The length of words reflects their conceptual complexity

Molly L. Lewis and Michael C. Frank

Psychology Department, Stanford University,
450 Serra Mall, Stanford, CA 94305, USA

*To whom correspondence should be addressed; E-mail: mll@stanford.edu.

Are the forms of words systematically related to their meaning? The arbitrariness of the sign has long been a foundational part of our understanding of human language.[1,2] Theories of communication predict another relationship between form and meaning, however: longer descriptions should convey more complex meanings.[3,4] Here we show that both the lexicons of human languages and their speakers encode the relationship between linguistic and cognitive complexity. Participants mapped longer words to more complex objects in comprehension and production tasks and across a range of stimuli. Explicit judgments of complexity were highly correlated with an implicit measure of study time in a memory task, suggesting that complexity is directly related to basic cognitive processes. In addition, judgements of complexity for a sample of real words correlated highly with their length across 80 languages, even controlling for frequency, familiarity, imageability, and concreteness. These results point to a general regularity in the design of lexicons and suggest the importance of cognitive constraints on language evolution.[5,6]

In a classic example of pragmatic reasoning,[3] the utterance "Lee got the car to stop" seems to imply an unusual state of affairs. Had the speaker wished to convey that Lee simply applied

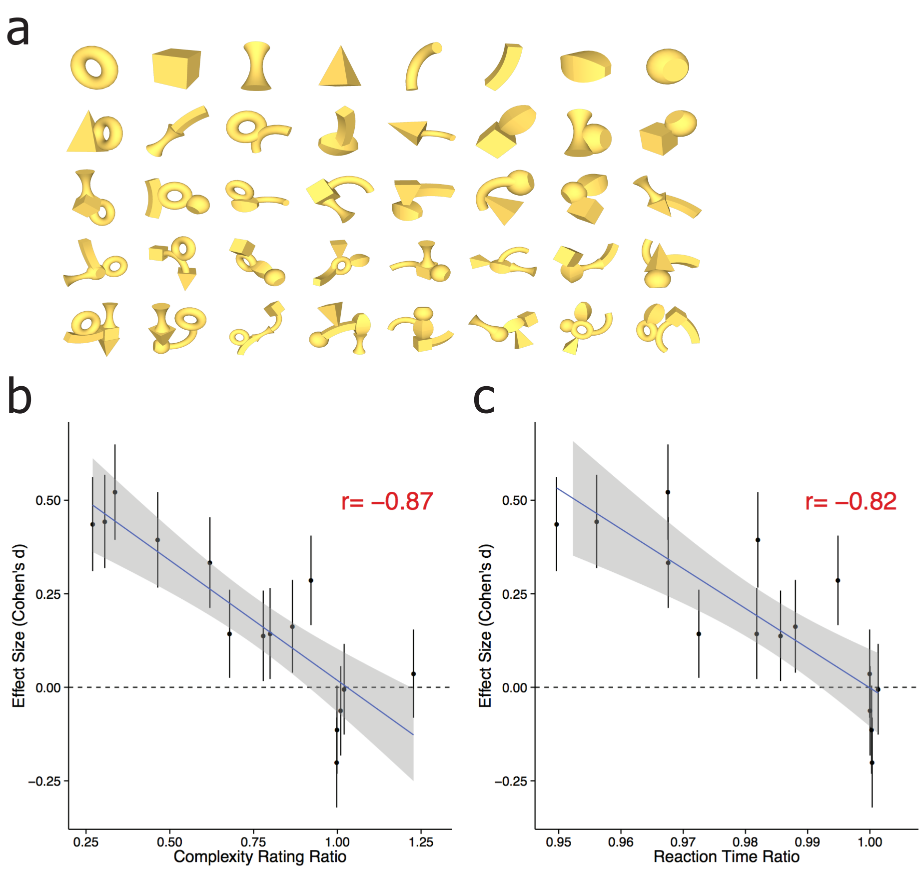

Figure 1

the brakes, the shorter and less exceptional “Lee stopped the car” would be a better description. The use of a longer utterance licenses the inference that there was some problem in stopping—perhaps the brakes failed—and that the situation is more complex. Does this same reasoning apply to the meanings of words? Is a “tupabugorn” more likely to be a complex, unusual object than a “bugorn”?

