How cognitive selection affects language change

Ying Li[a,b,1] (iD), Fritz Breithaupt[c,d,1] (iD), Thomas Hills[e] (iD), Ziyong Lin[f] (iD), Yanyan Chen[a,g], Cynthia S. W. Siew[h], and Ralph Hertwig[b] (iD)

Like biological species, words in language must compete to survive. Previously, it has been shown that language changes in response to cognitive constraints and over time becomes more learnable. Here, we use two complementary research paradigms to demonstrate how the survival of existing word forms can be predicted by psycholinguistic properties that impact language production. In the first study, we analyzed the survival of words in the context of interpersonal communication. We analyzed data from a large-scale serial-reproduction experiment in which stories were passed down along a transmission chain over multiple participants. The results show that words that are acquired earlier in life, more concrete, more arousing, and more emotional are more likely to survive retellings. We reason that the same trend might scale up to language evolution over multiple generations of natural language users. If that is the case, the same set of psycholinguistic properties should also account for the change of word frequency in natural language corpora over historical time. That is what we found in two large historical-language corpora (Study 2): Early acquisition, concreteness, and high arousal all predict increasing word frequency over the past 200 y. However, the two studies diverge with respect to the impact of word valence and word length, which we take up in the discussion. By bridging micro-level behavioral preferences and macro-level language patterns, our investigation sheds light on the cognitive mechanisms underlying word competition.

language evolution | serial reproduction | psycholinguistics | corpus analysis | age of acquisition

> A struggle for life is constantly going on among the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue.
>
> M. Müller (1)

Languages change due to social, cultural, and cognitive influences. Social and cultural environments can often evolve rapidly due to the impact of, for example, war, pandemic disease, technological innovation, population changes, or changes in norms and values. By contrast, the mind's cognitive structures and constraints, due to its deep roots in our biological endowment, are much more stable over time and are therefore potentially capable of imposing lasting and coherent impacts on language evolution. Cognitive selection during interpersonal communication, even when its effects are very small, may be magnified into macro-level patterns of linguistic phenomena through generations of language speakers (for a review of evidence from experimental studies, see ref. 2). In this research, we asked how cognitive selection affects the preservation of English word forms. We addressed this question by exploring a set of psycholinguistic factors that impact the survival of word forms in two studies: first in the context of interpersonal communication using data from a large experiment on story retellings and second in the context of diachronic language change over 200 y using three large historical language corpora.

One approach to the study of language change is to view language as a complex "organism," constantly evolving under selection pressures from human learning and processing mechanisms over generations of language speakers (3, 4). This view on the cultural evolution of language can be traced back to Darwin (5), who wrote in The Descent of Man: "The formation of different languages and distinct species, and the proofs that both have been developed through a gradual process, are curiously the same... The survival and preservation of certain favored words in the struggle for existence is natural selection." Evidence from modern science supporting this view has been accumulating from various research paradigms, including computational simulations (6, 7), laboratory experiments of artificial languages (8), and analysis of natural language corpora (9–12). In light of this work, the present study combines laboratory and historical language approaches to examine the preservation of word forms (not morphological or semantic changes). Therefore, the language change process we

Significance

Words compete for survival in each language. This research investigates the cognitive mechanisms underlying the rise and fall of English word forms using two complementary research paradigms: a serial reproduction experiment where a story is passed on along a diffusion chain and a quantitative linguistic analysis of historical corpora that spans the past two centuries. We found that the competition among word forms is closely associated with how humans use language: words that are acquired earlier in life, more concrete, and more arousing are more likely to survive. Our results suggest that micro-level patterns of language production may scale up to macro-level patterns of language change over generations of language speakers.

described has several divergences from natural evolution, including lack of fitness-blind mutation and accumulation of mutations in a lineage that result in the adaption of individuals.

From Individual-Level Cognitive Selection to Diachronic Language Change. One commonly used experimental paradigm aimed to explore language evolution is serial reproduction—a process whereby information produced by one individual is observed by another, who then communicates the information to a third individual, and so on along a transmission chain. This method of serial reproduction harks back to the works by early British anthropologists (as explained in ref. 13) and became popular in psychology through the work of Bartlett (14). Bartlett measured change over successive reproductions of stories and analyzed what components of the original story are forgotten or retained. Those that survived the test of time should, by definition, be more memorable. Both the original Bartlett study and its recent follow-ups found that people selectively retain information that is coherent with the narrator and re-narrator's prior conceptions (schemata) and therefore reproduce and even distort information in ways that assimilate the stories to the schemata available to them (e.g., refs. 15–18). For example, when folk tales from another culture, "weird stories" for Bartlett's British audience, were passed down a transmission chain, many of the folk tales' supernatural elements were removed and replaced (14). Similarly, Bangerter (16) found that abstract scientific concepts unfamiliar to participants gradually become anthropomorphic after repeated iterations. These studies suggest that there are a set of underlying cognitive factors that impact which information is kept, removed, or distorted when it is transmitted from one person to the next. Similarly, cognitive selection—the mind's tendency to preferentially attend to certain kinds of information (19)—may systematically favour words with certain properties in daily communication and, over time, impact the survival of words on a longer time scale of language change.

