SOCIAL SCIENCES

Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche

Christophe Coupé[1,2]*, Yoon Mi Oh[3,4]*, Dan Dediu[1,5], François Pellegrino[1]†

Language is universal, but it has few indisputably universal characteristics, with cross-linguistic variation being the norm. For example, languages differ greatly in the number of syllables they allow, resulting in large variation in the Shannon information per syllable. Nevertheless, all natural languages allow their speakers to efficiently encode and transmit information. We show here, using quantitative methods on a large cross-linguistic corpus of 17 languages, that the coupling between language-level (information per syllable) and speaker-level (speech rate) properties results in languages encoding similar information rates (~39 bits/s) despite wide differences in each property individually: Languages are more similar in information rates than in Shannon information or speech rate. These findings highlight the intimate feedback loops between languages' structural properties and their speakers' neurocognition and biology under communicative pressures. Thus, language is the product of a multiscale communicative niche construction process at the intersection of biology, environment, and culture.

INTRODUCTION

Language is universally used by all human groups, but it hardly displays undisputable universal characteristics, with a few possible exceptions related to pragmatic and communicative constraints (*1*, *2*). This ubiquity comes with very high levels of variation across the 7000 or so languages (*3*). For example, linguistic differences between Japanese and English lead to a ratio of 1:11 in their number of distinct syllables. These differences in repertoire size result in large variation in the amount of information they encode per syllable according to Shannon's theory of communication. Despite those differences, Japanese and English endow their respective speakers with linguistic systems that fulfill equally well one of the most important roles of spoken communication, namely, information transmission. We show here that the interplay between language-specific structural properties (as reflected by the amount of information per syllable) and speaker-level language processing and production [as reflected by speech rate (SR)] leads languages to gravitate around an information rate (IR) of about 39 bits/s. This finding, based on quantitative methods applied to a large cross-linguistic corpus of 17 languages, highlights the intimate feedback loops between languages and their speakers due to communicative pressures. We suggest that this phenomenon is rooted in the human neurocognitive capacity, probably present in our lineage for a long time (*4*), and that human language can be analyzed as the product of a multiscale communicative and cultural niche construction process involving biology, environment, and culture (*5*).

Each human language provides its speakers with a communication system that fulfills their needs for transmitting information to their peers. The Uniform Information Density hypothesis (*6*) and similar approaches [e.g., (*7*) and (*8*)] suggested that speakers distribute information along the speech signal following a smooth distribution rather than high-amplitude fluctuations. Compatible with Shannon's theory, this optimization process guarantees the robust information transmission at a rate close to the channel capacity. We adopt here a quite different perspective, where we compare, across very different languages, the average rates at which information is emitted. This approach enables us to estimate the channel capacity and to assess whether the large differences observed among languages in terms of encoding result in analog differences in channel capacity or, conversely, whether there exist compensating strategies that go beyond the local adaptation operating during speech production. Therefore, we investigate the interaction between information encoding and average SR and, more specifically, whether the variation among languages in IR is regulated by communicative constraints. Thus, does too low an IR hinder communicative efficiency? And, at the other extreme, does pushing it too high incur too heavy physiological and cognitive costs? While a negative correlation between average SR and the informativeness of linguistic constituents has been demonstrated in a small multilanguage corpus (*9*), the distribution of IRs across human languages is almost totally unknown despite its crucial importance for understanding human spoken communication. While our data here come only from speech production (information encoding), our results, nevertheless, implicitly address also speech perception (information retrieval) and processing, as they are all intimately coupled and coevolve during language acquisition, use, and change (*10*).