We tested this hypothesis by asking whether speakers would be biased to interpret a long novel word as being more likely to refer to a more complex novel referent. We presented participants on Amazon Mechanical Turk ($N = 750$) with a novel word of either 2 or 4 syllables and two possible alternative object referents. Possible referents were novel artificial objects whose complexity we manipulated by varying the number of parts the object contained (1 - 5 "geons";[7] Fig. 1a; these judgements were highly correlated with explicit complexity judgments, $r = .90$, $p < .0001$). For every unique combination of object complexities (1 vs. 2 geons, 1 vs. 3 geons, 1 vs. 4 geons, etc.), participants were asked to select which object the word named.

Across conditions, the more complex object was more likely to be judged the referent of the longer word. For each condition (e.g., 1 vs. 2 geons), we calculated the effect size for participants' complexity bias — the degree to which a choice of the more complex object was more likely for a longer, rather than a shorter, word. Effect size was highly correlated with the ratio of object complexities: The greater the mismatch in object complexity, the more the longer word was paired with the more complex object ($r = .87$, $p < .0001$; Fig. 1b). This bias also varied systematically with the length of the word: participant were more likely to select the more complex referent In a control experiment, we presented participants with words composed of 1,3, or 5 randomly concatenated syllable and two referents: a single geon object and a five geon objects ($N = 200$, $\beta = -.34$, $p < .0001$).

Next we asked whether this bias extended to more naturalistic objects. We gathered a sample of real objects without canonical labels (Fig. 2a) and asked participants to rate their complexity. These judgements were highly reliable across two independent samples (N = 60 in each, $r = .93$). We then divided the objects into quintiles based on these ratings, and used them as stimuli in a mapping task identical to the one used with the artificial objects. As with the artificial objects, effect size was negatively correlated with the complexity judgment ratio between the referent alternatives ($N = 1500$; $r = .70$, $p < .005$; Fig. 2b). Finally, we also found the same bias in language production: Participants produced novel coinages that were longer for the top quartile of objects compared to the bottom quartile ($N = 59$; $t(57) = 3.91$, $p < .001$).

If complexity is related to a basic cognitive process, we should be able to measure it using an implicit task, not just via explicit ratings. In visual cognition, stimuli that contain more

