability (Griffiths & Kalish, 2007; Hurford, 1989; Kirby, 2001; Smith & Kirby, 2008). To date, however, there is limited evidence for diachronic changes in learnability in a naturally occurring language. The challenge to documenting such changes rests in identifying a clear linguistic marker of learnability and tracking this marker in relation to diachronic corpora over a sufficient time scale and in a language environment likely to be under selective forces sufficient to produce linguistic change.

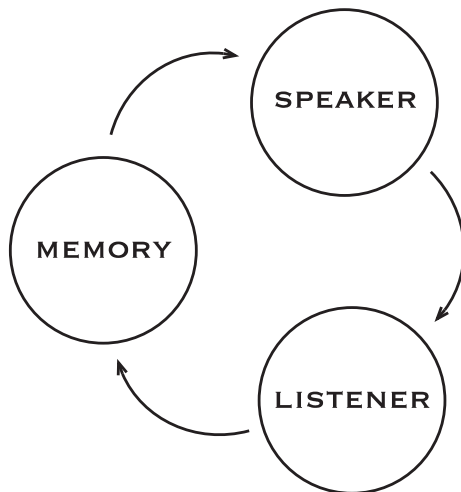
Two potential forces on the evolution of communication include the role of language learners and the influence of crowding. Language learners, because of the rate at which they learn linguistic properties, favor linguistic simplicity over complexity, and existing languages reflect this at multiple levels (Lupyan & Dale, 2010; Monaghan, 2014). Crowding may play a similar role. Recent trends in information production are likely to be associated

with analogous and highly competitive environments for communicators in English (Eppler & Mengis, 2004; Varian & Lyman, 2000). Information creates markets for attention (Davenport & Beck, 2001; Hansen & Haas, 2001) and in competitive situations these should drive evolutionary forces in communication. Similar to a global cocktail party problem—where communicators compete for the attention of listeners—crowding and the resultant attention economies should drive language evolution in the direction of learnability (Shannon & Weaver, 1949). This has been demonstrated in birds by an increase in the acoustic distinctiveness of bird song in response to crowding (Grant & Grant, 2010; Luther, 2009). If language responds to the number of competing language producers as well as the demands of language learners, then evolution in language should occur in communities where information crowding is most likely to be taking place.

It is a basic principle of evolution that replicators that have selective advantages in survival and reproduction should be more likely to be found in future populations. Analogously, language that comes to mind faster, is more easily understood, and is more readily remembered is more likely to be produced by future speakers. Each of these features is well known by psycholinguists to be associated with concreteness. Concreteness refers to the capacity for a concept to be vividly imagined and identified with a specific referent. Words like *dog* and *computer* are more vividly imagined than words like *truth* and *feeling*, and people easily report this difference. In studies of language processing, concrete words are more rapidly recognized as valid words in lexical decision tasks (James,

\* Corresponding author.

E-mail address: [t.t.hills@warwick.ac.uk](mailto:t.t.hills@warwick.ac.uk) (T.T. Hills).



**Fig. 1.** Stages where forces favoring learnability can select for concreteness in the cognitive life cycle of language. Because of cognitive capacity limitations, each arrow in the diagram represents a stage in the life cycle of language where selection may favor more learnable concepts: Speakers recall concepts from memory and then produce them for listeners, who must attend to the message and then store the information for later production. Selection will occur anytime there is more information to attend to, encode, and produce than language users can cognitively process.

1975), more easily recalled in memory tasks (Miller & Roodenrys, 2009; Romani, McAlpine, & Martin, 2008), and language composed of more concrete language is both more interesting and easier to understand (Sadovski, 2001). Concrete words are also more readily learned by both second and first language learners (e.g., De Groot & Keijzer, 2000). Together, these provide a cognitive basis for predicting selection for more concrete language during the stages of iterated-learning necessary for language evolution (e.g., Kirby, Cornish, & Smith, 2008; Smith & Kirby, 2008). In particular, these represent (1) what listeners attend to and comprehend, (2) what listeners encode, and (3) what speakers recall and, in turn, produce (see Fig. 1).

