Cognition 165 (2017) 45–52

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication

Jasmeen Kanwal *, Kenny Smith, Jennifer Culbertson, Simon Kirby

Centre for Language Evolution, University of Edinburgh, Edinburgh, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 25 October 2016
Revised 24 March 2017
Accepted 1 May 2017
Available online 8 May 2017

Keywords:
Zipf's Law of Abbreviation
Principle of Least Effort
Language universals
Efficient communication
Information theory
Artificial language learning

ABSTRACT

The linguist George Kingsley Zipf made a now classic observation about the relationship between a word's length and its frequency; the more frequent a word is, the shorter it tends to be. He claimed that this "Law of Abbreviation" is a universal structural property of language. The Law of Abbreviation has since been documented in a wide range of human languages, and extended to animal communication systems and even computer programming languages. Zipf hypothesised that this universal design feature arises as a result of individuals optimising form-meaning mappings under competing pressures to communicate accurately but also efficiently—his famous Principle of Least Effort. In this study, we use a miniature artificial language learning paradigm to provide direct experimental evidence for this explanatory hypothesis. We show that language users optimise form-meaning mappings only when pressures for accuracy and efficiency *both* operate during a communicative task, supporting Zipf's conjecture that the Principle of Least Effort can explain this universal feature of word length distributions.

1. Introduction

In 1935, the linguist George Kingsley Zipf pointed out what he claimed to be a universal property of human language: that "the magnitude of words stands in an inverse (not necessarily proportionate) relationship to the number of occurrences" (Zipf, 1935; p. 23). In other words, the more frequent a word is, the shorter it tends to be. This "Law of Abbreviation" has now been verified in a wide range of human languages, including: Chinese, Croatian, Czech, Dutch, English, French, German, Greek, Hungarian, Indonesian, Italian, Polish, Portuguese, Romanian, Russian, Slovenian, Slovak, Spanish, Sundanese, and Swedish (Ferrer-i-Cancho & Hernández-Fernández, 2013; Piantadosi, Tily, & Gibson, 2011; Sigurd, Eeg-Olofsson, & Van Weijer, 2004; Strauss, Grzybek, & Altmann, 2007; Teahan, Wen, McNab, & Witten, 2000). For example, one can clearly see this relationship for English words in Fig. 1. Interestingly, there is even evidence for its broader application in animal communication systems (in the vocalisations of common marmosets and formosan macaques, and in the surface behavioural patterns of dolphins; Ferrer-i Cancho et al., 2013) and in computer programming (e.g., use of the *alias* function in Unix to abbreviate frequent commands; Ellis & Hitchcock, 1986).

Zipf hypothesised that this universal pattern arises as a result of a tradeoff between two competing pressures: a pressure for accurate (successful) communication and a pressure for efficiency or less effort.[1] The idea is that together, these pressures would shape how forms are mapped to meanings, because languages have a finite inventory of discrete sounds that can be recombined to form words. This results in a lexicon with a limited number of words of a given length. Importantly, the shorter the length, the fewer distinct possible words there will be of that length, and the greater the potential confusability—shorter forms have less space for signal redundancy and thus are more likely to be confused in noisy signal transmission. Therefore, while a pressure for efficiency should favour these short words since they require less effort to produce (all things being equal), this is in direct conflict with the pressure for accurate communication. The latter should instead favour unique form-meaning mappings which minimise potential ambiguity—from this perspective, longer words have the clear advantage. How, then, can a language use the available short forms optimally, while still keeping ambiguity in check? The solution is to assign the shortest words to the most frequent meanings, leaving longer words for less frequent meanings, as in variable-length, e.g. Huffman, coding (Huffman,

* Corresponding author.
E-mail addresses: jasmeen.kanwal@ed.ac.uk (J. Kanwal), kenny.smith@ed.ac.uk (K. Smith), jennifer.culbertson@ed.ac.uk (J. Culbertson), simon@ling.ed.ac.uk (S. Kirby).

[1] The assumption that information is packaged into repeating words of variable length, and not fixed-length blocks—as in, e.g., block codes such as Hamming codes (Hamming, 1950)—is also necessary to make this prediction. Thanks to an anonymous reviewer for pointing this out.

http://dx.doi.org/10.1016/j.cognition.2017.05.001
0010-0277/

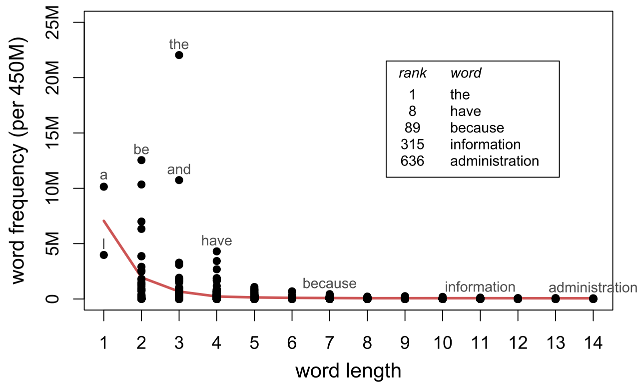

Fig. 1. The 1000 most frequent words in English. Each point represents an individual word (some points are labeled). The red line marks the mean frequency for the words of each length (here, orthographic length is used, but the same overall pattern would be seen if phonetic length were used instead.) The more frequent a word is, the shorter it tends to be. According to Zipf's Law of Abbreviation, this is a universal pattern of human languages. Frequency counts used here are from the 450 million word COCA corpus (Davies, 2008). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1952). Zipf called this hypothesised tendency to produce short utterances wherever possible the "Principle of Least Effort".