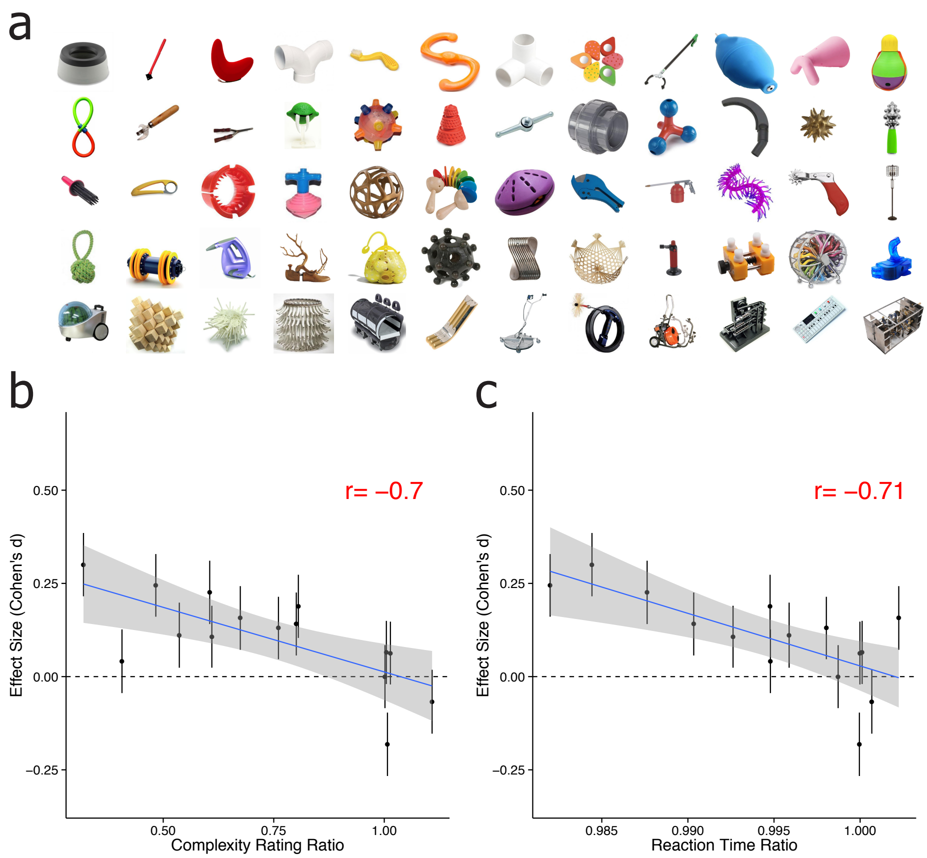

Figure 2

information require more processing time in search.[8] To test this prediction, we measured participants' reaction time to objects in a memory task. Each participant viewed half of the objects in the stimulus set, one at a time, and then made old/new judgments for the entire set. Critically, the training phase was self-paced, such that participants were allowed to study each object for as much time as they wanted.

Mean study time was highly correlated with explicit complexity norms for both artificial objects ($N = 250$; $r = .89$, $p < .0001$) and novel real objects ($N = 500$; $r = .54$, $p < .0001$). In addition, the ratio of study times for the two objects was correlated with the bias to choose a longer label for both the artificial objects ($r = .82$, $p < .001$; Fig. 1c) and the novel real

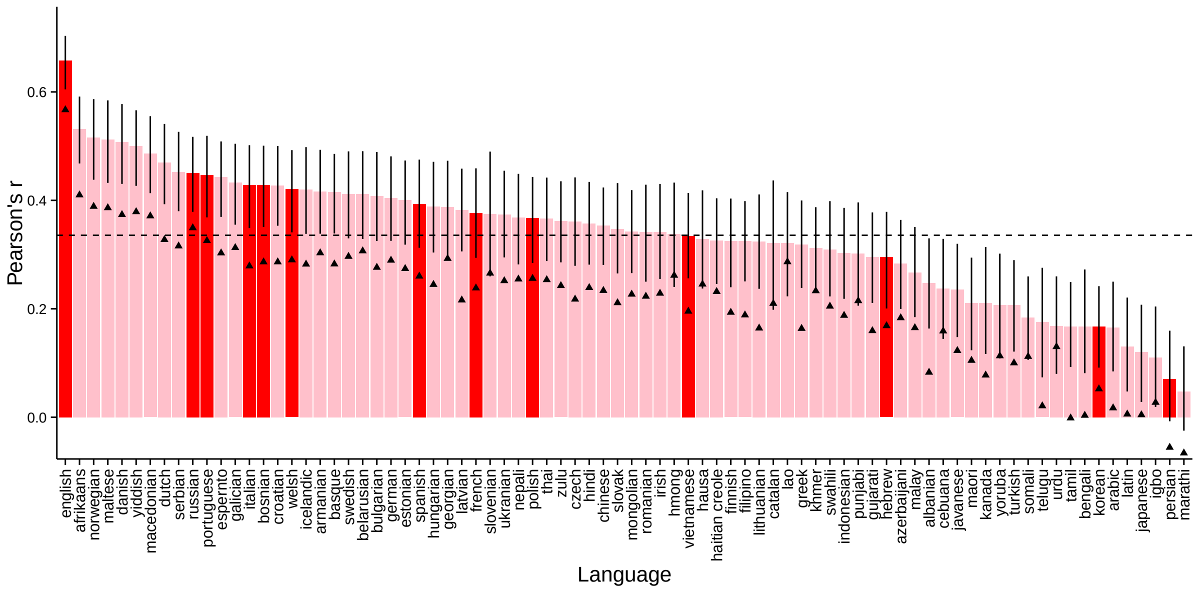

Figure 3

objects ($r = .71$, $p < .005$; Fig. 2c): Longer study times predicted longer labels. These findings suggest that label judgments are supported by basic cognitive processes related to the complexity or information content of a stimulus.