In the present work, we investigate historical changes in the distribution of concrete language in American English by using large collections of text combined with psycholinguistic measures of concreteness. Our approach follows on the “culturomics” approach of using “big data” to measuring behavioral and cultural change over time (e.g., Dodds, Harris, Kloumann, Bliss, & Danforth, 2011; Greenfield, 2013; Michel et al., 2011; Petersen, Tenenbaum, Havlin, Stanley, & Perc, 2012). In particular, we extend this approach to the study of a cognitive capacity-induced change in language by combining historical changes in text with recent large-scale collections of psycholinguistic norms (e.g., Brysbaert, Warriner, & Kuperman, 2013). This approach offers psychologists the opportunity to develop historical psychological accounts of behavioral change over cultural time and, as we demonstrate here, offers new questions and insights into how large-scale social change may influence communication in relation to cognitive processing.

## 1. Methods

To investigate diachronic changes in concreteness, we computed a concreteness rating over time for corpora using a recent collection of large-scale concreteness norms consisting of more than 40,000 words, each ranked for concreteness on a 5-point scale by multiple online participants (Brysbaert et al., 2013). These norms cover more than 50% of each of the corpora in any year. Concreteness ratings were combined with diachronic language

corpora, consisting of book collections archived in the Google Ngrams corpus of American English (355 billion words, see Michel et al., 2011), newspaper and magazine language collected in the Corpus of Historical American English (400 million words, see Davies, 2009), and inaugural addresses produced by American presidents. Presidential speeches represent much smaller language samples ranging from 487 words for Roosevelt in 1945 to 7189 words for Harrison in 1841. To investigate the influence of gender, we used a corpus of approximately one million words from 2943 blogs coded for gender of the author (1393 female, 1550 male; Mukherjee & Liu, 2010). We compared language data with historic records of U.S. immigration (Migration Policy Institute, 2014). In the supplemental online material, we provide similar results to those presented below for English Fiction and state of the union addresses.

The concreteness scores,  $C_y$ , for each year,  $y$ , in the diachronic corpora were computed as a weighted average,

$$C_y = \sum_i^n c_i p_{i,y}$$

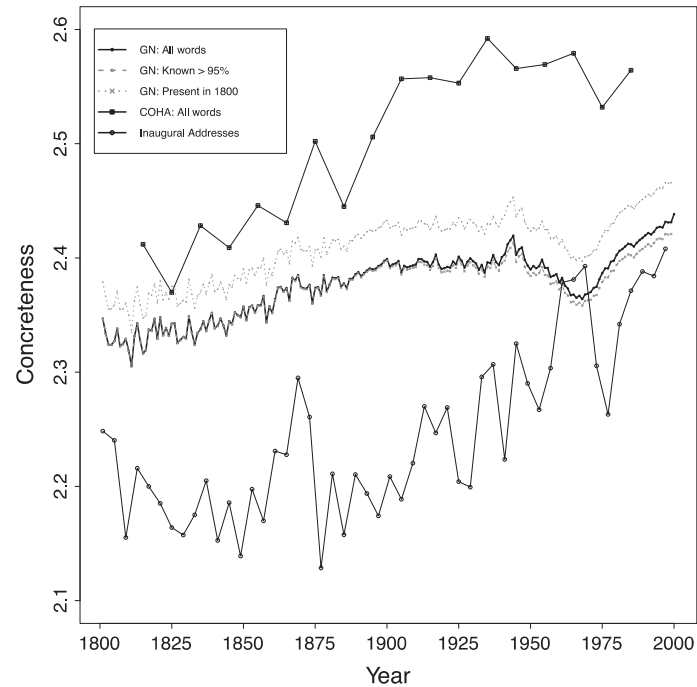
where  $c_i$  is the concreteness for word  $i$  taken from the concreteness norms (Brysbaert et al., 2013) and  $p_{i,y}$  is the probability of word  $i$  in year  $y$ . This was computed for all  $n$  words in the concreteness norms. We also computed concreteness on a per document basis, as opposed to per word as we do above; the same patterns of results were found as we report here.