The Principle of Least Effort offers a functional explanation for the Law of Abbreviation, if we imagine it playing out through incremental changes over time. If language users track frequency differences between meanings (consciously or otherwise), then processes of change may differentially affect words whose frequencies differ. For example, if a word is more frequently used, then it may be more likely to be targeted for reduction or shortening (e.g., 'information' becomes 'info'). Form-meaning mappings would then gradually shift toward more optimal alignment of frequency with length (Zipf, 1935).

While this is an attractive explanatory account, several researchers have raised the possibility that the inverse relationship between word length and word frequency could emerge instead from simple constraints on randomly generated systems. For example, a lexicon generated through a random typing process, in which 'words' are produced by pressing keys (including the space bar) at random, has properties that are consistent with the Law of Abbreviation (Ferrer-i-Cancho & Moscoso del Prado, 2012; Moscoso del Prado, 2013). While we know that languages are not actually generated at random in this way, it nevertheless remains a possibility that the Law of Abbreviation could result from some yet-unidentified statistical process, unrelated to optimisation behaviour on the part of language users.

Several studies provide indirect evidence connecting competing pressures for accurate and efficient communication to properties of linguistic systems introduced by language users. For example, previous research has shown that learners restructure case marking systems such that case markers are preferentially used when grammatical roles are ambiguous and omitted when other disambiguating information is present (Fedzechkina, Jaeger, & Newport, 2012). This is consistent with the idea that effort (here, producing case markers) is reduced in a way which preserves communicative function. Language learners have also been shown to capitalise on differences in the length of novel labels to make pragmatic inferences about the communicative intentions of speakers (Degen, Franke, & Jäger, 2013). A computational model of iterated learning (Kirby, 2001) shows that short, non-compositional morphological forms are more likely to evolve for frequent meanings, while longer, compositional 'regular' forms are more likely to persist for infrequent meanings, due to a tradeoff between the pressure for learnability and the pressure for producing shorter, more replicable forms.

A direct link between frequency and utterance-length shortening in actual language users has been shown in studies such as Krauss and Weinheimer (1964) and Clark and Wilkes-Gibbs (1986). In these studies, participants played a dyadic communication game, where 'directors' used English to describe objects for their partners ('matchers') to identify from a set. The objects being communicated about were abstract geometrical shapes lacking canonical English names. The director would typically begin by using a long, elaborate phrase to help the matcher identify the correct object. However, on repeat occurrences of the object, the director would take advantage of a growing base of shared knowledge, established through communication, to gradually shorten the descriptive phrase and thereby reduce the effort expended. For example, an object described as "upsidedown martini glass in a wire stand" on its first occurrence ultimately became shortened to just "martini" after several repeat occurrences. The more times an object reoccurred, the shorter its average length by the end of the experiment. These results depended on the director receiving positive, real-time feedback from the matcher during the signalling game (Hupet & Chantraine, 1992; Krauss & Weinheimer, 1966), suggesting that it is a communicative context which triggers the drive to reduce effort. Thus, this result suggests one mechanism by which the Law of Abbreviation could arise: if the form associated with a meaning becomes shorter the more times it occurs in conversation, and these mappings are retained and spread across speakers, then in the lexicon overall, more frequent meanings will end up with shorter forms than less frequent meanings.

However, as we mentioned above, there is competition for the short forms in a lexicon. For example 'info' refers to 'information', and not 'informality', 'infoliation', or 'infoedation'. Why is this? In the Krauss and Weinheimer (1964) and Clark and Wilkes-Gibbs (1986) studies, participants were communicating about a small set of meanings using a large space of possible utterances. All labels could thus be shortened in this task without resulting in ambiguity. However, when several meanings are in direct competition for a single short label—a problem that arises at the level of an entire lexicon—the mechanism shown in these studies is not sufficient to account for why one meaning gets mapped to the short form and not the others.[2]

Thus, while these previous studies are consistent with the idea that something like the Principle of Least Effort operates during language use, they do not explicitly target the hypothesised role of competing communicative pressures—the pressure for reduced effort versus the pressure against ambiguous form-meaning mappings—in modulating word length within the lexicon. In our study, we make use of a miniature artificial language learning paradigm to create a setting in which these two pressures are directly in conflict: a reduction in effort cannot be achieved without also increasing the ambiguity of form-meaning mappings. Crucially, our set-up allows us to isolate these different pressures in order to determine their individual contribution to the overall behaviour of a miniature artificial lexicon. Following Zipf, we hypothesise that only when these pressures are both present—and thus in direct con-

[2] Interestingly, not all possible short forms in a language actually get used. This could be a consequence of noisy communication—using short forms sparingly would further minimise potential confusability. However, it has been found that frequent (and by proximity short) forms tend to be tightly clustered together in the phonological space, in seeming opposition to this end (Dautriche, Mahowald, Gibson, Christophe, & Piantadosi, 2017). This may be due to the influence of constraints on learning, memory, and production, which favour lexicons with high phonetic regularity. Thus, even though not all possible short forms are used, there will be particularly tough competition for those forms that fall within the more densely-populated regions of the phonological space. Thanks to an anonymous reviewer for raising this topic.

flict—will language users restructure their input to align shorter forms with more frequent meanings. In this way, our study aims to provide a concrete link between optimisation behaviour at the level of the individual and the global pattern Zipf first observed.

