The length of words reflects their conceptual complexity

Supplementary Materials

TABLE OF CONTENTS

This document was created from an R Markdown file. The R Markdown file can be found here. All analyses and plots can be reproduced from the raw data with the code in this file. This document also contains links to the experimental tasks.

Methods

In Studies 1 and 5, we manipulated word length (2 vs. 4 syllables) and the relative complexity of the referent alternatives within participants. There were 15 complexity conditions, corresponding to every possible combination of object quintiles. In Study 1, the quintiles were determined by the number of geons in the object. In Study 5, the quintiles were determined by the norms obtained in Study 4. 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 Studies 2 and 4, we presented participants 12 objects from 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.” The first two objects were images of a ball and a motherboard to anchor participants on the scale.

In Studies 3 and 6, participants completed six forced-choice trials in which they saw two possible referents, from the top and bottom quintiles. The novel words were created by randomly concatenating 2, 4, or 6 consonant-vowel syllables. The last syllable of all words ended in a consonant. Each participant completed 2 trials for each word length.

In Study 7, participants were presented 10 objects from the set of objects normed in Study 4 and asked to generate a novel single-word label for the object. Five of the objects were from the bottom quantile of complexity norms, and 5 of the objects were from the top quantile of complexity norms. Order of objects was randomized.

In Studies 8 and 9, 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 study 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 Study 10, we selected 499 relatively high-frequency English words. 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." The first two words were always "ball" and "motherboard" to anchor participants on the scale. Each participant rated a sampleof 32 words.

In Study 12 (n = 477), we presented participants with 10 objects on a single screen. The objects were composed of a single geon. There were two types of objects. One object type appeared nine times and the second object type appeared once. After this training period, participants completed a forced choice mapping task, as in Studies 1 and 5. We presented a word that was either 2 or 4 syllables long and asked participants to make a judgment about whether the word referred to the low or high frequency object. Each participant completed a single mapping trial, and word length was manipulated between participants. There was no difference between the long and short word conditions ($\chi 2(1) = 0.02$, $p = .89$).

In Study 13 (n = 97), we manipulated object frequency by sequentially presenting objects. Participants saw 60 objects from the set of normed real objects one at a time. One object was presented 10 times and a second object was presented 40 times. Ten additional objects were included as fillers. After this training phase, participants completed a single mapping trial as in Study 12. Word length was manipulated between participants. There was no difference between the long and short word conditions ($\chi 2(1) = 0.01$, $p = .92$).

The Stanford University Review Board approved the study protocol for all experiments, and informed consent was obtained from participants prior to their participation. Sample sizes and exclusion criteria were pre-specified on the basis of pilot studies. Exclusion criteria are described in the Supplementary Information. The data meet the assumptions of the statistical tests applied, and all statistical tests were two-tailed.

Supplemental Text

All experimental studies (Studies 1-10 and 12-13) were completed on Amazon Mechanical Turk (AMT). AMT is an online crowdsourcing platform that provides a reliable subject pool for web-based studies (*17*). Participants were paid US$0.15-0.30 for their participation, depending on the length of the task.