## 2. Results

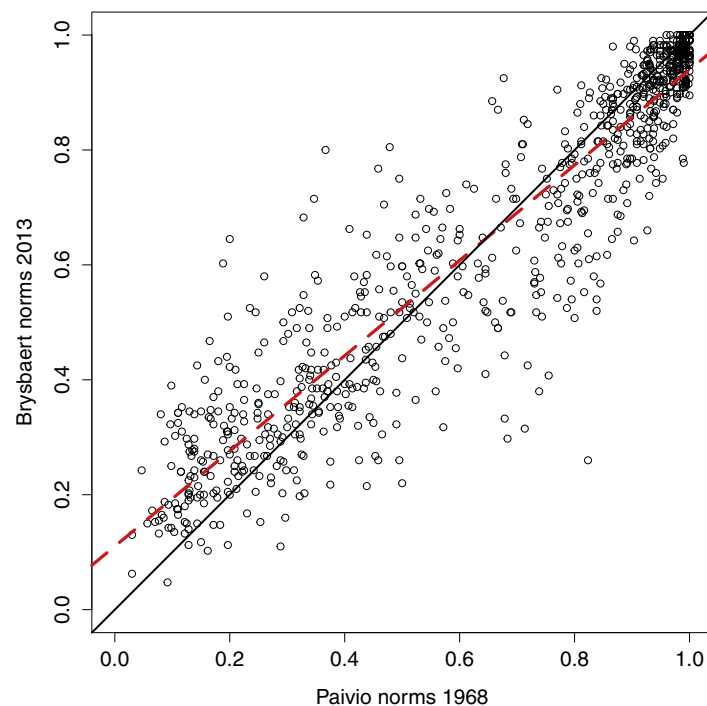
All the corpora we examined show an increase in concreteness over time (Fig. 2; see also Figs. S2 and S4). The Google Ngrams show this rise over the last 200 years, with a temporary drop following the Second World War. The Corpus of Historical American English and presidential inaugural addresses show a similar rise. Fig. 2 also shows the Google Ngram data with concreteness computed only for words present in the lexicon at year 1800 or using only those words that were known by more than 95% of participants in the concreteness norms (Brysbaert et al., 2013). In both cases, the rise in concreteness remains, indicating that the rise is not due to new words entering the lexicon or unfamiliar words leaving the lexicon.

One possible explanation for the general rise in concreteness is that some words may tend to lose concreteness over time. This is referred to as semantic bleaching, desemanticization, or grammaticalization (Aitchison, 2003; Hopper & Traugott, 2003). An example is the word *disaster*, which originally referred to dire and acute events, but can now usefully refer to topics as diverse as hairstyles and public policy. An additional example is the word *going*, which in the early nineteenth century lost its connection with motion and became a grammatical marker associated with events in the near future (Millar, 2015). Not all diachronic linguistic patterns are associated with semantic bleaching. Metonymy is a common figure of speech where some specific aspect of an abstract concept is used in place of a more abstract reference, as in *The White House*. Nonetheless, if bleaching were sufficiently strong, it might explain the observed rise in concreteness: if people used words in the past because they were more concrete at the time, then if these words became more abstract through use, and were therefore replaced with other more concrete words in more contemporary speech, the net result would be that older language appears more abstract, even though it may have been equally concrete. Note that if bleaching took place equally across all words, then this would create the opposite effect of reduced concreteness over time.

Concreteness norms were not collected as far back as the 1800s. However, the Paivio norms (Paivio, Yuille, & Madigan, 1968)



**Fig. 2.** Concreteness in the Google Ngrams, Corpus of Historical American English, and presidential inaugural addresses. The Google Ngram (GN) corpus of American English provides a collection of over 355 billion words published in books digitized by Google (Michel et al., 2011). We limit our analysis to years following 1800 to ensure appropriate representation. The Corpus of Historical American English (COHA), collected independently of the Google Ngram corpus, represents a balanced and representative corpus of American English containing more than 400 million words of text from 1810 to 1990, by decade, and composed of newspaper and magazine articles (Davies, 2009). Presidential inaugural addresses are computed from the entire text of the speech. Also shown is the Google Ngram corpus using only words that were present in the 1800 lexicon or only words that were known by more than 95% of participants in the concreteness norms (Brysbaert et al., 2013). Pearson correlations with time (95% CI): Google Ngrams,  $r = .86$  (.82, .89); COHA,  $r = .89$  (.73, .96); Inaugural Addresses,  $r = .77$  (.63, .86).



**Fig. 3.** Comparison of the Paivio and Brysbaert norms. The Paivio concreteness norms (Paivio et al., 1968) consist of 925 nouns, collected in the laboratory and using a 7-point scale. The Brysbaert et al. norms were collected on a 5-point scale. Both norms are normalized to be between zero and one. Each dot represents an individual word. The solid line shows the 45-degree line. The dotted line shows the best-fit regression line,  $r = .94$  (0.927, 0.943).

provide a 45-year window of comparison with the Brysbaert et al. norms. Normalizing the ratings for both shows that there are no systematic changes in word concreteness over the approximately

900 words used for comparison (Fig. 3). The median change is centered around zero and a paired  $t$ -test finds no significant difference in concreteness ( $t(876) = -0.79$ ,  $p = .45$ ). This suggests that a

**Table 1**

Concreteness for a selection of word classes in the Brysbaert et al. norms. Closed word classes are shown in *italics*.