2. Miniature artificial language learning experiments

We use a miniature artificial language learning paradigm, which has previously been used to shed light on the cognitive mechanisms and environmental pressures that shape language structure (e.g., Culbertson, Smolensky, & Legendre, 2012; Fedzechkina et al., 2012; Kirby, Cornish, & Smith, 2008). In this paradigm, participants learn a miniature artificial language, and then we observe how they reshape their input as they use the language, in this case to communicate with a partner (see also Fehér, Wonnacott, & Smith, 2016; Kirby, Tamariz, Cornish, & Smith, 2015; Winters, Kirby, & Smith, 2015).

2.1. Participants

124 participants (51 females, 64 males; a further 9 chose not to report their gender) were recruited through Amazon Mechanical Turk. 106 of these reported themselves as native English speakers, of which 88 were monolingual. A broad range of other languages were represented across the remaining participants. Ages ranged from 18 to 73 (mean = 33).

2.2. Materials

Participants were trained on two names for each of two plant-like alien objects, by repeatedly being shown pictures of the objects labeled with their names on a computer screen (see also Reali & Griffiths, 2009; Vouloumanos, 2008). Crucially, one of the two objects appeared three times more frequently than the other—specifically, one object appeared 24 times and the other 8, for a total of 32 training trials.

Each object appeared half the time labeled with its long name, a 7-letter word, and half the time with its short name, a 3-letter word derived by clipping the last two syllables off the long name. The process of clipping, or word-truncation, is a common word-shortening device in many languages (e.g. *info* for *information* in both English and French; Antoine, 2000). In natural languages, shorter words are subject to greater confusability for a number of reasons. They have less space for signal redundancy and are therefore more likely to be misinterpreted or lost in noisy transmission. There are also more unique possible 7-letter strings than 3-letter strings, and thus word shortening can often result in outright ambiguity. Indeed, shorter words are more likely to be polysemous and homophonous (Piantadosi, Tily, & Gibson, 2012). To model these phenomena in our miniature lexicon, we designed the names such that the short name for both objects was identical (*zop*), while the long names were unique (*zopekil* and *zopudon*). A schematic diagram of the object frequencies and labels is provided in Fig. 2a.

Which object (the blue fruit or the red stalk) was more frequent, as well as which object was paired with each label, were both counterbalanced between participants, giving a total of 4 possible object-frequency-label pairings which a participant might be trained on. This ensured that potential factors such as sound symbolism, or higher saliency of one of the objects, could not systematically bias our results.

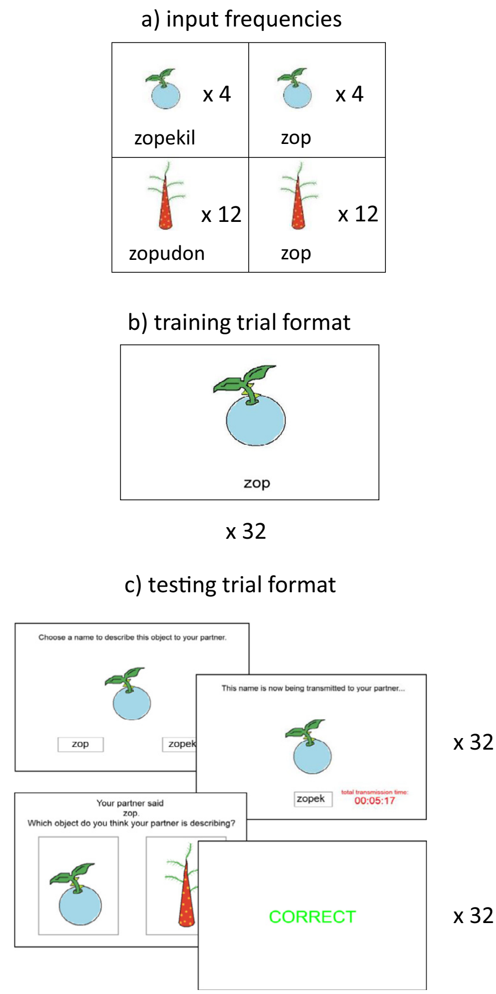

Fig. 2. (a) A schematic diagram of the frequencies of the objects and labels presented during the training trials in all four experimental conditions. One object appeared three times more frequently than the other. Each object was labeled half the time with its unique long name, and half the time with its ambiguous short name, which was a clipped version of the long name. (b) An example training trial. (c) An example of a director trial in the Combined condition (top) and a matcher trial followed by feedback (bottom).

2.3. Procedure

Participants were assigned to one of four conditions, where we manipulated the presence of pressures to communicate accurately and quickly in a between-subjects 2 × 2 design. In all conditions, the experiment consisted of two phases: training and testing. The training phase was identical for all four conditions, but the testing phase differed across conditions.

2.3.1. Training phase

On each training trial, an object was presented on screen alone for 700 ms. The appropriate label then appeared beneath the object for a further 2000 ms, yielding a total trial duration of 2700 ms. A blank screen showed for 500 ms between each trial. The 32 training trials were presented in a different randomised order for each participant.

2.3.2. Testing phase

After the training phase, testing procedures varied depending on the experimental condition. In the Combined condition, participants were under a pressure to communicate accurately *and* to communicate efficiently, as according to Zipf's hypothesis, both of these competing pressures must be present for the Law of Abbreviation to emerge. The remaining three conditions removed one or both of these accuracy and time pressures. In all conditions, the testing trials contained the same frequency ratio over objects as the training trials: the frequent object appeared three times more frequently than the infrequent object.