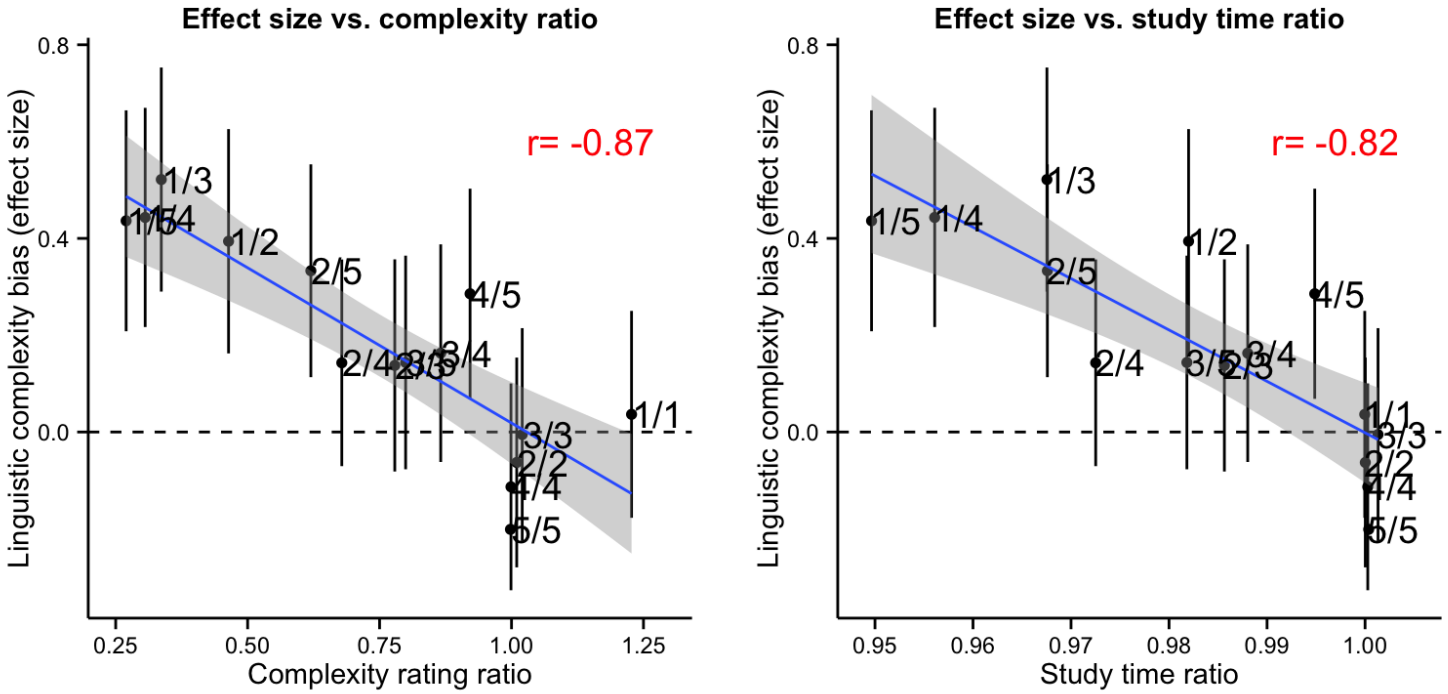
Study 1: Geon mapping task

The task can be found here.

The short word items were: “bugorn,” “ratum,” “lopus,” “wugnum,” “torun,” “gronan,” “ralex,” “vatrus.” The long word items were: “tupabugorn,” “gaburatum,” “fepolopus,” “pakuwugnum,” “mipatorun,” “kibagronan,” “tiburalex,” “binivatrus.”

Across all experiments, some participants completed more than one study. The results presented here include the data from all participants, but all reported results remain reliable when excluding participants who completed more than one study. Participants were counted as a repeat participant if they completed a study using the same stimuli (e.g., completed both Studies 1 and 2 with geons).

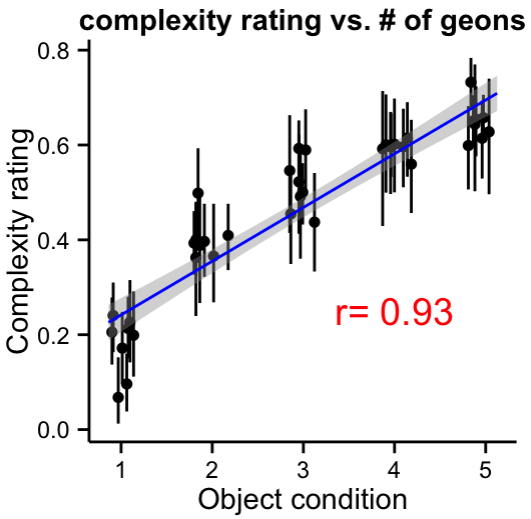
Plotted below is the effect size (bias to select complex alternative in long vs. short word condition) as a function of the complexity ratio between the two object alternatives. Each point corresponds to an object condition. Conditions are labeled by the quintiles of the two alternatives. For example, the “1/5” condition corresponds to the condition in which one alternative is from the first quintile and the other is from the fifth quintile. In the left plot, complexity is operationalized as the explicit complexity norms (Study 2). On the right, complexity is operationalized in terms of study times (Study 8). Effect sizes were calculated using the log odds ratio (*18*). In this and all subsequent plots, errors bars reflect 95% confidence intervals.



Study 2: Geon complexity norms

The task can be found here.

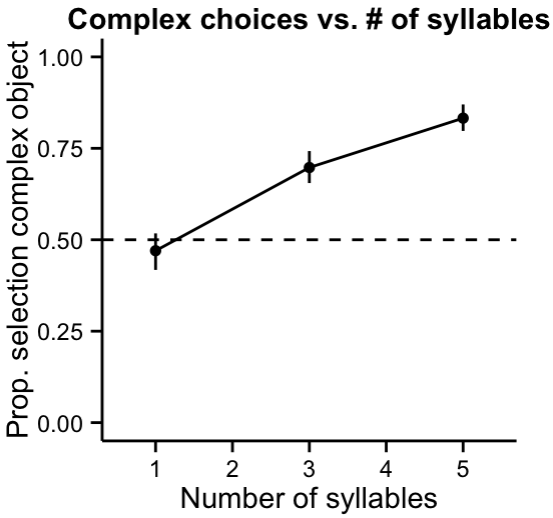
The relationship between number of geons and complexity rating is plotted below (*M* = .47, *SD* = .18). Each point corresponds to an object item (8 per condition). The x-coordinates have been jittered to avoid over-plotting. The confidence intervals are calculated via non-parametric bootstrapping.



Study 3: Geon mapping task control

The task can be found here.

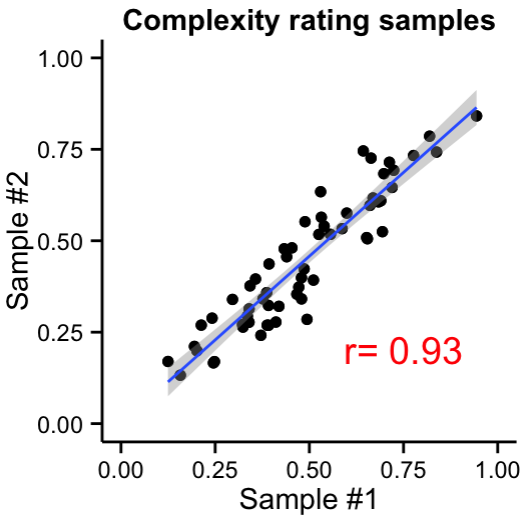
Plotted below is the proportion complex object selections as a function of the number of syllables in the target label. The dashed line reflects chance selection between the simple and complex alternatives.