Together, these experiments point to a complexity bias in interpreting novel labels: Words that are longer tend to be associated with meanings that are more complex, as reflected in both explicit and implicit measures. Is this bias only relevant to judgments of unfamiliar words, or does it apply to familiar labels as well?

We collected ratings of the complexity of the meaning of 499 English words ($N = 250$). Longer English words had meanings that were rated as more complex in a similar explicit complexity rating procedure. Complexity judgements were positively correlated with word length, measured in syllables, phonemes, and morphemes ($r_{syllables} = .63$, $r_{phonemes} = .66$, $r_{morphemes} = .43$, all ps $< .0001$), even when closed-class words were excluded ($N = 453$; $r_{syllables} = .62$, $r_{phonemes} = .66$, $r_{morphemes} = .43$, all ps $< .0001$). Importantly, these relationships also remained reliable after controlling for the word's concreteness, imageability, and familiarity.

If the complexity bias relies on a universal cognitive process, it should generalize to lexicons beyond English. We explored this prediction in 79 additional languages, using Google Translate to translate our word set. Native speakers checked the accuracy of these translations for 12 of the 79 languages, finding an accuracy of .92 within this sample. For each language, we calculated the correlation between word length in terms of number of characters (to allow comparison between languages for which no phonetic dictionary was available) and mean complexity rating. All 79 languages showed a positive correlation between length and complexity ratings (Fig. 3). The grand mean correlation across languages was .34.

Word length is also strongly related to linguistic predictability, operationalized via simple frequency[9] or using a language model.[10] But the regularity we describe—a relationship between conceptual complexity and word length—holds even when controlling for frequency. In English, the correlation was only slightly reduced when controlling for log frequency ($r = .56$, $p < .0001$). Across languages, partialling out log frequency (estimated in English), the grand mean correlation was .22. In addition, in a number of experiments, when we manipulated the observed frequencies of novel objects, we found no effects on judgments of word length (see SI).

Languages also show phonological iconicity effects, such that semantic features[11,12] and even particular form classes[13] are marked by particular sound patterns. Our findings could be due to some broad iconic relationship between abstract measures of complexity and amount of verbal or orthographic effort—indeed this is precisely the relationship that Horn originally noticed. But specific iconic hypotheses that posit a parallel between an object's parts and the number of phonemes, morphemes, or syllables in its label do not account for the patterns in the English lexicon, which hold for monomorphemic words alone ($N = 387$; $r_{syllables} = .46$,

$r_{phonemes} = .53$, all ps $< .0001$) and for abstract ideas that have no obvious part structure ($r = .75$, $p < .0001$, for the 249 least concrete nouns).

What are the origins of structure in human language? Our findings reveal a broad systematicity in the design of languages — a bias for longer words to map to more complex meanings — and suggest that properties of the human cognitive system may underly this systematicity. We find that referents that contain more information, and consequently take longer for the cognitive system to process (as measured by study time), tend to be associated with longer labels. This broad systematicity both within and across languages fundamentally challenges the assumption that lexicons are arbitrary, and suggests that the structure of the lexicon reflects principles of communication.

Methods All participants in the experimental studies were recruited on Amazon Mechanical Turk and paid $0.15-0.30 for their participation, depending on the length of the task.

The methods were identical for the set of studies with the geons and the naturalistic objects, except for the stimuli used. In the norming task, we present participants with 10 objects from the the full stimulus set one at a time. For each object, we asked "How complicated is this object?," and participants responded using a slider scale, anchored at "simple" and "complicated."

In the study time task, participants were told they were going to view some objects and their memory of those exact objects would later be tested. In the study phase, participants were presented with half of the full stimulus set one at a time (20 geon objects and 30 naturalistic objects) and allowed to click a "next" button when they were done studying each object. After the training phase, we presented participants with each object in the full stimulus set (40 geon object and 60 naturalistic objects), and asked "Have you seen this object before?." Participants responded by clicking a "yes" or "no" button.