Recent decades have seen a surging interest in explaining how individual-level behaviors result in population-level linguistic phenomena. This interest is deeply inspired by work and theorizing in evolutionary biology that investigates links between small-scale micro-evolutionary processes (e.g., natural selection and mutation) and population-level macro-evolutionary patterns in time and space, with the latter patterns generated in part by the former. The same rationale can be applied to language evolution. The advantage of adapting a serial-reproduction approach to language evolution is that it bridges the gap between behavioral patterns in small-scale information transmission that can be observed in laboratory experiments and population-level patterns arising from generations of language users in real life (20). Previous laboratory studies using the method of serial reproduction to explore questions of language change typically focus on whether inter-personal communication is sufficient to give rise to the basic design features of human language, such as the emergence of abstract symbols (21, 22), compositionality[*] (8), arbitrariness[†] and systematicity[‡] (23). Few studies, however, have investigated the survival of word forms of real language in serial-reproduction experiments.

Psycholinguistic Properties that Predict Language Change. The present study focuses on the evolutionary success of word forms. Words compete with each other for the limited cognitive capacities of language users. Successful words are those used in high frequency, whereas words that are no longer in use go extinct. What makes some words more likely to survive than others? To answer this question, consider the cognitive life cycle of information: To make it from one individual to another, information must be attended to, comprehended, encoded in memory, and later reproduced (10). Correspondingly for the life cycle of language communication, words that are favored at each of these stages should, ceteris paribus, be used more frequently and be less likely to become extinct. Below we review a set of psycholinguistic properties that impact each of the aforementioned stages of language communication.

Age of acquisition (AoA). Words that are acquired early in childhood have more stable representations, making early acquired words easier and more accurately retrievable (24), and more resistant to aging-related language depletion (25) and cognitive impairment (26, 27). Consequently, early acquired words are more resistant to evolutionary change in both form (9) and meaning (12).

Concreteness. Concrete words are those that refer to specific objects or entities that can be perceived through our senses. Following this definition, all words are abstract: Even a word like dog does not refer to any individual dog that one can see, touch, smell, hear, or interact with. However, some words seem to be more concrete than others, and this might mean that they seem to be close to our senses (when I see a dog, the word dog comes to mind), are more vividly imagined than others (most people can picture a dog), or seem less abstract (dog in comparison to animal). That is, while the precise standard of concreteness is elusive, people reliably agree on more or less concrete words.

Compared to abstract words, words that people rate as more concrete are more rapidly recognized as valid words (28), more easily recalled in memory tasks (29, 30), and easier to learn (31). Language composed of more concrete words is also more interesting and easier to understand (32). These results together provide a cognitive basis for predicting a selective advantage for more concrete language, relative to more abstract language, during the life cycle of language communication. Consequently, the advantage for concrete words in language communication predicts an increased usage frequency of concrete words in the long run (10, 33).

Emotion. The emotional connotation of words affects attention, processing, and memory. Emotions are characterized by two primary and orthogonal dimensions, namely, valence and arousal (34). Negative words can be either calming (e.g., dirt) or arousing (e.g., snake), and positive words can also be calming (e.g., sleep) or arousing (e.g., sex). Both negative and arousing stimuli engage attention longer than other stimuli (35). Consequently, lexical decisions[§] often take more time for negative words (36, 37) and arousing words (36).

When processing the meaning of individual words, words with greater emotionality—nearer the ends of the valence scale (positive or negative)—are easier to process (38, 39); only for abstract words. Similarly, words with more emotionality are better remembered (40, 41). Though the above studies found null effects for arousal in language, when examining pictures, videos, or events, arousal facilitates memory, in fact, current theories often attribute emotional memory enhancement mostly

---

[*]Key design feature of language whereby the meaning of an expression is a function of its constituent parts and the way the constituents are combined.

[†]A sign is arbitrary when there is no inherent relationship between its form and its meaning.

[‡]Key design feature of language whereby a feature (e.g., house) that is common to multiple items (red house, blue house) is represented by the same sign (house).

[§]i.e., Participants decide whether a string of letters is a word or nonword. It requires correct recognition of word forms. It does not necessarily involve semantic processing.

to arousal (42–44), with valence having little or no influence independent of arousal (45).

The precise role of emotion on language change remains unclear. Part of the challenge in exploring this question is that language change responds not only to human cognitive constraints but also to shifting external socio-cultural environments. This issue may be more salient for valence than other linguistic properties. Consider, for instance, the evidence that life satisfaction has increased while the rate of violence has gone down, especially since the end of World War II (46). Hence, one may speculate that these trends imply less demand for negative words. Consistent with this, Hills et al. (47) found that the average valence of historical language predicts change in national well-being in four countries. Therefore, change in the prevalence of positive and negative emotional words may also reflect cultural changes in their relative need.

Overview of the current study. We used two studies to explore the role of psycholinguistic properties on the competition among word forms. In Study 1, we used data from a serial-production experiment in which stories were passed along a transmission chain of three generations of narrators. Based on the reviewed psycholinguistic studies on language communication, we hypothesize that words that are more concrete and acquired earlier in life have greater chance of survival in the transmission chain. Additionally, we explore the role of emotional connotations (i.e., valence and arousal) without a clear hypothesis due to mixed prior findings on how emotion impacts various stages of language communication. In Study 2, we test whether findings from Study 1 can be extrapolated to the longer time scale of language change. We hypothesize that the same set of psycholinguistic properties that predict word survival in the serial-production experiment also predict the increase of word frequency in natural language corpora over the past 200 y.