Study 4: Real object complexity norms

The task can be found here.

Plotted below is the correlation between the two samples (n = 60 each, $M1$ = .49, $SD1$ = .18, $M2$ = .44, $SD2$ = .18) of complexity norms. Each point corresponds to an object (n = 60).

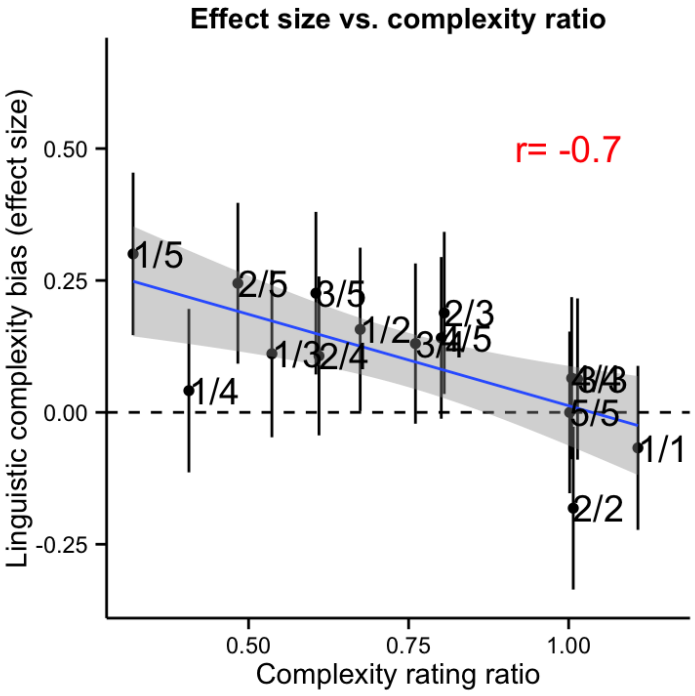


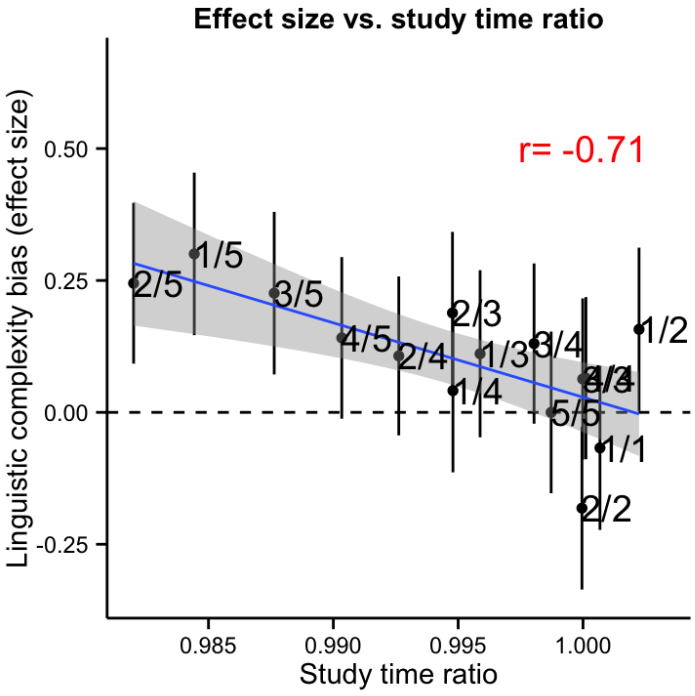
Study 5: Real object mapping task

The task can be found here.

The linguistic items were identical to Study 1.

Plotted below is the effect size (bias to select complex alternative in long vs. short word condition) as a function of the complexity ratio between the two object alternatives. Each point corresponds to an object condition. In the left plot, complexity is operationalized as the explicit complexity norms (Study 4). In the right plot, complexity is operationalized in terms of study times (Study 9).

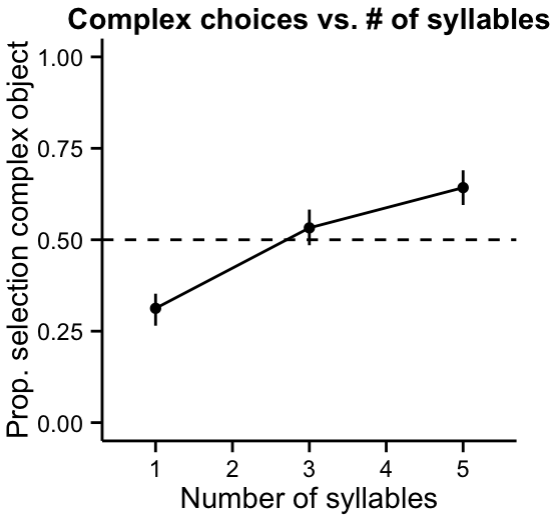




Study 6: Real object mapping task control

The task can be found here.

Plotted below is the proportion of complex object selections as a function of number of syllables. The dashed line reflects chance selection between the simple and complex alternatives.