2.3.2.1. Condition 1: Combined. In the testing phase of this condition (henceforth referred to as the Combined condition), participants were paired with a partner to play a communication game. This was done by putting participants in a virtual queue, managed by a central server script, after completing the training trials. Participants were paired sequentially as they finished training; once a participant entered the queue, the server would pair them with the next participant to finish training after them. To encourage participants to wait as long as possible in the queue without leaving the game, they were shown a humorous cat video while they waited. However, if participants had still not been paired with a partner after 5 min, they were removed from the queue and paid for their time. This method allowed us to successfully run a dyadic artificial language communication experiment online using a crowdsourcing platform. We were therefore able to relatively quickly and easily collect data from a more culturally and linguistically diverse group of participants than is usually possible with traditional lab-based experiments that draw mainly from a university's undergraduate population.

Once paired with a partner, participants began the communication game. On each trial, the 'director' was shown an object on the screen and told to transmit its name to the 'matcher'. The director always had two options for which name to send: the long name for the object or the (ambiguous) short name. The director chose a name by clicking on it, and was then given instructions for how to actually transmit the name to the matcher. This was done by pressing and holding the mouse in a central transmission box in which each letter in the name appeared one by one, at 1200 ms intervals. Note that participants never had to type the names or necessarily remember their correct spelling; once they chose a name from the two options on the screen, the letters would appear sequentially in the transmission box as they held down the mouse. Only once all the letters had appeared in the box was the name transmitted to the matcher. If the mouse was released before all letters had been transmitted, the participant would have to start again from the first letter (but the total transmission time was only counted for the successful transmission). This belaboured method of transmission, in which the long name was significantly slower to transmit than the short name, introduced an element of effort into communication, modelling the difference in effort in spoken communication associated with producing long versus short utterances.

Once the matcher received the name from the director, the matcher was asked to choose which of the two objects they thought the director was referring to. Both players were then given feedback as to whether the matcher chose the correct object.

The players alternated roles after every trial, with the matcher becoming the director and the director becoming the matcher, until both completed 32 director trials and 32 matcher trials. The frequency with which each object appeared in each player's director trials matched those of the training frequencies: 24 occurrences of the frequent object, and 8 of the infrequent object. The order of these 32 director trials was randomly shuffled for each participant. The member of the pair who entered the queue first was the first player to direct.

To model the pressures in spoken communication to be both efficient and accurate, pairs were told at the beginning of the game that they would be rewarded a bonus payment if they were the pair to complete the game in the quickest time with the highest number of correct match trials. Time was only counted during name transmission, and the time count was displayed next to the transmission box as the participant was transmitting a name, to underline the time pressure. Example screenshots of a director trial and matcher trial are shown in Fig. 2c.

In order to tease apart the influence of the two pressures on the participants' patterns of behaviour, we included three further experimental conditions, described below, for a full 2×2 manipulation of the pressures for accuracy and efficiency.

2.3.2.2. Condition 2: Accuracy. In this condition, participants were paired to play a communication game as described above, but in the director trials, there was no intermediate step between the director choosing a name to send and the matcher receiving the name; the names were sent instantaneously, thus removing any difference in effort between transmitting long or short names. Pairs were told that the goal of the game was to have their partner make as many correct guesses as possible. There was no bonus reward given for the most accurate pair, as the task was extremely easy and we predicted that most pairs would achieve maximum accuracy, which turned out to be the case.

2.3.2.3. Condition 3: Time. In this condition, communication was taken out of the game entirely; participants played a one-player game consisting of 64 director trials only. In each director trial, participants were told to choose a name to describe the object shown on the screen, but there was no subsequent communicative task. As in the previous conditions, the choice was always between the long name and the short name. Once chosen, the name had to be entered as in the Combined condition, by pressing and holding the mouse in a transmission box, with each letter appearing at 1200 ms intervals. The next trial began only when all the letters had appeared in the box. Thus, the long name was significantly slower to produce than the short name. The transmission process was also timed with an on-screen timer as in the Combined condition, and participants were told at the beginning of the game that they would be rewarded a bonus payment if they were the player with the shortest overall transmission time.

2.3.2.4. Condition 4: Neither. The fourth and last condition contained neither a pressure for efficiency nor a pressure for accuracy. As in the Time condition, participants played a one-player game with no explicit communicative element, but additionally there was no time difference associated with transmission; once a label was chosen to describe an object, long or short, it was instantaneously recorded and the player was advanced to the next trial. We included this condition in order to provide a baseline for participants' behaviour from which to assess the effects of the accuracy and time pressures in the other three conditions.

2.3.3. Payment

Participants were paid depending on the condition they were in, commensurate with the average time it took to complete that condition. Participants in the Combined condition, the longest to complete due to both the slow transmission process and having to wait for the partner's response after each trial, were paid $2; participants in the Accuracy and Time conditions were paid $1, and participants in the Neither condition, the shortest to complete, were paid $0.50.

2.4. Predictions

Our predictions for the Neither condition were that participants would either probability-match—i.e. use the long and short forms for both objects with equal frequency, as in the training trials (see Hudson Kam & Newport, 2005)—or their behaviour would reveal prior biases language users bring to the task, such as a preference against using ambiguous forms.

In the Accuracy condition, we predicted that participants would be more likely to use the long names for both objects compared to the baseline condition, given the potential loss of accuracy from using the ambiguous short name, and with no time considerations to favour the use of short but ambiguous labels. Given the task demands, this would therefore be the best strategy to use in this condition.

In contrast, in the Time condition, we predicted that participants would use the short name for both objects: with no communicative purpose attached to the transmissions, and an incentive to be as quick as possible, using the short name in every trial is the best strategy in this condition.

In the critical Combined condition, with both a time and an accuracy pressure, we predicted that participants would converge on the optimal strategy, in which the frequent object is consistently mapped to the ambiguous short name, and the infrequent object to its unique long name, in line with Zipf's Law of Abbreviation. Using this strategy, transmission time is minimised as much as possible while still maintaining one-to-one form-meaning mappings, thereby also ensuring accurate communication.