Results of Study 1

We used a mixed effect logistic regression model to predict preservation of words in a serial-reproduction story retelling task (Table 1). No evidence of multicollinearity was found (*SI Appendix,* section 1). The model shows that all independent variables (i.e., AoA, concreteness, arousal, emotionality, length, frequency) except valence are significant predictors of whether a word is preserved in a story retelling. Overall, the model explains 29.6% of the variance (Conditional R^2), with fixed effects explaining 11.7% of the variance (Marginal R^2) and random effects explaining the remaining 17.9% of the variance. Among the four random effects, the emotion category of the initial story (joy, sadness, disgust, embarrassment, risk) has a much smaller SD than the other three, suggesting that there is little variability in word preservation across different emotion categories. In contrast, there is much larger variability in word preservation across different stories, different retellers, and words of different grammatical categories. Further analysis shows that nouns and verbs are more likely to be preserved than other grammatical categories (*SI Appendix,* Fig. S1).

Looking at the fixed effects, the odds ratio shows the relative change in the likelihood of word preservation for a one SD increase in the predictor. The closer the odds ratio is to 1, the smaller the effect of the predictor. Age of acquisition (AoA) shows a small but significant effect on word preservation, with a one SD increase in AoA corresponding to a 2% decrease in the likelihood of word preservation. This suggests that earlier acquired words (smaller AoA) have a larger chance of being

Table 1. Logistic regression model that predicts preservation of words in the story retelling task (Study 1)

Fixed effects	Odds ratios	95% CI	P
(Intercept)	0.39	0.30 to 0.50	<0.001
Age of acquisition	0.98	0.97 to 0.98	<0.001
Valence	1.00	0.99 to 1.00	0.568
Arousal	1.17	1.16 to 1.18	<0.001
Concreteness	1.45	1.44 to 1.73	<0.001
Emotionality	1.55	1.39 to 1.73	<0.001
Length	1.05	1.04 to 1.06	<0.001
Word count	1.71	1.70 to 1.73	<0.001
Log frequency	1.05	1.04 to 1.06	<0.001
Iteration	1.34	1.29 to 1.40	<0.001

Random effects	Variance	SD	Number
Story ID	0.23	0.48	9265
Individual ID	0.49	0.7	6126
Grammatical category	0.10	0.31	9
Emotion category	0.01	0.09	5
Observations		537,055	
Marginal R^2		11.7%	
Conditional R^2		29.6%	

Note: All variables (except iteration) are scaled and centered. Word count refers to number of times a word appears in the input story. Iteration refers to the position of the stories in the diffusion chain.

preserved in the context of a story-retelling task. Concrete words are also more likely to be preserved, with a one SD increase corresponding to a 45% increase in the likelihood of word preservation. In addition, words of greater emotionality (as opposed to emotionally neutral words) and words of higher arousal are found to have a better chance of being preserved. Valence, however, is not a significant predictor, suggesting that negative words do not have a selective advantage over positive words.

We also entered word length, count (number of times a word appears in the input story), and word frequency as control factors. The results show that longer words, words that appeared multiple times in a story, and words of higher frequency in the Google Ngram Corpus are more likely to be preserved. Last, we include iteration number as the only non-linguistic factor. It shows that the probability of word preservation increases as stories were passed down the diffusion chain. This suggests stories become increasingly stable after each additional retelling. Also, stories became shorter in the process of retelling, a process referred to as "leveling" (48).

Results of Study 2

We used general additive model to predict the increase of logged word frequency from year 1800 to year 2000 (Table 2). We reasoned that an upward frequency trend signals greater advantage in the competition for existence while a plummeting frequency foreshadows extinction. The combined model with both the Corpus of Historical American English (COHA) and the Google Fiction accounts for 43.6% of the variance. We did not include grammatical category as a random effect because it complicates the model while only explaining 0.6% additional variance.

Fig. 1 shows the partial effect of each psycholinguistic property on change of word frequency, with shading indicating 95% CIs. We also examined the one-to-one relationship between

Table 2. General additive mixed model that predicts increase of word frequency from 1800 to 2000 (Study 2)

Combined (Fiction & COHA)

Smooth terms	Estimated DF	F value	P-value
Age of acquisition	4.46	1467.7	<0.001
Length	4.21	61.2	<0.001
Valence	4.95	81.6	<0.001
Arousal	4.95	43.3	<0.001
Concreteness	5.09	46.1	<0.001
Log frequency (year 1800)	6.97	1436.5	<0.001
Number of observation	23,306		
Marginal R^2	0.367		
Conditional R^2	0.436		

Note: All variables are scaled and centered.

the change of word frequency and each independent variable separately and present the results in the embedded miniature plots.

For each variable, in order to determine which region of the slope was significantly increasing or decreasing, we determined the derivatives and SEs for the derivatives along the gradient of the slope. The derivative at one point of a curve means the slope of the tangent line at that point. Where the 95% CI for the derivatives does not overlap zero, there is a significant increase or decrease in the slope. Fig. 1 shows the region of the slopes that indicate significant change with a solid line and regions of slope that show no significant change with a dashed line. Therefore, the advantage of general additive model is that it shows in which specific range an independent variable has significant impact on the dependent variable.

The same scale of the y-axis is used for all subplots for easier comparison of the impact on frequency change: The steeper the slope, the larger the impact. Consistent with Study 1, we found that earlier acquired (smaller AoA value), more concrete, and more arousing words were more likely to increase in frequency. Unlike in Study 1 where AoA has a relatively small effect on word survival, when looking at language change across two centuries, early acquisition favors word preservation. The effect of AoA is monotonic across its range. In contrast, the effect of concreteness and arousal on change of word frequency is less straightforward. For low- to middle-concrete words, the relationship is positive

and monotonic, with increasing concreteness relating to greater increase in word frequency. However, for middle- to high-concrete words, there is no significant effect of increased word concreteness on the change of word frequency. Similarly, the main influence of word arousal occurs at the two ends of the arousal scale.