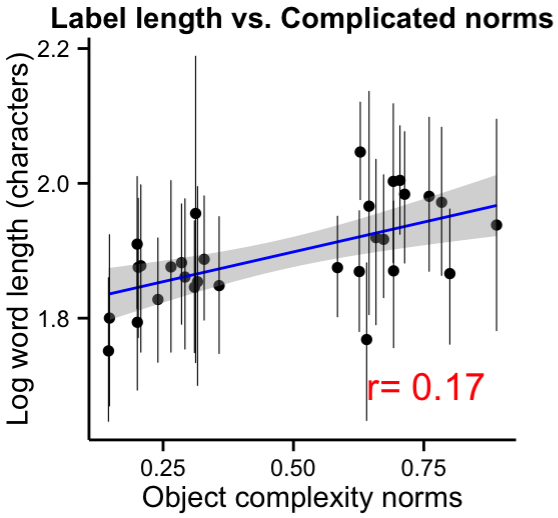
Study 7: Real object production task

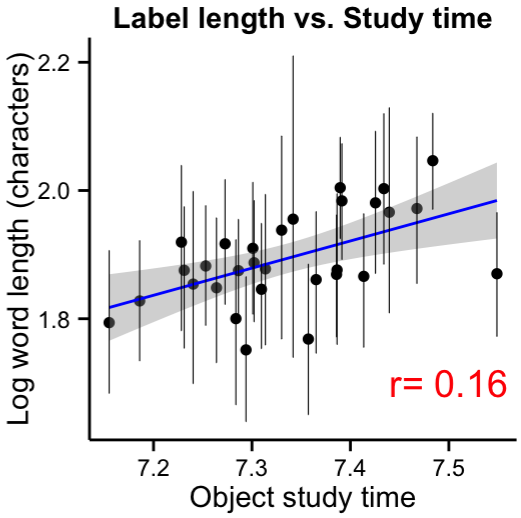
The task can be found here.

There were 26 productions (4%) that included more than one word. These productions were excluded.

For each object, we analyzed the log length of the production in characters as a function of the complexity norms (Study 4, left below). Length of production was correlated with the complexity norms: Longer labels were coined for objects that were rated as more complex (r=.17, p<.0001).

We also analyzed the log length of the production in characters (M = 1.89, SD = .26) as a function of study times (Study 9, right below). Length of production was correlated with study times: Longer labels were coined for objects that were studied longer (r = .16, p<.001).



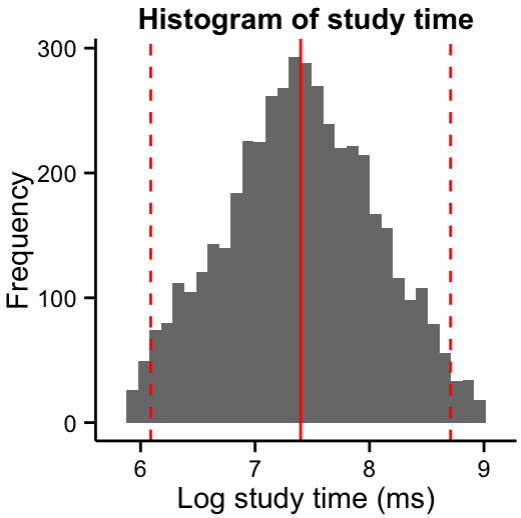


Study 8: Geon study time task

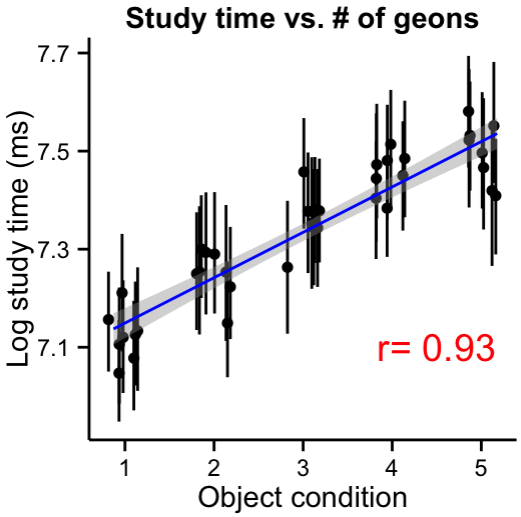
The task can be found here.

We excluded subjects who performed at or below chance on the memory task (20 or fewer correct out of 40). A response was counted as correct if it was a correct rejection or a hit. This excluded 9 subjects (4%). With these participants excluded, the mean correct was 72%.

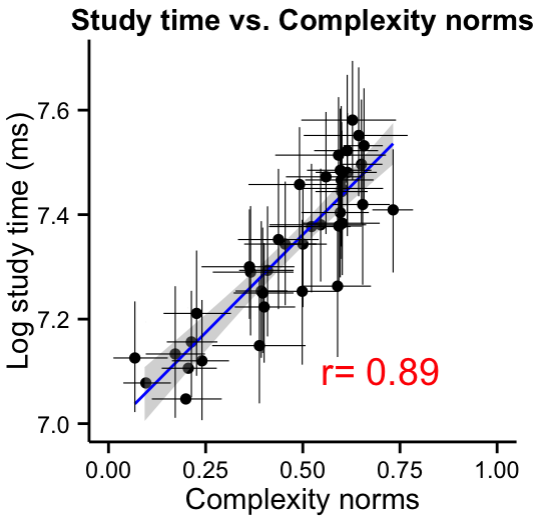
Participants were also excluded based on study times. We transformed the time into log space, and excluded responses that were 2 standard deviations above or below the mean. This excluded 4% of responses. Below is a histogram of study times after these exclusions (*M* = 7.40, *SD* = .66). The solid line indicates the mean, and the dashed lines indicate two standard deviations above and below the mean.