Word class	Concreteness	
	<i>M</i>	<i>SD</i>
Names	3.73	0.86
Nouns	3.53	1.02
Numbers	3.49	0.38
Verbs	2.92	0.76
<i>Pronouns</i>	2.76	0.71
<i>Adjectives</i>	2.50	0.72
<i>Prepositions</i>	2.29	0.64
<i>Determiners</i>	2.11	0.55
<i>Adverbs</i>	2.06	0.53
<i>Articles</i>	1.66	0.54
<i>Conjunctions</i>	1.64	0.54

systematic change of concreteness within words is unlikely to explain the apparent rise in concreteness we see across words in the data.

Changes in the word classes people use may also change concreteness over time and may be influenced by the same selective forces described above. For example, articles—such as *the*, *a*, and *one*—have low relative concreteness in comparison with other word classes. Names, on the other hand—like *China* and *Butterfingers*—have the highest concreteness. Articles and names belong to closed and open word classes, respectively, which refer to the finite or potentially unlimited size of word classes. Closed word classes are generally less concrete than open word classes (see Table 1). If preferences for different word classes changed over time—e.g., if people dropped more articles and used more proper names—then this would increase concreteness.

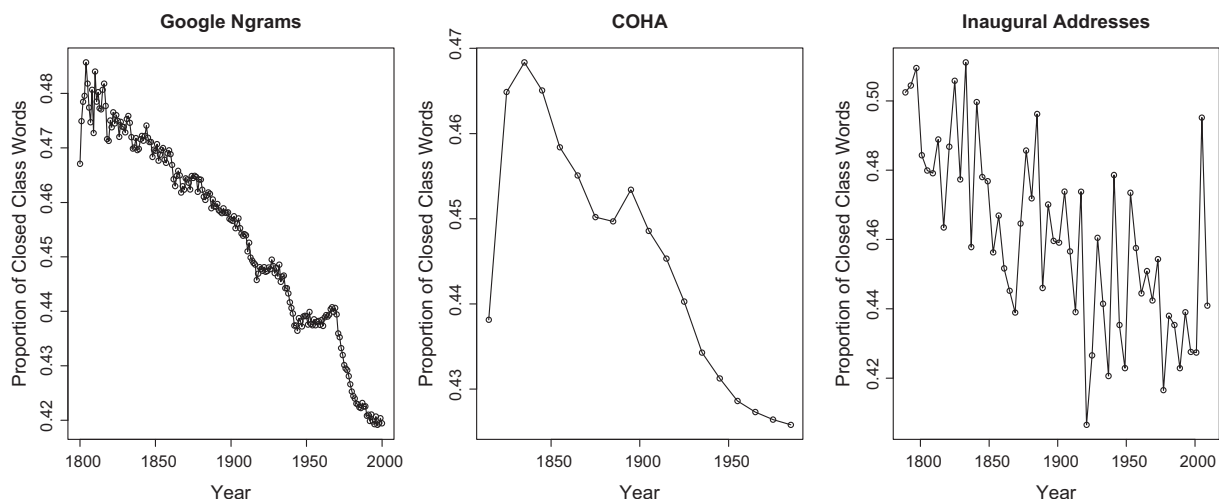
Consistent with our general hypothesis that American English has changed to increase its overall concreteness, closed class words have declined over the past two hundred years (see Fig. 4). These changes also explain some of the overall change in concreteness presented in Fig. 2 (predicting concreteness based on proportion of closed class words: Google Ngrams:  $\beta = -1.39$ ,  $R^2 = .74$ ,  $p < .001$ ; Corpus of Historical American English:  $\beta = -3.91$ ,  $R^2 = .57$ ,  $p < .001$ ; Inaugural Address:  $\beta = -1.79$ ,  $R^2 = .36$ ,  $p < .001$ ). However, as shown in Fig. 5, within the three word classes representing the largest portion of the data (nouns, verbs, and prepositions)—representing more than 50% of the words in each

corpus—concreteness rises (with some variation) over the period from 1800 to 2000, in all cases. Thus, concreteness also tends to increase within word classes, indicating that rising concreteness is systemic—operating at multiple levels within the structure of language.