3. Results

Fig. 3 shows the proportion of trials on which the short (ambiguous) label was selected by the director, for high- and low-frequency objects. As predicted, in the Accuracy condition, most participants retained the unique long names for both objects, while in the Time condition, most participants mapped both objects to the ambiguous short name. Crucially, in the Combined condition, where participants were under pressure to communicate both accurately and efficiently, most pairs converged on the optimal strategy wherein the most frequent object was mapped to the ambiguous short name, and the infrequent object to its unique long name. This made the participants' lexicon both efficient and expressive, in line with Zipf's Law of Abbreviation. Finally, the Neither condition revealed an underlying bias towards avoiding ambiguity.[3]

A logistic regression model was fit in R (R Core Team, 2015) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), with short name use (as contrasted with long name use) as the binary dependent variable, object frequency, experimental condition, and their interaction as fixed effects, and by-participant random intercepts and random slopes for object frequency. This model yielded a significant positive interaction for the frequent object in the critical Combined condition. Thus, in this condition, participants were significantly more likely to assign the short name to the frequent object than in the baseline condition. Participants were significantly *less* likely to assign the short name to either object in the Accuracy condition, and significantly more likely to assign it to both objects in the Time condition, as reflected by the large negative coefficient for the former condition, and the large positive coefficient for the latter. Finally, the intercept is significantly negative, indicating that in the Neither condition, there is a baseline preference for avoiding the short form (see Table 1 for a full list of model coefficients).

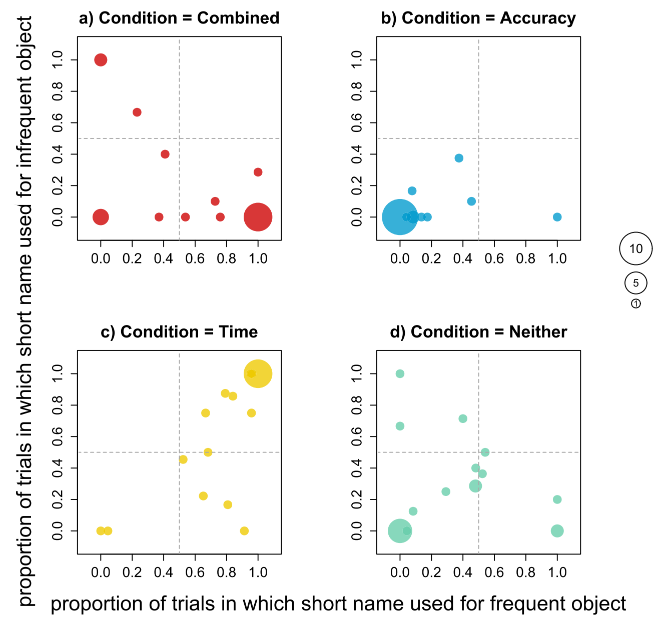

Fig. 3. The proportion of trials in which the short name was used to label the frequent object versus the proportion of trials in which it was used to label the infrequent object. For the Combined (a) and Accuracy (b) condition, each data point combines a pair of communicating players, representing the sum of their director trial productions. For the Time (c) and Neither (d) condition, each data point corresponds to an individual player's productions. The size of the circles is perceptually scaled (Tanimura, Kuroiwa, & Mizota, 2006) to reflect the number of data points coinciding at each value. Data from only the second half of testing trials is shown here, as participants were more likely to have converged on a stable mapping by this time. Data points in the top right quadrant indicate participants who are mostly using the short name for both objects; participants are clustered in this quadrant in the Time condition. Data points in the bottom left quadrant indicate those who are mostly using the unique long names for both objects; participants are clustered here in the Accuracy condition. Data points in the bottom right quadrant indicate participants who are mostly using the short name for the frequent object and the long name for the infrequent object. This behaviour, consistent with the Law of Abbreviation, only reliably arises in the Combined condition, where both pressures are present.

Table 1
Summary of fixed effects for a binomial regression model with short name use as the binary dependent variable, and by-participant random effects for object frequency. The predicted effects are shown in bold. Like Fig. 3, this model is fit using only the second half of each participant's training trial data, as participants were more likely to have converged on a stable linguistic mapping by then.

	β	SE	p
intercept (object = infrequent, condition = Neither)	−2.225	0.501	<0.001
object = frequent	1.392	0.484	0.004
condition = Accuracy	**−5.149**	**0.781**	**<0.001**
condition = Time	**6.031**	**1.207**	**<0.001**
condition = Combined	0.343	0.746	0.645
object = frequent & condition = Accuracy	−0.722	0.751	0.337
object = frequent & condition = Time	−1.079	1.180	0.360
object = frequent & condition = Combined	**2.573**	**0.709**	**<0.001**

In Fig. 4 we plot partipants' 'languages' (the collection of form-meaning mappings produced in their director trials) according to their average token length and the mutual information between their forms f and meanings m: $\sum_f \sum_m p(f,m) \log \frac{p(f,m)}{p(f)p(m)}$.[4] The mutual information between the forms and meanings in a participant's lexicon gives us a measure of how predictable the meanings are given the forms and vice versa, and thus tells us how *expressive* a language is, i.e. how much information is expressed by the forms in the lexicon. The average token length of director trial productions serves as a measure for the effort expended. According to the Principle of

[3] The complete set of raw data from this experiment can be accessed using the following link: http://datashare.is.ed.ac.uk/handle/10283/2702.