Like for the complexity norms, study times were highly correlated with the number of geons in each object (*r*=.93, *p*<.0001; see plot below, x-coordinates jittered to avoid over-plotting). Objects that contained more geons tended to be studied longer.



Study times were also highly correlated with complexity norms. Objects that were rated as more complex tended to be studied longer.



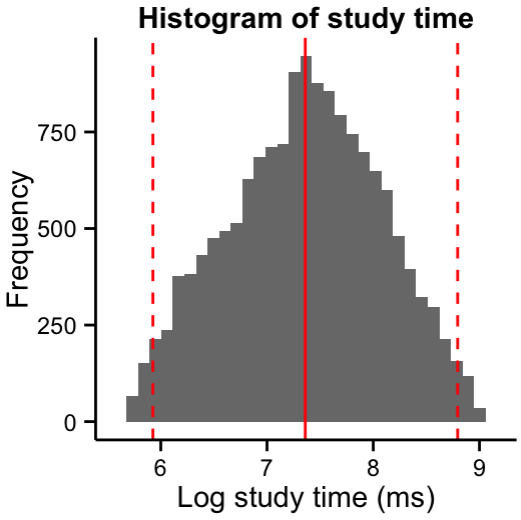
Study times did not predict memory performance. The study times for hits (correct "yes" responses; *M* = 7.33, *SD* = .52) did not differ from misses (correct "no" responses; *M* = 7.34, *SD* = .59; *t*(223) = .61, *p*=.54).

Study 9: Real object study time task

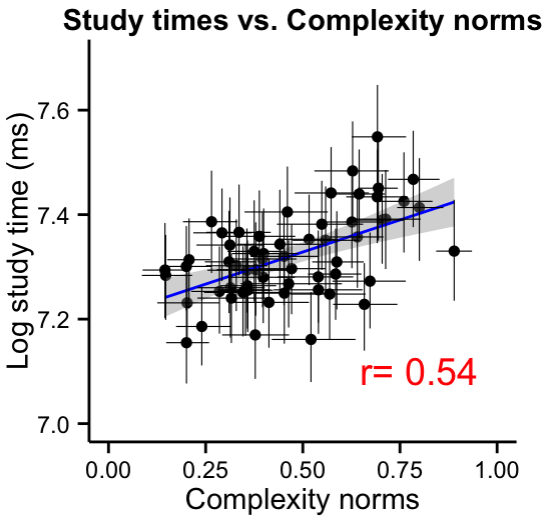
The task can be found here.

We excluded subjects who performed at or below chance on the memory task (30 or fewer correct out of 60). A response was counted as correct if it was a correct rejection or a hit. This excluded 6 subjects (1%). With these participants excluded, the mean correct was 84%.

Participants were also excluded based on study times. We transformed the time into log space, and excluded responses that were 2 standard deviations above or below the mean. This excluded 4% of responses. Below is a histogram of study times after these exclusions (*M* = 7.36, *SD* = .72). The solid line indicates the mean, and the dashed lines indicate two standard deviations above and below the mean.



The plot below shows the correlation between study times and explicit complexity norms for each object. Like for the geons, objects that were rated as more complex were studied longer.

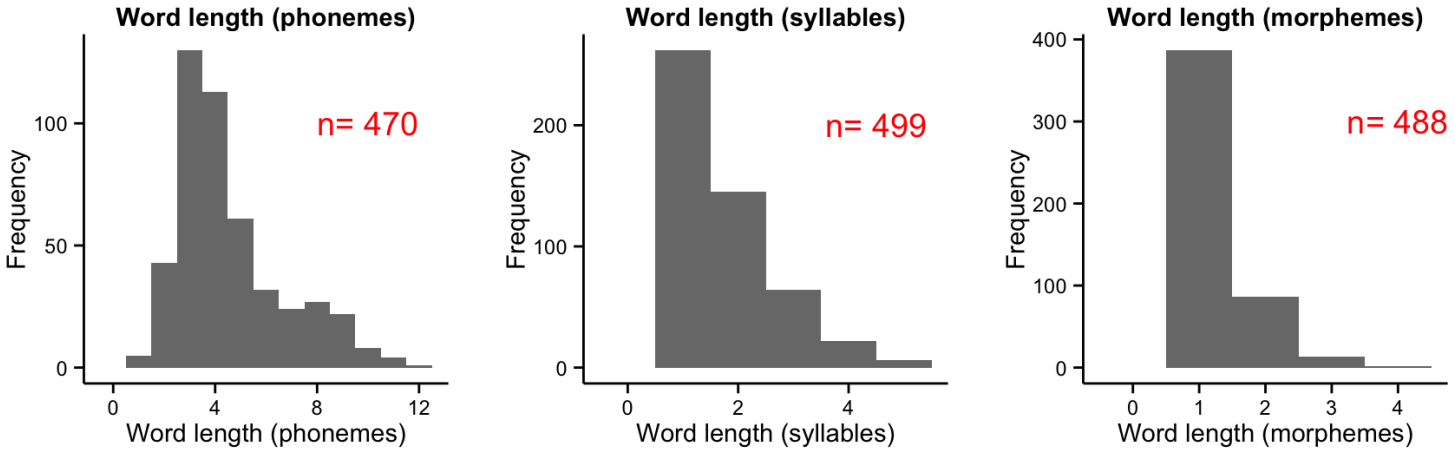


For the real objects, study times predicted memory performance. Study times for hits (correct "yes" responses; *M* = 7.24, *SD* = .60) were greater than for misses (correct "no" responses; *M* = 7.11, *SD* = .66; *t*(393) = 9.74, *p*<.0001).