Word length and valence show divergent patterns from Study 1 (Fig. 1). Word length has a monotonically declining effect on word frequency, with longer words seeing greater reductions in word frequency. More positively valenced words also see a greater increase in their frequency, with the change primarily occurring near the middle of the valence range. High valence words increase in frequency more than low valence words.

We also found that words of lower frequency at year 1800 are more likely to increase in frequency over the past two centuries. This pattern is indicative of regression to the mean. This is counter-intuitive at first glance because one may predict the opposite pattern given the negative correlation between log frequency in 1800 and age of acquisition. Caution must apply when interpreting this result because our data only contain words that are still in use in recent decades and, therefore, systematically exclude words that are no longer in use (simply because linguistic features such as AoA and valence norms are only available for words in use today). Therefore, we include frequency in the model simply as a control factor in this analysis.

General Discussion

What makes one word more likely to survive than another? We argue that the answer to this question lies in both micro-level patterns of language usage and macro-level patterns of language change, with the effect of the former possibly being magnified into the latter through the behavior of generations of language users. We took a methodologically pluralistic approach to examine this question. In Study 1, we examined word survival rate in a serial-reproduction experiment where people read stories and retold them to others. In Study 2, we analyzed the rise and fall of word frequency across 200 y, using English corpora. Both studies converged on the same conclusion: Early acquisition, concreteness, and high arousal are the linguistic properties that give a word a selective advantage. The two studies diverged on the effect of word length and emotional valence, which we discuss
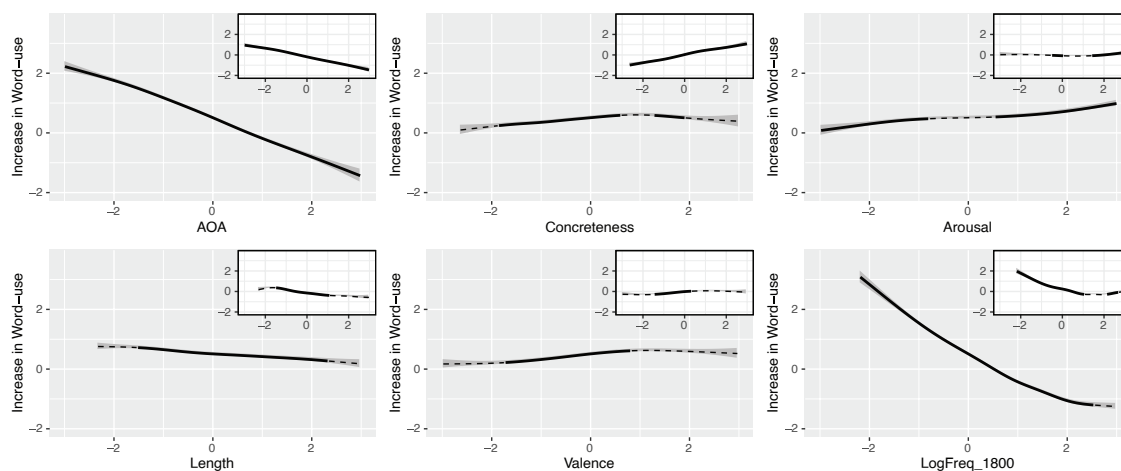

Fig. 1. The partial relationship between log frequency changes between 1800 and 2000 (y-axis) and six psycholinguistic properties (x-axis). The inset plots show one-to-one relationship between log frequency changes and each psycholinguistic property. All independent variables are scaled and centered around 0. Lines are model estimates (dotted lines are estimates and solid lines indicate where the estimates are significantly increasing or decreasing). Shading indicates 95% CIs around the estimates.

further below. Our results provide empirical evidence supporting the emerging language theory that suggests languages become continuously easier for humans to learn, process, and produce (3, 4). Moreover, our result is strengthened by its strong external validity, achieved by using large-scale data, examining survival of words in context (instead of learning and producing individual words independent of context), and validation of results across different methodologies and time frames.

Words change in both form and meaning. While the current study focuses on the preservation of words, our recent work (12) examined the underlying driving force of diachronic semantic change. As languages face challenges of expressing an infinite range of ideas using a finite set of words, a common strategy is to assign new meanings to existing words. The authors found that words that are acquired later in life, less arousing, and more abstract are more likely to host new meanings. In addition, they also found mixed results on the role of length and valence: The effect varied depending on which corpus (Google Ngram All English or COHA) was used to quantify semantic change. Together, their study and ours demonstrate that early acquisition, higher arousal, and concreteness of words function as a preserver against change in semantics and protect against extinction.

In Study 2, we found longer words were used less often over the past 200 y. This is consistent with previous findings that longer words are more likely to change in their lexical form (9) and more likely to be loan words borrowed from another language (49). While loan words and morphological change in word forms are common in language evolution over hundreds or thousands of years, they are less likely to occur in an experiment involving story retellings. That may explain why Study 1 did not find longer words are more likely to be dropped. On the contrary, possibly because longer words in study 1 are more salient and informative to the gist of the story, they are more likely to be preserved when one retells a story. Consistent with this speculation is evidence that longer words often encode greater conceptual complexity (50).