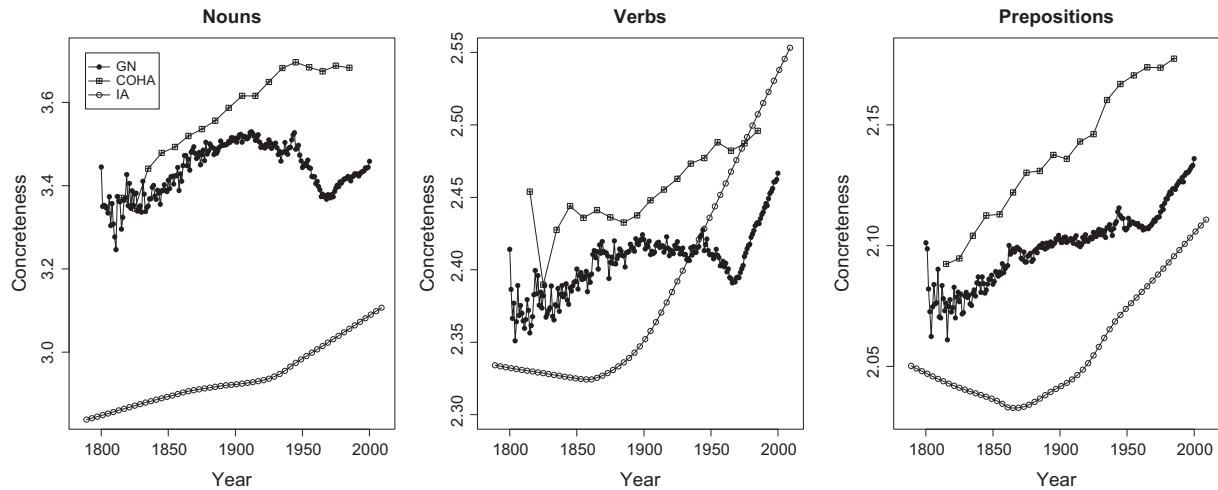
A potential driving force on the evolution of learnability in American English is related to the influx of second language learners. Though correlation does not imply causation, it is nonetheless useful with data of this kind to attempt to rule out such potential explanations. As shown in Fig. 6, the rise in concreteness in Google Ngrams and the Corpus of Historical American English are tightly correlated with U.S. immigration patterns. The Pearson correlation with Google Ngrams concreteness and total U.S. population is slightly weaker,  $r = .76$  (.42, .91). This is consistent with theories of language change in response to the cognitive abilities of language learners (e.g., Lupyán & Dale, 2010).

Several additional forces can be ruled out as complete explanations for the rise in concreteness. According to data from the National Center for Education Statistics, literacy has been increasing decade over decade in the U.S. since its earliest recorded date in 1870, with measures above 95% since the 1930s (Snyder, 1993). Comparing this decadal literacy data from 1870 to 1980 with our concreteness data, literacy is not statistically correlated with concreteness in the Google Ngrams,  $r = .45$  (−.16, .81), though it is with the Corpus of Historical American English,  $r = .77$  (.36, .93). Matching our analysis to the same time period, this is lower than the correlation between immigration and concreteness in the Google Ngrams,  $r = .76$  (0.32, 0.92), but approximately the same as that found between immigration and concreteness in the Corpus of Historical American English,  $r = .74$  (.29, .92). Immigration remains the stronger predictor, even though literacy may be a contributing factor.