Study 10: English complexity norms

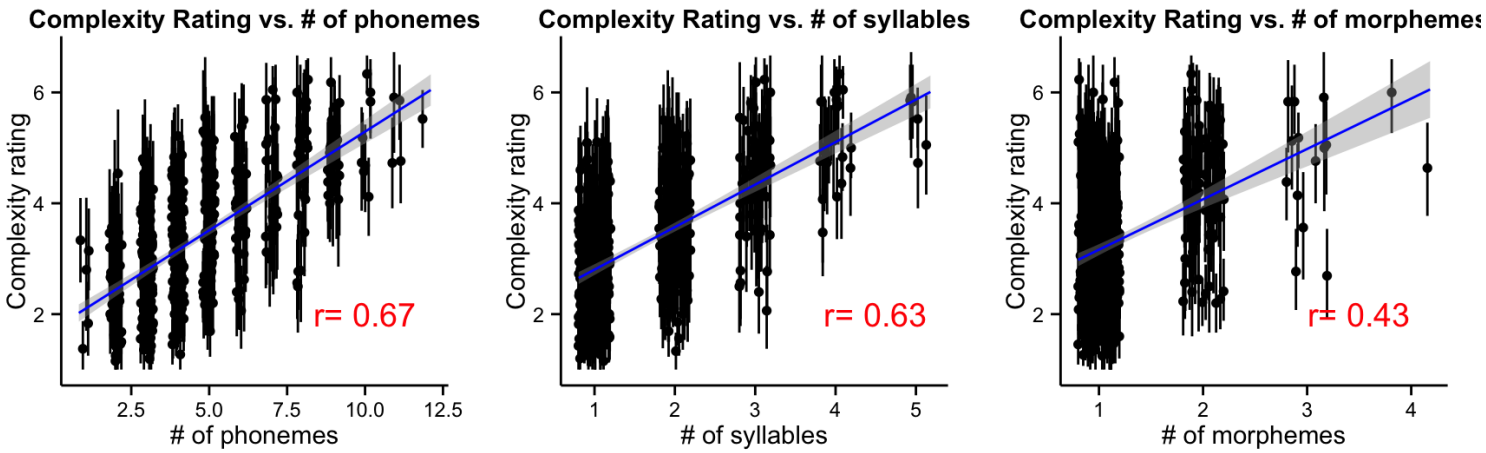
The task can be found here.

We selected 499 English words that were broadly distributed in their length. All of these words were included in the MRC Psycholinguistic Database (*19*). We considered three different metrics of word length: phonemes, syllables, and morphemes. Measures of phonemes and syllables were taken from the MRC corpus and measures of morphemes were taken from CELEX2 database (*16*). Below are histograms of the number of words as a function of each of the three length metrics. All three metrics were highly correlated with each other (phonemes and syllables: *r* = .89; phonemes and morphemes: *r* = .65; morphemes and syllables: *r* = .67). All three metrics were also highly correlated with number of characters, the length metric we use for the cross-linguistic analyses in Study 11 (phonemes: *r* = .92; morphemes: *r* = .69; syllables: *r* = .87).