In the mapping task, we manipulated word length (2 vs. 4 syllables) and relative complexity of the referent alternatives within participants. There were 15 complexity conditions, corresponding to every possible combination of object quintiles (determined by the norming task described above). Each participant completed 4 short and 4 long trials in a random order, where each word was randomly associated with one of the complexity conditions. No participant saw the same complexity condition twice and no word or object was repeated across trials.

In the language control task for the naturalistic objects, participants completed a single trial in which they saw two possible referents, from the top and bottom quantiles of the complexity norms. Between participants, we manipulated the length of the word. The words were composed of 2, 4 or 6 syllables that were created by randomly concatenating CV syllables. Participants selected a referent by betting on the two alternatives, with the constraint that the bets must sum to 100.

In the label production task, participants were presented with a random subset of 10 of the 60 naturalistic objects and asked to generate a novel single-word label for the object.

In the studies involving real words, we selected 500 relatively high-frequency English words that were included in the MRC corpus. For each word, we asked "How complex is the meaning of this word?," and participants indicated their response on a 7-pt Likert scale, anchored at "simple" and "complex." Each participant rated 30 words.

References

1. F. Saussure, *London: Peter Owen* (1916).

2. C. Hockett, *Scientific American* **203**, 88 (1960).

3. L. Horn, *Meaning, form, and use in context* **42** (1984).

4. T. Jaeger, R. P. Levy, *Advances in Neural Information Processing Systems* (2006), pp. 849–856.

5. M. H. Christiansen, N. Chater, *Behavioral and Brain Sciences* **31**, 489 (2008).
6. E. Lieberman, J.-B. Michel, J. Jackson, T. Tang, M. A. Nowak, *Nature* **449**, 713 (2007).
7. I. Biederman, *Psychological Review* **94**, 115 (1987).
8. G. A. Alvarez, P. Cavanagh, *Psychological Science* **15**, 106 (2004).
9. G. Zipf, *The Psychobiology of Language* (Routledge, London, 1936).
10. S. Piantadosi, H. Tily, E. Gibson, *Proceedings of the National Academy of Sciences* **108**, 3526 (2011).
11. D. Maurer, T. Pathman, C. J. Mondloch, *Developmental Science* **9**, 316 (2006).
12. C. Westbury, *Brain and Language* **93**, 10 (2005).
13. T. A. Farmer, M. H. Christiansen, P. Monaghan, *Proceedings of the National Academy of Sciences* **103**, 12203 (2006).

Supplementary Information

Acknowledgements

Author Contributions

Author Information

Figure Legends

Figure 1. Artificial "geon" stimuli. Each row shows a different level of complexity, determined by the number of geon parts in the objects. (b, c) Experimental results from task in which participants were asked to map a novel word of varying length to one of two possible referents. Effect size between the long and short language conditions is plotted against the complexity ratio of the two referent alternatives. Fig. 1b shows the referents plotted in terms of explicit complexity judgements, and Fig. 1c shows the referents plotted in terms of reaction time. Error bars show 95% confidence intervals.

Figure 2. (a) Naturalistic, novel stimuli. Each row corresponds to a quintile determined by the explicit complexity judgements (top: least complex; bottom: most complex). (b, c) Experimental results from task in which participants were asked to map a novel word of varying length to one of two possible referents. Effect size between the long and short language conditions is plotted against the complexity ratio of the two referent alternatives. Fig. 2b shows the referents plotted in terms of explicit complexity judgements, and Fig. 2c shows the referents plotted in terms of reaction time. Error bars show 95% confidence intervals.

Figure 3. Correlations between conceptual complexity norms and word lengths, across languages. Dark red bars indicate languages for which translations were checked by native speakers; all other bars show translations obtained via Google Translate. Error bars show 95% confidence intervals obtained via non-parametric bootstrap. Triangles indicate correlation value partialling out English log frequency. The dashed line indicates the grand mean correlation across languages.