Why does the effect of emotional valence on language production differ across the two studies? The divergent result is a reminder that language change does not only follow cognitive mechanisms in language production; it also responds to the shifting sociocultural environment. A world that has become unprecedentedly peaceful and safe (46) is likely to reflect itself in language, with greater demands for positive words and less for negative words (47). Perhaps that is why we observe that positive words experience larger increases of word frequency over the past two centuries. In contrast, in the serial-reproduction experiment where the influence of historical improvement of life satisfaction is precluded, we found emotionality (i.e., distance from emotionally neutral state), instead of valence, predicts survival of words. This result is consistent with previous research showing that emotionally extreme words are better remembered (40) and easier to process (38) than neutral words.

We also have a serendipitous finding on arousal: Both studies converge on the selective advantage of arousing words. This result is in line with current theories that attribute emotional memory enhancement to arousal (42–44). The underlying assumption is that our limited memory resources are preferentially allocated to significant stimuli (51), with arousal acting as a primary index of behavioural significance (52). However, it is worth pointing out that most studies that explored the role of arousal on memory used non-verbal input (e.g., images, videos). Studies using words as stimuli are scarce and often find null effects of arousal on memory

(40) and semantic processing (38). It seems that the effect of arousal is harder to detect when words are used as stimuli. This is probably because individual words out of context are substantially less arousing than images and videos, so that arousal levels were generally too low to exert effects. In contrast, both of our studies examined survival of words within the context of meaningful communication. With an unprecedented large amount of data, we are able to detect effects of arousal on word survival, and this effect is larger than emotionality.

Why have the patterns of language change observed here not stabilized? That is, why has language evolution not already arrived at a stable equilibrium among the various cognitive selective forces for word features observed in Study 1? Recent studies suggest that English has been under greater force of cognitive selection over the last 200 y. This is likely caused by an increasing amount of language production facilitated by more numerous and faster means of language communication as well as a growing population of English language speakers (10). Partly isolated populations similar to Darwin's finches may also have accelerated the production of new words that may have been brought back to the common language pool by new forms of communication. If language production exceeds human attention capacity, this excess increases competition and what Darwin (53), speaking of biological species, called the "struggle for existence": As a result, selective forces become stronger. For our purposes, more competition for language means that more words must compete to pass through the life cycle of language communication, successfully acquiring listeners who later become producers, much like a parasite seeking a host. The consistency we observe between individual story retellings and the historical change of word frequency is coherent with this language competition hypothesis. That is, the pattern of language change we observe over historical time finds its root in individual cognitive selection for words that are easier to process, recall, and reproduce: Words that more successfully achieve that are, by definition, better adapted for the life cycle of language production under times of heavy competition.

Clearly, this is not the only driver of language change—the English language is not baby talk. But what forces enrich it? These could include the adaption to changing social, emotional, and technological environments. It also includes the ever-changing demand for expressing new ideas and cutting through time-worn phrasings that have lost their ability to engage the listener. In addition, speech communities may be enjoying an increasingly diverse set of subgroups, each of which creates technical terms, jargon, and in-group slang to satisfy their idiosyncratic communicative needs. These new words may replenish the vocabulary that cognitive selection culls in times of increasing competition. Future research could extend our analyses to include not only existing words during a historical period of time, but also word births and deaths, to explore how the dynamics between the two forces maximize a language's communicative power under cognitive constraints.

Finally, future research could delve one step deeper by investigating what might concreteness, early acquisition, and arousal have in common that lead to their influence on word preservation. One possibility is that they better engage cognitive resources, leading to more elaborated cognitive representations (54). Concreteness as proposed by Paivio (55) better engages multiple representational pathways (such as visual, auditory, and sensory motor representations), leading to, for example, the picture superiority effect. Arousal may further elaborate concepts through their relationships with emotional centers in the brain.

And age of acquisition may reflect more foundational terms, which—because they are learned earlier—are more broadly associated with other terms. Therefore, each of these linguistic properties, though correlated, may capture a different aspect of cognitive elaboration, which helps these terms to persist in the face of repeated use, under a wide variety of contexts, and in environments of potentially increasing competition.

Materials and Methods

Measures of Psycholinguistic Properties.

Age of acquisition. To evaluate the age at which a word is learned, we used the Age of Acquisition ratings (AoA) collected by Kuperman et al. (56). In their study, participants were asked to recall the age (in years) at which they thought they had learned a word[¶]. While memory can never be perfectly accurate (Kuperman et al. noted that people underestimate vocabulary growth before age of 3 and overestimate vocabulary growth between 9 and 14), a follow-up study shows that the subjective, retrospective AoA estimates still have a high correlation with a more objective measure based on vocabulary assessment among pupils [Pearson $r = 0.76$, $N = 18,139$ (57)].

Emotion and concreteness. Warriner et al. (58) provided valence and arousal[#] ratings for 13,915 English words; and Brysbaert et al. (59) provided concreteness ratings for 37,058 English words. Using a neural network model to capture word meanings in vectors, Hollis et al. (60) extrapolated the above human judgments of words to over 70,000 English words. The out-of-sample correlation between human ratings and machine-learnt ratings is 0.79 for valence, 0.61 for arousal, and 0.83 for concreteness.

In the present study, we take valence, arousal, and concreteness norms from Hollis et al. (60). In addition, we compute emotionality by taking the absolute value of the difference between the word's valence and the average valence in the dataset so that the most negative and the most positive words have the largest scores on emotionality.