[4] We computed the mutual information directly from the empirical distributions, rather than using a bias-corrected estimate; since our use of this measure is for purposes of comparison between participants, we are not concerned with the absolute values, which would be lowered by roughly the same factor across all participants using a bias-correction method such as the Miller-Madow method.

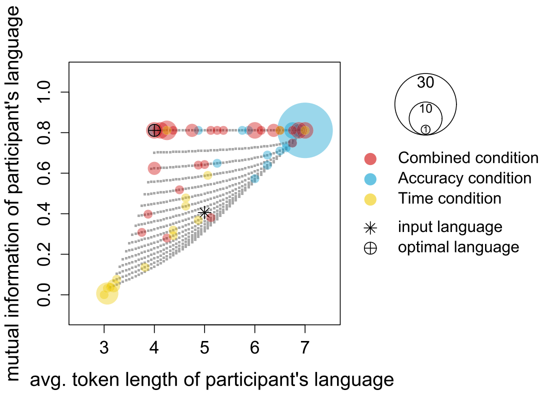

Fig. 4. The average token length of an individual participant's 'language' (the full set of all their director trial productions) plotted against the expressivity (the mutual information between the forms and meanings) of their language. The size of the circles is perceptually scaled (Tanimura et al., 2006) to reflect the number of data points coinciding at each value. The input language that participants are exposed to in training trials is marked with an asterisk, and the grey points represent possible output languages. (Possible output languages are constrained by the number of different expressivity values that are possible for a language with a given average token length. For example, there is only one possible configuration for both the shortest and longest average token lengths—all objects are either mapped to the short name or the long name, respectively—and thus only one possible expressivity value at the endpoints.) The optimal language—the language with the minimum avg. token length while achieving maximum expressivity—is marked with a target symbol.

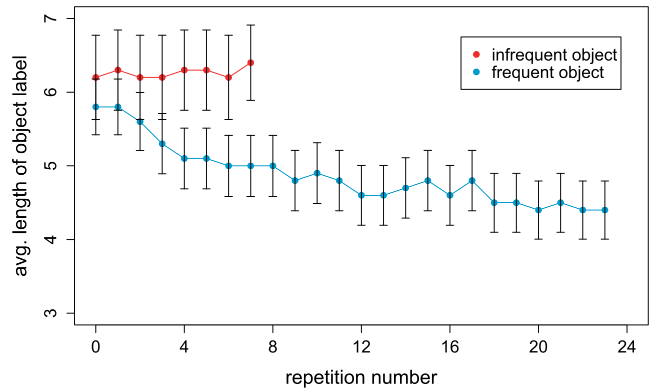

Fig. 5. Timecourse of productions in the critical Combined condition. Each data point shows the average word length taken over all participants' productions at a given repetition number of an object.

Least Effort, an optimal language would maximise expressivity while minimising effort. Only participants in the critical Combined condition produce languages which are optimal in this way. Participants in the Accuracy condition gravitate overwhelmingly towards the strategy that maximises expressivity *and* average token length, and participants in the Time condition maintain minimal average token length but sacrifice expressivity to do so; these were the optimal strategies to use in these respective conditions, given the different task demands.

In Fig. 5, we take a closer look at the possible mechanisms behind participants' trial-by-trial production choices in the Combined condition, by measuring the average length of each object's label over successive repetitions. (Note that participants' frequent and infrequent object production trials are randomly shuffled, and thus repetition number does not correspond with a specific spacing of trial numbers.) As discussed in Section 1, earlier studies by Krauss and Weinheimer (1964) and Clark and Wilkes-Gibbs (1986) show that object descriptions tend to shorten with repetition, and that more frequent objects end up with shorter descriptions simply because they go through more repetitions. In these studies, because the meaning space was small compared to the large descriptive space available (i.e., English phrases with no length restriction), all descriptions could be shortened somewhat without producing ambiguous form-meaning mappings. In our study, we investigated the case where a pressure to use shorter forms comes into direct conflict with the pressure to avoid ambiguity: in this miniature lexicon, shortening yields the same, ambiguous label for the two objects in the meaning space.

If participants are simply more likely to use a shorter form for an object the more times they communicate about that object, then we would expect the average label length for both the frequent object and the infrequent object to decrease at a similar rate as the number of repetitions increases. However, as Fig. 5 shows, this is not what we find. Only the average label length of the *frequent* object decreases with successive repetitions; the average label length of the infrequent object remains roughly constant over the course of the trials. A logistic regression model fit to just the data from the Combined condition, with short name use as the binary dependent variable, object frequency, trial number and their interaction as fixed effects, and by-participant random intercepts and slopes for object frequency and trial number, confirms this. The model results (Table 2) show an overall significant positive effect of trial number on short form use only when the object is frequent. Note that there is also a marginal difference between the two objects at repetition number 0. Thus, in the critical Combined condition, while most participants switch to using the short form for the more frequent object at some point during production trials, most also maintain the long form for the infrequent object throughout the trials—the threat of ambiguity appears to block shortening altogether for this object. This suggests that, in cases where the pressure to decrease effort and the pressure to avoid ambiguity come into direct conflict, language-users' production choices result in systems which maximise expressivity while minimising effort, optimising across the lexicon as a whole.