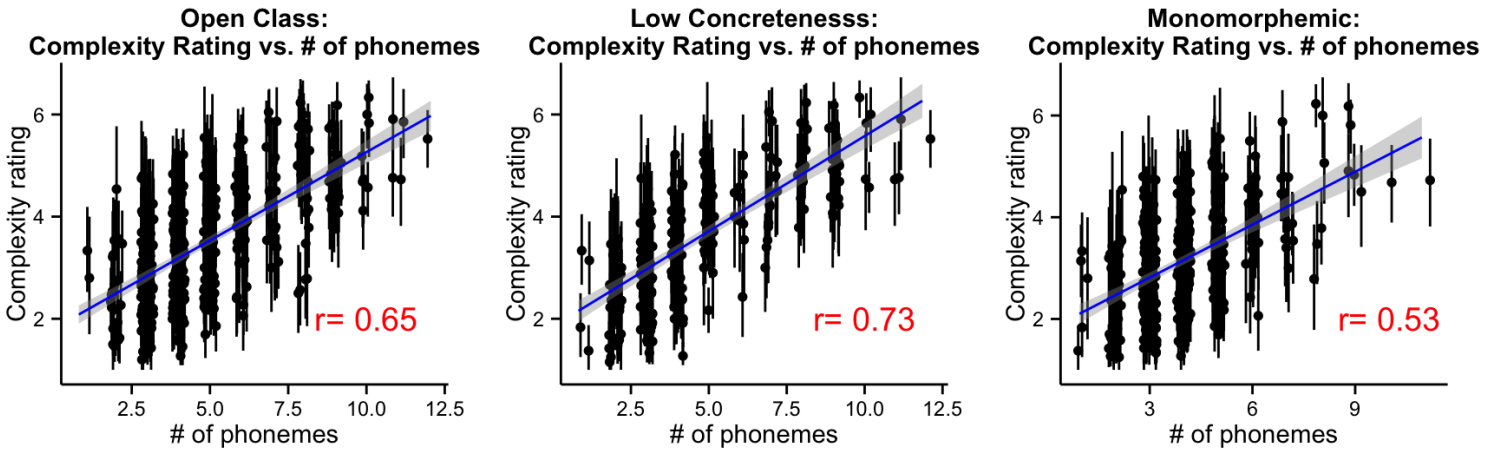


246 participants completed the rating task. We excluded participants who missed a simple math problem in the middle of the task that served as an attentional check. This excluded 6 participants (2%). Complexity ratings (*M* = 3.36, *SD* = 1.93) were highly correlated with length. Below we plot complexity as a function of each of the three length metrics. Each point corresponds to a word. The x-coordinates have been jittered to limit over-

plotting.



The relationship between length and complexity remained reliable for the subset of words that were open class, low in concreteness, and monomorphemic. The subset of low-concreteness words was determined by a median split based on the concreteness norms in the MRC corpus (*19*). Word class was coded by the authors. Plotted below are complexity ratings versus number of phonemes for closed class words (left), low concreteness words (center), and monomorphemic words (right).



Complexity and length are intuitively related to a number of other psycholinguistic variables. We estimated concreteness, familiarity and imageability from the MRC corpus (*19*), and word frequency from a corpus of transcripts of American English movies (Subtlex-us database; (*20*). All of these variables were reliably correlated with complexity (concreteness: *r* = -.27; familiarity: *r* = -.43; imageability: *r* = -.21; frequency: *r* = -.42, all *ps* <.0001). Length was also highly correlated with frequency (phonemes: *r* = -.53, *p* <.0001).

Nonetheless, the relationship between word length and complexity remained reliable controlling for all four of these factors. We created an additive linear model predicting word length in terms of phonemes with complexity, controlling for concreteness, imageability, familiarity, and frequency. Model parameters are presented below.

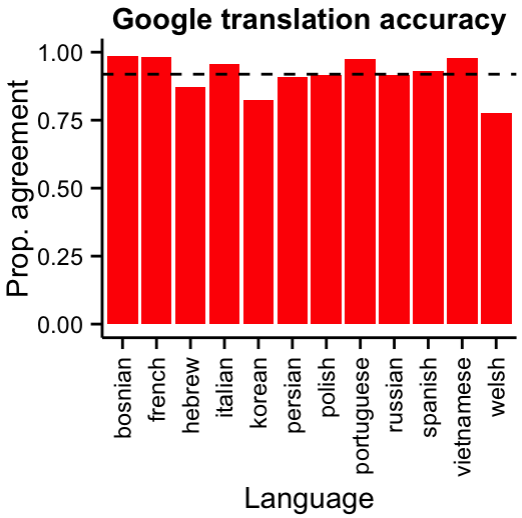
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.5020	0.2061	36.40	0.0000
complexity	0.2429	0.0116	20.86	0.0000
mrc.fam	0.0024	0.0005	4.80	0.0000
mrc.imag	-0.0003	0.0004	-0.81	0.4183
mrc.conc	-0.0033	0.0004	-9.16	0.0000
subt.log.freq	-1.1556	0.0332	-34.80	0.0000

This pattern held for the other two metrics of word length (morphemes and syllables).