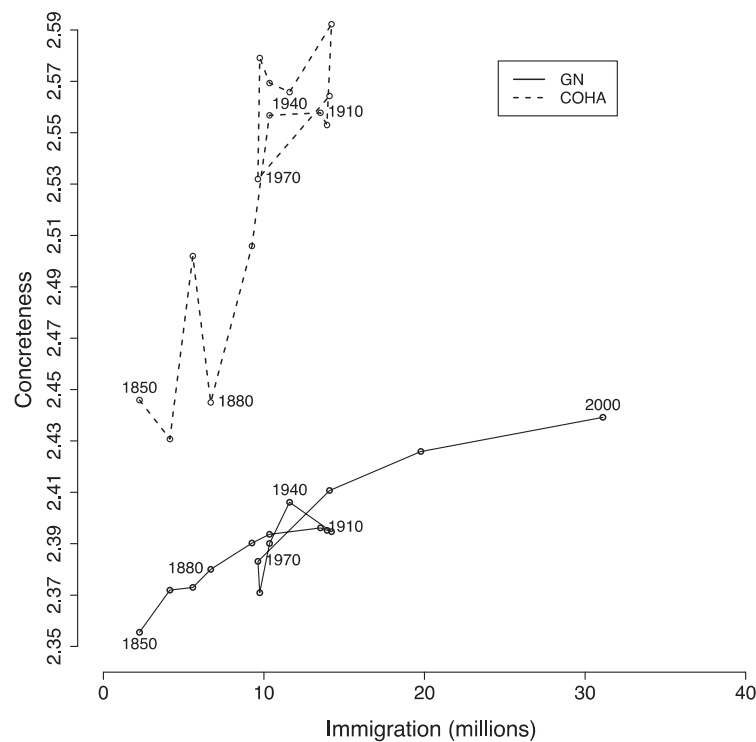
Gender may have also influenced the historical trajectory of language change. There has undoubtedly been an increase in the proportion of language media produced by females over the last two centuries. We applied our concreteness score to a corpus of blogs made available by Mukherjee and Liu (2010) and found that females' language was more concrete than males' language ( $M = 2.50$  for females and  $M = 2.48$  for males,  $t(2941) = 5.54$ ,  $p < .001$ , Cohen's  $d = 0.20$ ). The approximate difference between males and females in concreteness—0.02 units on the concreteness scale—is small relative to the overall rise in concreteness shown in Fig. 2. Changes in the ratio of males to females in language media is



**Fig. 4.** The decline in closed class words (determiners, prepositions, articles, conjunctions, and pronouns) over the last two hundred years in three corpora of American English. The open-classed words show a symmetrical increase. Pearson correlation with time (95% CI): Google Ngrams,  $r = -.98$  (−.99, −.97); COHA,  $r = -.85$  (−.94, −.63); IA,  $r = -.68$  (−.80, −.50).



**Fig. 5.** Concreteness within word classes (i.e., nouns, verbs, or prepositions) for the Google Ngram corpus (GN), the Corpus of Historical American English (COHA), and presidential inaugural addresses (IA). Because of the variation in the IA data, we present the smoothed data using a locally-weighted polynomial ('lowess' in R, scaling factor  $f = 2/3$ ). Pearson correlations with time (95% CI): nouns, GN,  $r = .39$  (.27, .50); COHA,  $r = .96$  (.88, .98); IA,  $r = .42$  (.17, .61); verbs, GN,  $r = .79$  (.74, .84); COHA,  $r = .86$  (.65, .95); IA,  $r = .68$  (.51, .80); prepositions, GN,  $r = .94$  (.92, .95); COHA,  $r = .99$  (.97, 1.00); IA,  $r = .17$  (-0.10, .41).



**Fig. 6.** Comparison of immigration patterns in the U.S. with concreteness computed from Google Ngrams (GN) and the Corpus of Historical American English (COHA). Pearson correlations between immigration and concreteness (95% CI): Google Ngrams,  $r = .96$  (.62, 1.00); COHA,  $r = .86$  (.61, .96).

therefore likely be a contributor to the rise in language concreteness, even if it only contributes to a fraction of the effect.

### 3. Conclusions

Though numerous accounts of language evolution have suggested that language should evolve to be learnable (Christiansen & Chater, 2008; Deacon, 1997; Smith & Kirby, 2008), the present work is the first to demonstrate the diachronic evolution of a property associated with learnability in a natural language. In particular, the evidence demonstrates a systematic rise over two centuries

in one of the most well understood psycholinguistic variables associated with learnability—concreteness. We further demonstrate that this effect appears to permeate language in a variety of forms (e.g., newspapers, books, and public speeches) and across word classes (e.g., nouns, verbs, and prepositions). We also show that this change is unlikely to be explained by a prominent observation from historical linguistics, semantic bleaching, because a comparison between concreteness norms collected 45 years apart shows no evidence for a systematic loss of concreteness among nouns, while during the same time period we observe a systematic rise in concrete nouns (see Fig. 5). Finally, we describe some suggestive



evidence that the observed changes in concreteness are unlikely to be fully explained by factors such as literacy and gender, but are nonetheless highly correlated with changes in population size and immigration. These latter results are consistent with recent hypotheses that language adapts to the cognitive demands of its users (e.g., Lupyan & Dale, 2010).