Interestingly, there were a small number of participants (for example in the Combined condition) who consistently mapped the short form to the *infrequent object*. While shortening the label for either object does satisfy the time pressure to some extent, why might this sub-optimal strategy be used? One possibility is that a participant's strategy is not to optimise based on overall frequency distributions within the signalling game, but simply to shorten the first object they are presented with in production trials, which then blocks shortening of the other object. However, of the 10 participants who were presented with the infrequent object first, 30% converged on a 'reversed' or other non-optimal strategy as opposed to the optimal strategy. Of the remaining 30 participants who saw the frequent object first, 37% converged on a reversed or other non-optimal strategy. Thus, which object appeared in the first production trial (or even the first several trials, which we also checked) is not predictive of which strategy (optimal or otherwise) the participants converged on in the critical condition. We believe these occasional reversed lexicons are thus more likely due to an effect of the cost of switching an incipient convention during the task. For example, if a participant starts out producing labels probabilistically, following the language they were trained on, they will sometimes produce a short name for the infrequent object. If this results in successful communication, and is picked up by a communicative partner, then this pattern may become conventionalised. However, once such a pattern is established, the cost of switching to a different mapping becomes an obstacle. The pressure to maximise the number of correct guesses in the testing trials means the cost of switching labels would further penalise participants who attempted to abandon an incipient sub-optimal convention midway through the task.

Table 2
Summary of fixed effects for a binomial regression model with short name use as the binary dependent variable, and by-participant random effects for object frequency and trial number. This model is fitted to the data from all participants' production trials in the Combined condition.

	β	SE	p
intercept (object = infrequent)	−7.115	2.067	0.001
object = frequent	3.949	2.251	0.079
trial number	0.064	0.059	0.279
object = frequent × trial number	0.137	0.046	0.003

4. Discussion

More than 80 years ago, Zipf hypothesised that the inverse relationship between word length and word frequency was a universal feature of human language, resulting from language users optimising form-meaning mappings for efficient communication. Our study provides direct experimental evidence linking pressures that operate at the level of the individual during communication to the Law of Abbreviation, an emergent structural feature of languages. In particular, language users converge on an optimally-configured lexicon, preferentially using short but potentially ambiguous labels for frequent objects and long labels for infrequent objects. Importantly, this holds only when both a pressure to communicate accurately and a pressure to communicate efficiently are present.

When these pressures were isolated, the Law of Abbreviation did not emerge; an accuracy pressure alone led participants to use the longer non-ambiguous forms regardless of frequency, while a time pressure alone led them to use the short forms. Some participants mapped the short form to the more frequent object in the Neither condition, however the effect was much weaker. Thus, while biases towards accuracy and efficiency might be implicitly present in any linguistic task, emphasising these pressures significantly amplified the effect, as predicted. Even though this experiment involved a miniature lexicon consisting of three possible forms, our result is a proof of concept that such pressures can push a lexicon to align with the Law of Abbreviation. We expect the results to scale up to lexicons with more forms and meanings; with the groundwork in place we can now test this in future studies.

It is important to note, however, that there is a distinction between a language-user's *mental representation* of the lexicon, and the form-meaning mappings they actually produce in communication. Participants using the short form for the frequent object and the long form for the infrequent object may still retain associations of both forms with both objects in their mental lexicon—however, the nature of the communicative task in this experiment may have caused them to produce only the short form for one object and the long form for the other based on purely *pragmatic* considerations (see, e.g., Franke, in press). Given that our experiment only recorded participants' actual productions, we cannot with certainty distinguish between these two possible explanations for the observed behaviour. However, we did include an exit survey which asked participants to explain their strategies during the production stage. Some of the language used in the responses suggested that some participants *had* remapped their mental lexicons. E.g., "I waited until my partner sent Zop twice for the blue round object and then we had a mutual understanding that that's what the Zop was" and "the small round object was Zop, and the orange tall figure was the longer word." However, some other participants indicated that they interpreted the short form as either a prefix or convenient shortening—e.g., "one of the objects had to use the long name, as the short Zop was the same prefix for both" and " [I] used just Zop when transmitting Zopekil [as] the other needed more transmission time"—suggesting that they still retained the long form in their mental lexicon even if they stopped using it.[5]

Our interpretation of such cases is that, while this pragmatics-driven asymmetry in usage may or may not lead to an immediate shift in lexical representations, it may be an important first step in such a change. In English, many words exist that initially began as convenient shortenings of longer forms, which are now either no longer in use, or no longer associated with the same meaning as the short forms. Some examples are: *bus* (from *omnibus*); *wig* (from *periwig*); *pram* (from *perambulator*); *pub* (from *public house*); and *pants* (from *pantaloons*). In all these cases, the clipped form has undergone "opacification", i.e. it is no longer widely recognised as a derivation of the full form, and exists autonomously in the lexicon as an unmarked, standard form (Jamet, 2009). Likewise, even if participants in our experiment are retaining the long form in their mental lexicon, the rapid decrease in its frequency of use over successive generations of learners would likely lead the long form to eventually drop out of the lexicon, with the short form becoming lexicalised as the standard form. Indeed, studies in the iterated learning paradigm show that, in the lexicons produced by successive generations of participants, those in which two labels map to the same meaning are dispreferred (e.g., Reali & Griffiths, 2009; Smith & Wonnacott, 2010). In short, permanent lexical changes often begin life as pragmatics-driven asymmetries in usage (Bybee, 2010). Thus, even if the alignment with the Law of Abbreviation that we observe in participants' usage is not yet accompanied by a corresponding shift in their mental lexicons, it is an important intermediary stage on the way to this outcome.

It is also worth noting that across conditions we found evidence for a baseline preference against ambiguity: when no pressures were present, participants tended towards retaining the unique long forms for *both* objects, and no participants used the ambiguous short names for both objects simultaneously. Indeed, in both conditions featuring a time pressure, a few participants nevertheless used the long names across the board. These results suggest that for some participants, the framing of the task as one of learning a language carries with it some expectation of communicative utility.