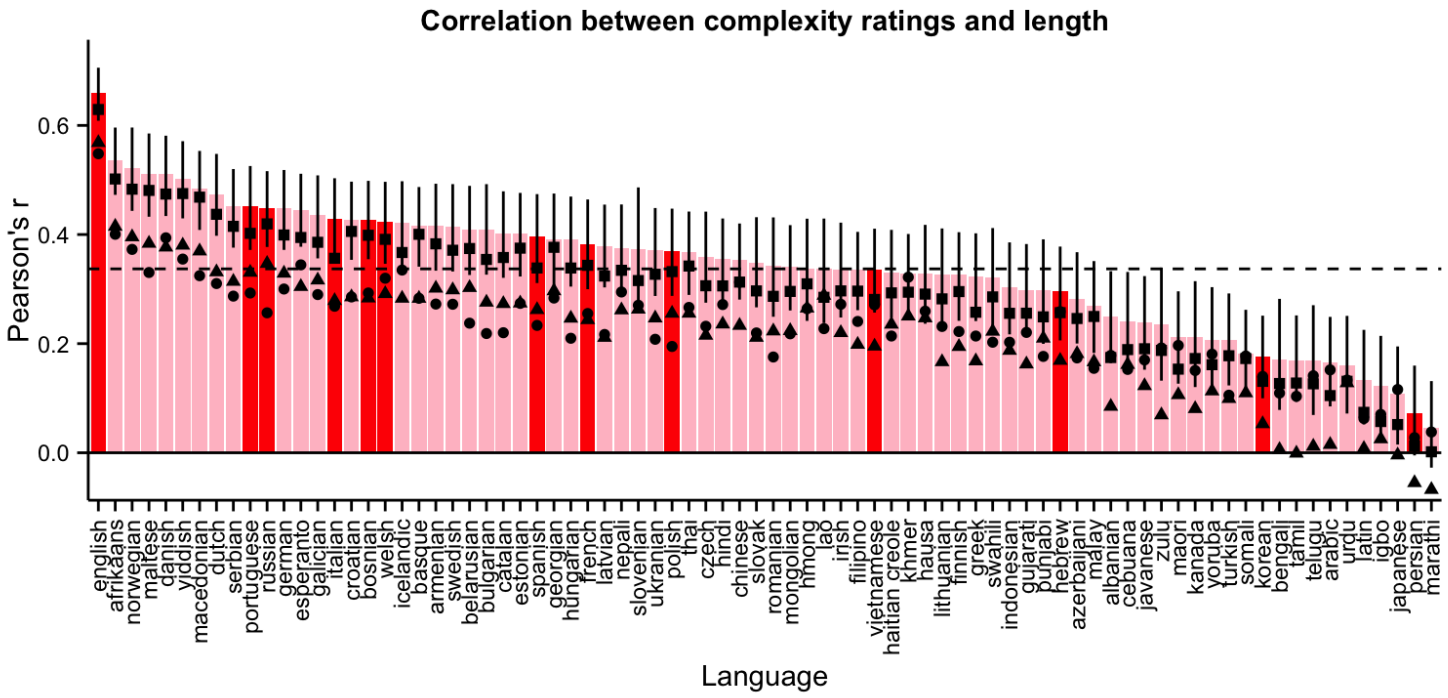
Study 11: Cross-linguistic analysis

We translated all 499 words from Study 10 into 79 languages using Google translate (retrieved March 2014). We translated the set of words into all languages available in Google translate. Words that were translated as English words were removed from the data set. We also removed words that were translated into a script that was different from the target language (e.g. an English word listed for Japanese).

Native speakers evaluated the accuracy of these translations for 12 of the 79 languages. Native speakers were told to look at the translations provided by Google, and in cases where the translation was bad or not given, provide a "better translation." Translations were not marked as inaccurate if the translation was missing. Plotted below is the proportion native speaker agreement with the Google translations across all 499 words. The dashed line indicates the mean (*M* = .92).



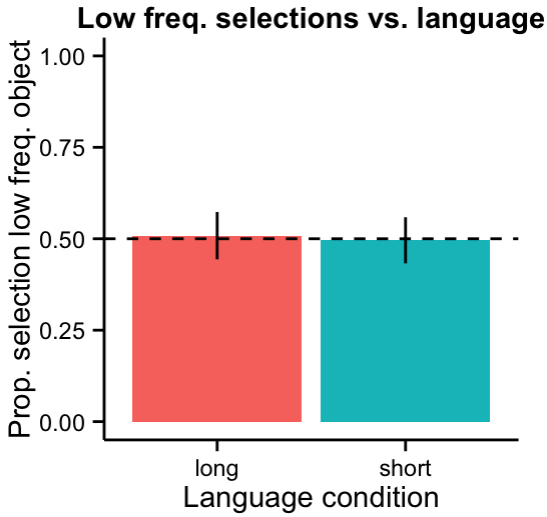
We counted the number of unicode characters for each translation. Variability in word length within languages was positively correlated with complexity ratings. Below the correlation coefficients are plotted for each language. Red bars indicate languages where the accuracy was checked by a native speaker and pink bars indicate unchecked languages. The dashed line indicates the grand mean correlation across languages. Triangles indicate the correlation between complexity and length, partialling out log spoken frequency in English. Circles indicate the correlation between complexity and length for the subset of wordskk that are monomorphemic in English. Squares indicate the correlation between complexity and length for the subset of open class words.



Study 12: Simultaneous frequency task

The task can be found here.

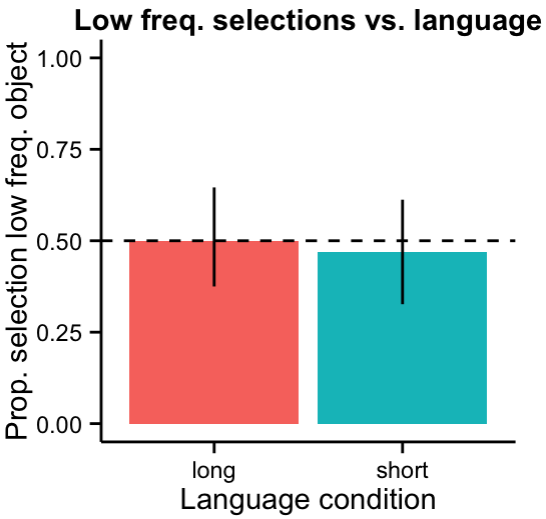
Plotted below is the proportion of low frequency object selections as a function of language condition (long vs. short). Selections between the two conditions did not differ.



Study 13: Sequential frequency task

The task can be found here.

Plotted below is the proportion of low frequency object selections as a function of language condition (long vs. short). Selections between the two conditions did not differ.


Low freq. selections vs. language
Prop. selection low freq. object
1.00
0.75
0.50
0.25
0.00
long
short
Language condition