Culture is a marketplace for information. This leads to the prediction that any broad medium of communication should evolve over time in response to changes in the communication environment. Moreover, aspects of these signals should be enhanced in the presence of competition or among new learners. In particular, the distribution of information that is processed more quickly, better remembered, and more easily learned should increase as information markets become more crowded and involve more learners. In the present work we take a culturomics approach to address these issues—which has the strength of big data and the weakness of limited information about individual data points. This approach has substantial potential for addressing existing hypothesis in cognition as well as producing new lines of research. In particular, it is quite likely that concreteness is but one change among many in modern American English; changes in rhetorical style or topics of conversation that increase memorability, processing, and retrieval would all be equally consistent with reducing processing difficulty, which is the broader concept we hypothesize to be causative. These represent exciting research prospects and, in combination with laboratory studies on language memorability, also offer new inroads to understanding how to enhance the effectiveness of communication for key messages in competitive language environments such as in education or emergency situations. In addition, language efficiency is likely to have pronounced effects on the cognitive capacities of its users. For example, the documented rise in symbolic intelligence associated with the Flynn effect (Flynn, 2012; Fox & Mitchum, 2013) could be both cause and consequence of increased concreteness in language. That is, an increase in concrete language may have enhanced our individual capacity to process complex information (Sadoski, 2001) at the same time as it allows us to communicate clear but increasingly complicated ideas.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2015.06.009>.

## References

- Aitchison, J., & Lewis, D. M. (2003). Polysemy and bleaching. In B. Nerlich, Z. Todd, V. Herman, & D. Clarke (Eds.), *Polysemy: Flexible Patterns in the Mind* (pp. 253–265). Berlin: Mouton De Gruyter.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904–911.
- Chomsky, N. (1966). *Language and mind*. New York: Harcourt Brace Jovanovich.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489–509.
- Davenport, T. H., & Beck, J. C. (2001). *The attention economy: understanding the new economy of business*. Cambridge, MA: Harvard Business School Press.
- Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008+): design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14, 159–190.
- De Groot, A., & Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, 50, 1–56.
- Deacon, T. W. (1997). *The symbolic species: the co-evolution of language and the brain*. New York: Norton.
- Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. *PLoS One*, 6(12), e26752.
- Eppler, M. J., & Mengis, J. (2004). The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20, 325–344.
- Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the twenty-first century*. Cambridge University Press.
- Fox, M. C., & Mitchum, A. L. (2013). A knowledge-based theory of rising scores on “culture-free” tests. *Journal of Experimental Psychology: General*, 142, 979–1000.
- Grant, B. R., & Grant, P. R. (2010). Songs of Darwin’s finches diverge when a new species enters the community. *Proceedings of the National Academy of Sciences*, 107, 20156–20163.
- Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24, 1722–1731.
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31, 441–480.
- Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, 46, 1–28.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.
- Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77, 187–222.
- James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 130.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5, 102–110.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105, 10681–10686.
- Labov, W. (2011). *Principles of linguistic change, cognitive and cultural factors*. John Wiley & Sons.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713–716.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, 5, e8559.
- Luther, D. (2009). The influence of the acoustic community on songs of birds in a neotropical rain forest. *Behavioral Ecology*, 20, 864–871.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176–182.
- Millar, R. M. (2015). *Trask’s historical linguistics*. London: Routledge.
- Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, 37, 850–865.
- Migration Policy Institute. (2014) <[www.migrationpolicy.org/](http://www.migrationpolicy.org/)>. Accessed: 30.09.14.
- Monaghan, P. (2014). Age of acquisition predicts rate of lexical evolution. *Cognition*, 133(3), 530–534.
- Mukherjee, A., & Liu, B. (2010). Improving gender classification of blog authors. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 207–217). Cambridge, MA: Association for Computational Linguistics.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76, 1–25.
- Petersen, A. M., Tenenbaum, J. N., Havlin, S., Stanley, H. E., & Perc, M. (2012). Languages cool as they expand: allometric scaling and the decreasing need for new words. *Scientific Reports*, 2.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707–727.
- Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, 61, 292–323.
- Sadoski, M. (2001). Resolving the effects of concreteness on interest, comprehension, and learning important ideas from text. *Educational Psychology Review*, 13, 263–281.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 3591–3603.
- Snyder, T. D. (1993). *120 years of American education: a statistical portrait*. Washington, D.C.: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
- Varian, H. R., & Lyman, P. (2000). *How much information*. University of California, Berkeley: UC Press.