Frequency. Log frequency was retrieved from the Google Ngram Books Corpus (61). It has a strong correlation with log frequency derived from two other corpora: Corpus of Historical American English [COHA (62); Pearson $r = 0.80$] and British National Corpus [BNC (63); Pearson $r = 0.91$].

Length. Length of word was computed by counting the number of letters in a word. We also extracted grammatical category for each word using the NLTK python package (64).

Study 1.

Serial reproduction experiment. The story retelling data were collected by Breithaupt et al. (65). Participants were recruited from Amazon Mechanical Turk. In one of their studies, participants were asked to read stories created by researchers (seed stories) and retell the stories in their "own words." They typed the retellings on their own computer. These retold stories were then presented to the next generation of narrators who performed the same task. In total, each seed story was passed down through 3 generations of narrators. Following these procedures, the 97 seed stories were retold for a total of 2,695 times in the first iteration, 6,474 times in the second iteration, and 7,428 times in the third iteration. (see *SI Appendix*, section 5 for more details).

Analysis. For each story retelling, we first extract word types (the distinct word forms in a text). Whether a word type in the input story is preserved in the output story was taken as the dependent variable, with preservation coded as 1 and drop-out coded as 0. The set of psycholinguistic properties mentioned

above were included in a regression model as predictors. In addition, we also included two variables: the number of times a word appeared in the input story and the position of the stories in the transmission chain (This is denoted as iteration in Table 1. The value takes 1, 2, or 3.). We included participants' ID, story ID, emotion category, and tagged part of speech as random effects. Participants' ID was included as a random effect to model individual differences among participants whereby some participants use more words from the input story than others (due to better memory or devoting more attention to the task). Story ID as a random effect models the idiosyncratic features of a story that makes its words more likely to be reproduced than other stories. Moreover, since the original stories were created by researchers to contain one of the five emotion appraisals, we entered emotion category as a random effect to model potential differences in the role of different emotions on the probability of a word's preservation. Last, we include grammatical category as a random effect since it has been found to predict language evolution, such as borrowing of loan words between languages (49) and change of lexical forms (9).

Study 2.

Historical corpora. To evaluate the life and death of words over historical time, we extracted word frequency from historical language corpora. Three corpora were used: the Corpus of Historical American English [COHA (62)] and two corpora from Google Books Ngram corpus (61): the All English corpus and the English Fiction corpus.

Compared to COHA (carefully curated to be genre-balanced across decades) and the English Fiction corpus, the All English corpus was found to contain an explosion of scientific writings since around 1930 (66); also see additional evidence in *SI Appendix*, Fig. S5). This shifting sampling paradigm may lead to results different from those found in our other two corpora. Thus, we provide the Google Ngram All English corpus for comparison in (*SI Appendix*, Fig. S2, *Bottom*). In the main text, we focus on COHA and the Google Fiction corpus, which are not subject to this known sampling bias and therefore may better reflect the language use of its time period. We describe further details of the three corpora in *SI Appendix*, section 6.

Quantifying Frequency Change. The increase of logged word frequency from year 1800[‖] to year 2000 were taken as the dependent variable:

$$\Delta Frequency = \log(Freq_{2000}) - \log(Freq_{1800})$$
$$= \log\left(\frac{Freq_{2000}}{Freq_{1800}}\right).$$

We are aware that historical word frequency may not always follow a monotonic change. If there is a large proportion of words with non-monotonic frequency change, representing change of word frequency as the difference between the two endpoints could be problematic. Using a principal component analysis (PCA), we confirm that for most words, frequency either remains stable or follows a monotonic trend over the 200 y. Only 5% of the variance is explained by non-monotonic frequency change (U-shaped or inverted U-shaped curve; see *SI Appendix*, section 7). Therefore, our analysis focusing on word frequency difference between the years 1800 and 2000 captures the dominant pattern of word frequency change.

General additive model. Age of acquisition, concreteness, arousal, valence, and word length were taken as independent variables. We controlled for log frequency at year 1800. All variables are scaled and centered. We are aware that word meanings are constantly in flux. Therefore, for those words that substantially changed their meaning, psycholinguistic norms such as valence, arousal, and concreteness may not apply because these norms were rated by people living today and therefore do not necessarily reflect word meanings from 200 y ago. For this reason, we quantified semantic shift following procedures described in Li and Siew (12) and only keep words with stable semantic meaning (ca.

[¶] In Kuperman et al. (56), participants were told to use the following definition when providing AoA: "We mean the age at which you would have understood that word if somebody had used it in front of you, EVEN IF YOU DID NOT use, read, or write it at the time."

[#] Valence and arousal are the two primary and independent dimensions of emotion that have been found universal across cultures (34). Valence taps into pleasantness that a word evokes, with higher-valenced words evoking pleasant emotions and lower-valenced words evoking unpleasant emotions. Arousal taps into feelings of being excited/calm with highly arousing words evoking excited, stimulated, frenzied, jittery, and wide-awake feelings and low arousing words evoking calm, sluggish, dull, and sleepy feelings.

[‖] For COHA, we take year 1820 because COHA starts at year 1810 and the 1810 data were much smaller in size than the 1820 data (words in 1820 are approximately 6 times the number of words in 1810).

92.5% of the vocabulary; see technical details in *SI Appendix*, section 8). We also performed the analysis without screening the semantically unstable words and found the same trend (*SI Appendix*, Fig. S3).