Returning to the issue of the explanation for the widespread application of the Law of Abbreviation, our results demonstrate that optimisation behaviour on the part of language-users can lead to the production of lexicons which align with this law. Our study expands on previous work that investigates the relationship between frequency and utterance length, by setting up a small lexicon in which the pressures for efficiency and expressivity in a communicative task come sharply head-to-head. We find that these conflicting pressures do indeed lead language-users to map shorter forms to more frequent meanings, as Zipf hypothesised. However, this result does not rule out that additional processes are involved in shaping this global linguistic pattern as well. Indeed, we expect there are many other factors that come into play as the size of the lexicon is scaled up and the conditions become closer to actual language-use: for example the bottlenecks of learning and memory; the influence of predictability in context; constraints of speech production; and the propagation of errors. There may be a role for random statistical processes to play as well. Future work should focus on how the pressures involved in this task interact with these and other factors, and especially on how the behaviour of individuals communicating in a pair spreads outside this context to the level of an entire population.

[5] All the exit survey responses are available along with the full dataset at: http://datashare.is.ed.ac.uk/handle/10283/2702.

5. Conclusions

Zipf's proposal—that the inverse relationship between a word's length and its frequency is a universal design feature of language—has been borne out repeatedly in observations of the world's languages (Ferrer-i-Cancho & Hernández-Fernández, 2013; Piantadosi et al., 2011; Sigurd et al., 2004; Strauss et al., 2007; Teahan et al., 2000). The long-standing explanation for this phenomenon appeals to the idea that language users want to communicate as efficiently as possible. However, the critical link between this Principle of Least Effort and the emergence of an optimal lexicon has remained largely untested. Our study explored the hypothesis that the mechanisms operating in individual language users during online language production can result in the active restructuring of a lexicon. Our findings reveal that when pressures to communicate accurately and efficiently are both present and in conflict, language users exploit information in the input about the frequency of meanings to converge on an optimally-configured lexicon. When only one of these pressures is present, the effect does not emerge. This result provides evidence that the universal pattern Zipf observed can indeed arise through individual-level optimisation of form-meaning mappings. More generally, this method provides a model for future work showing how explanations of population-level properties of languages can be grounded in the moment-to-moment behaviours of individuals.

Acknowledgments

Thanks to Michael Franke, an anonymous reviewer, Vanessa Ferdinand, Kevin Stadler, and other members of the Centre for Language Evolution for useful feedback and discussion on this work.

References

Antoine, F. (2000). *An English-French dictionary of clipped words* (Vol. 106). Peeters Publishers.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*, 306–329.

Dautriche, I., Mahowald, K., Gibson, E., Christophe, A., & Piantadosi, S. (in press). Words cluster phonetically beyond phonotactic regularities. *Cognition, 163*, 128–145.

Davies, M. (2008). *The corpus of contemporary american english: 520 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 376–381).

Ellis, S. R., & Hitchcock, R. J. (1986). The emergence of zipf's law: Spontaneous encoding optimization by users of a command language. *IEEE Transactions on Systems, Man and Cybernetics, 16*, 423–427.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109*, 17897–17902. http://dx.doi.org/10.1073/pnas.1215776109.

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularization of unpredictable variation. *Journal of Memory and Language, 91*, 158–180.

Ferrer-i-Cancho, R., & Hernández-Fernández, A. (2013). The failure of the law of brevity in two new world primates. statistical caveats. *Glottotheory International Journal of Theoretical Linguistics, 4*, 45–55 <http://www.degruyter.com/view/j/glot.2013.4.issue-1/glot.2013.0004/glot.2013.0004.xml>.

Ferrer-i Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science, 37*, 1565–1578.

Ferrer-i-Cancho, R., & Moscoso del Prado, F. (2012). Information content versus word length in random typing. *JSTAT*.

Franke, M. (2017). Game theory in pragmatics: Evolution, rationality & reasoning. In *Oxford Research Encyclopedia of Linguistics*, in press.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell Labs Technical Journal, 29*, 147–160.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE, 40*, 1098–1101.

Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research, 21*, 485–496.

Jamet, D. (2009). A morphophonological approach to clipping in English. Can the study of clipping be formalized? *Lexis: Journal in English Lexicology, HS 1*.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*, 102–110.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*, 10681–10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1*, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343.

Moscoso del Prado, F. (2013). The missing baselines in arguments for the optimal efficiency of languages. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1032–1037) <http://csjarchive.cogsci.rpi.edu/Proceedings/2013/papers/0203/paper0203.pdf>.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*, 3526–3529. http://dx.doi.org/10.1073/pnas.1012551108.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280–291. http://dx.doi.org/10.1016/j.cognition.2011.10.004.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing <https://www.R-project.org/>.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition, 111*, 317–328.

Sigurd, B., Eeg-Olofsson, M., & Van Weijer, J. (2004). Word length, sentence length and frequency-Zipf revisited. *Studia Linguistica, 58*, 37–52.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*, 444–449.

Strauss, U., Grzybek, P., & Altmann, G. (2007). Word length and word frequency. In *Contributions to the science of text and language* (pp. 277–294). Netherlands: Springer.

Tanimura, S., Kuroiwa, C., & Mizota, T. (2006). Proportional symbol mapping in R. *Journal of Statistical Software, 15*.

Teahan, W. J., Wen, Y., McNab, R., & Witten, I. H. (2000). A compression-based algorithm for chinese word segmentation. *Computational Linguistics, 26*, 375–393.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107*, 729–742.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition, 7*, 415–449.

Zipf, G. K. (1935). *The psycho-biology of language* (Vol. ix) Oxford, England: Houghton Mifflin.