Preliminary analyses revealed non-linear relationships in the data. General additive models can accommodate nonlinear relationships by adding extra fixed effects to represent quadratic, cubic, or higher-order terms of a variable to represent complex relationships. To avoid overfitting problems, general additive models introduce a penalty on more complex curves. For each independent term in the model, two basic statistics are provided: how well the term predicts the dependent variable, represented by an F statistic and a P-value, and how nonlinear the relationship is, represented by the estimated degree of freedom (edf). Edf of 1 indicates a linear relationship, with higher edf for more complex curves.

Given that results using Google Ngram English Fiction corpus and COHA are similar to each other, we combined these two corpora in one general additive model, including corpus as a random effect. See *SI Appendix*, Fig. S2 for model fittings to each of the three corpora independently.

Author affiliations: [a]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; [b]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin 14195, Germany; [c]Department of Germanic Studies, Indiana University, Bloomington, IN 001809; [d]Program of Cognitive Science, Indiana University, Bloomington, IN 001809; [e]Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom; [f]Center for Life Span Psychology, Max Planck Institute for Human Development, Berlin 14195, Germany; [g]Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China; and [h]Department of Psychology, National University of Singapore, Singapore 119077, Singapore

1. M. Müller, Darwinism tested by the science of language. Translated from the German of professor August Schleicher. *Nature* **1**, 256–259 (1870).
2. T. C. Scott-Phillips, S. Kirby, Language evolution in the laboratory. *Trends Cognit. Sci.* **14**, 411–417 (2010).
3. M. H. Christiansen, N. Chater, Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–509 (2008).
4. C. Beckner et al., Language is a complex adaptive system. *Lang. Learn.* **59**, 1–26 (2009).
5. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (John Murray, London, 1871).
6. S. Kirby, Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* **5**, 102–110 (2001).
7. P. Monaghan, M. H. Christiansen, S. A. Fitneva, The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *J. Exp. Psychol.: General* **140**, 325–3470 (2011).
8. S. Kirby, H. Cornish, K. Smith, Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10681–10686 (2008).
9. P. Monaghan, Age of acquisition predicts rate of lexical evolution. *Cognition* **133**, 530–534 (2014).
10. T. T. Hills, J. S. Adelman, Recent evolution of learnability in American English from 1800 to 2000. *Cognition* **143**, 87–92 (2015).
11. T. T. Hills, J. S. Adelman, T. Noguchi, "Attention economies, information crowding, and language change" in *Big Data in Cognitive Science*, M. N. Jones, Ed. (Psychology Press, 2016), pp. 279–302.
12. Y. Li, C. S. Siew, Diachronic semantic change of language is constrained by how people use and learn language. *Mem. Cognit.* **50**, 1284–1298 (2022).
13. A. C. Haddon, *Language and Mind* (Walter Scott Ltd., 1895).
14. F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press, Cambridge, 1932).
15. J. Barrett, M. Nyhof, Spreading non-natural concepts: The role of intuitive conceptual structures in memory and transmission of cultural materials. *J. Cognit. Cult.* **1**, 69–100 (2001).
16. A. Bangerter, Transformation between scientific and social representations of conception: The method of serial reproduction. *Br. J. Soc. Psychol.* **39**, 521–535 (2000).
17. A. Mesoudi, A. Whiten, The hierarchical transformation of event knowledge in human cultural transmission. *J. Cognit. Cult.* **4**, 1–24 (2004).
18. Y. Kashima, Maintaining cultural stereotypes in the serial reproduction of narratives. *Person. Soc. Psychol. Bull.* **26**, 594–604 (2000).
19. T. T. Hills, The dark side of information proliferation. *Perspect. Psychol. Sci.* **14**, 323–330 (2019).
20. A. Mesoudi, A. Whiten, The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. B: Biol. Sci.* **363**, 3489–3501 (2008).
21. S. Garrod, L. Fay, N. Oberlander J. J., T. MacLeod, Foundations of representation: Where might graphical symbol systems come from? *Cognit. Sci.* **31**, 961–987 (2007).
22. P. G. Healey, N. Swoboda, I. Umata, J. King, Graphical language games: Interactional constraints on representational form. *Cognit. Sci.* **31**, 285–3097 (2007).
23. C. A. Theisen, J. Oberlander, S. Kirby, Systematicity and arbitrariness in novel communication systems. *Int. Stud.* **11**, 14–32 (2010).
24. B. J. Juhasz, Age-of-acquisition effects in word and picture identification. *Psychol. Bull.* **131**, 684–712 (2005).
25. C. Hodgson, A. W. Ellis, Last in, first to go: Age of acquisition and naming in the elderly. *Brain Lang.* **64**, 146–163 (1998).
26. V. Bradley, R. Davies, B. Parris, I. F. Su, B. S. Weekes, Age of acquisition effects on action naming in progressive fluent aphasia. *Brain Lang.* **99**, 128–129 (2006). Special Abstract Issue Academy of Aphasia 2006 Program - Academy of Aphasia.
27. S. J. Holmes, F. J. Fitch, A. W. Ellis, Age of acquisition affects object recognition and naming in patients with Alzheimer's disease. *J. Clin. Exp. Psychol.* **28**, 1010–1022 (2006).
28. C. T. James, The role of semantic information in lexical decisions. *J. Exp. Psychol.: Hum. Percep. Perf.* **1**, 130–136 (1957).
29. L. M. Miller, S. Roodenrys, The interaction of word frequency and concreteness in immediate serial recall. *Mem. Cognit.* **37**, 850–865 (2009).
30. C. Romani, S. Mcalpine, R. C. Martin, Concreteness effects in different tasks: Implications for models of short-term memory. *Quart. J. Exp. Psychol.* **61**, 292–323 (2008).
31. A. De Groot, R. Keijzer, 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. *Lang. Learn.* **50**, 1–56 (2000).
32. M. Sadoski, Resolving the effects of concreteness on interest, comprehension, and learning important ideas from text. *Educat. Psychol. Rev.* **13**, 263–281 (2001).
33. B. Snefjella, M. Généreux, V. Kuperman, Historical evolution of concrete and abstract language revisited. *Behav. Res. Methods* **51**, 1693–1705 (2019).
34. J. A. Russell, A circumplex model of affect. *J. Person. Soc. Psychol.* **39**, 1161–1178 (1980).
35. F. Pratto, O. P. John, Automatic vigilance: The attention-grabbing power of negative social information. *J. Person. Soc. Psychol.* **61**, 380–391 (1991).
36. V. Kuperman, Z. Estes, M. Brysbaert, A. B. Warriner, Emotion and language: Valence and arousal affect word recognition. *J. Exp. Psychol.: General* **143**, 1065–1081 (2014).
37. Z. Estes, J. S. Adelman, Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion* **8**, 441–444 (2008).
38. P. M. Pexman, M. J. Yap, Individual differences in semantic processing: Insights from the calgary semantic decision project. *J. Exp. Psychol.: Learn. Mem. Cognit.* **44**, 1091–1112 (2018).
39. M. Moffat, P. D. Siakaluk, D. M. Sidhu, P. M. Pexman, Situated conceptualization and semantic processing: Effects of emotional experience and context availability in semantic categorization and naming tasks. *Psychon. Bull. Rev.* **22**, 408–419 (2015).
40. J. S. Adelman, Z. Estes, Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition* **129**, 530–535 (2013).
41. E. A. Kensinger, S. Corkin, Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Mem. Cognit.* **31**, 1169–1180 (2003).
42. S. Hamann, Cognitive and neural mechanisms of emotional memory. *Trends Cognit. Sci.* **5**, 394–400 (2001).
43. M. Mather, Emotional arousal and memory binding: An object-based framework. *Perspect. Psychol. Sci.* **2**, 33–52 (2007).
44. E. A. Phelps, Emotion and cognition: Insights from studies of the human amydala. *Annu. Rev. Psychol.* **57**, 27–530 (2006).
45. M. Mather, M. Sutherland, Disentangling the effects of arousal and valence on memory for intrinsic details. *Emot. Rev.* **1**, 118–119 (2009).
46. S. Pinker, *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes* (Penguin, UK, 2011).
47. T. T. Hills, E. Proto, D. Sgroi, C. I. Seresinhe, Historical analysis of national subjective wellbeing using millions of digitized books. *Nat. Hum. Behav.* **3**, 1271–1275 (2019).
48. G. W. Allport, L. Postman, *The Psychology of Rumor* (Henry Holt, Oxford, UK, 2011).
49. P. Monaghan, S. G. Roberts, Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition* **186**, 147–158 (2019).
50. M. L. Lewis, M. C. Frank, The length of words reflects their conceptual complexity. *Cognition* **153**, 182–195 (2016).
51. J. S. Nairne, "Adaptive memory: Evolutionary constraints on remembering" in *Psychology of Learning and Motivation: Advances in Research and Theory, Psychology of Learning and Motivation*, B. Ross, Ed. (2010), vol. 53, pp. 1–32.
52. J. L. McGaugh, Memory-a century of consolidation. *Science* **287**, 248–251 (2000).
53. C. Darwin, *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life* (John Murray, London, 1859).
54. F. I. M. Craik, E. Tulving, Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol.: Gen.* **104**, 268–294 (1975).
55. A. Paivio, Mental imagery in associative learning and memory. *Psychol. Rev.* **76**, 241–263 (1969).
56. V. Kuperman, H. Stadthagen-Gonzalez, M. Brysbaert, Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* **44**, 978–990 (2012).
57. M. Brysbaert, A. Biemiller, Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behav. Res. Methods* **49**, 1520–1523 (2017).
58. A. B. Warriner, V. Kuperman, M. Brysbaert, Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **45**, 1191–1207 (2013).
59. M. Brysbaert, A. B. Warriner, V. Kuperman, Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **46**, 904–911 (2014).
60. G. Hollis, C. Westbury, L. Lefsrud, Extrapolating human judgments from skip-gram vector representations of word meaning. *Quart. J. Exp. Psychol.* **70**, 1603–1619 (2017).
61. J. B. Michel et al., Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
62. M. Davies, *The Corpus of Historical American English: COHA* (Brigham Young University, BYE, 2010).

63. BNC Consortium, "British national corpus version 3 (BNC XMLth ed.)" (Tech. Rep., Oxford University Computing Services, Oxford, U.K., 2007).

64. S. Bird, E. Klein, E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* (O'Reilly Media Inc., 2009).

65. F. Breithaupt, B. Li, J. K. Kruschke, Serial reproduction of narratives preserves emotional appraisals. *Cognit. Emot.* **36**, 581–601 (2022).

66. E. A. Pechenick, C. M. Danforth, P. S. Dodds, Characterizing the Google books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PloS One* **10**, e0137041 (2015).