We have chosen to focus here on the syllable as the information-encoding unit for both linguistic and cognitive reasons. On the linguistic side, despite a long-lasting debate in phonology about whether the syllable is a universal unit in the world's languages (*11*) being a cornerstone of this controversy, analyzing the encoding of information in terms of syllables does offer several advantages over other levels of linguistic description (such as phonemes and morphemes). First, syllables are much less prone than phonemes to complete deletion in casual speech, allowing more robust estimates of SR (*12*) [readers can also refer to (*9*) for a more detailed discussion on this matter]. Moreover, we chose the syllable over meaning-bearing units (morphemes or words), as the latter levels rely more on a language-specific linguistic analysis, on top of various methodological difficulties affecting their robust counting in a cross-linguistic framework (see text S1 and fig. S1 for a discussion on the

[1]Laboratoire Dynamique du Langage UMR5596, CNRS, University of Lyon, 14 Avenue Berthelot, 69007 Lyon, France. [2]Department of Linguistics, The University of Hong Kong, Pokfulam Road, Hong Kong. [3]New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. [4]Department of French Language and Literature, Ajou University, Suwon, 16499 Gyeonggi-do, South Korea. [5]Collegium de Lyon, University of Lyon, 24 rue Baldassini, 69007 Lyon, France.
*These authors contributed equally to this work.
†Corresponding author. Email: francois.pellegrino@univ-lyon2.fr

relation between meaning and information encoding). On the neurocognitive side, the past decade has witnessed an abundance of models and studies that underpin the pivotal role of the syllabic time scale for speech comprehension, especially through the entrainment of cortical oscillations by the speech signal [see (*13*–*16*), among others]. These findings led to a view where "the sensitivity to syllable rate [is] arguably the most fundamental property of speech perception and production" (*16*), a view particularly relevant to our study here.

We studied a sample of 17 languages from 9 language families spread across Europe and Asia, showing a remarkable diversity in terms of linguistic and typological features at all levels, from phonetics and phonology to morphology and syntax and to semantics and pragmatics (see table S1). Focusing on their phonetics and phonology, these languages vary in their number of phonemes (from 25 in Japanese and Spanish to more than 40 in English and Thai), the number of distinct syllables (from a few hundred in Japanese to almost 7000 in English), tonal complexity (from none to six contrastive tones), and various other phonological phenomena (e.g., vowel harmony is present in Finnish, Hungarian, Korean, and Turkish). Thanks to its size and diversity, this sample is adequate to reveal robust trends reflecting phenomena that can potentially be extrapolated to human language in general.

We collected recordings of 170 native adult speakers of the aforementioned 17 languages, each reading at their normal rate a standardized set of 15 semantically similar texts across the languages (for a total amount of approximately 240,000 syllables). Speakers became familiar with the texts, by reading them several times before being recorded, so that they understand the described situation and minimize reading errors (see Materials and Methods below for more details). For each recording, we extracted the duration [in seconds, excluding pauses longer than 150 ms, i.e., longer than typical phonemic silences (*17*)] and the total number of syllables (NS) of the text's "canonical" pronunciation. This term refers to the standard pronunciation found in dictionaries and lexical databases (*12*). For instance, the word "probably" in English will be transcribed as [pɹɑ.bə.bli] and accounted for three syllables, even if some speakers adopted a pronunciation variant such as [pɹɑ.bli]. By adopting this convention, we considered the NS encoded in the signal and potentially retrieved from it. We computed the ratio between the NS and duration, which will be denoted as SR here (rather than the more precise but also less transparent "canonical articulation rate"). Using read speech (as opposed to spontaneous or conversational speech), we constrained the speakers in terms of lexical and syntactic strategies, and we encouraged them to adopt a clear speech pronunciation. Moreover (and very important here), we emphasized the cross-language comparability of the information encoded and retrievable (i.e., the canonical syllables), rather than the various reductions potentially performed by the speakers. Indeed "[s]peakers can produce utterances with more or less articulatory detail or even completely omit certain words, while still conveying the same message" (*18*). Last, for German, the canonical and realized SRs in a "normal, clearly spoken style" (which is somewhat similar to read style) have been shown to exhibit virtually no difference, even at the phonemic level, according to (*19*), providing yet another argument for considering *SR* as relevant here (unfortunately, because of the lack of cross-linguistic robustness of the automatic estimation of realized SR, we could not check this in our cross-linguistic database; see text S2 and fig. S2).

In parallel, from independently available written corpora in these languages, we estimated each language's information density (*ID*) as the syllable conditional entropy to take word-internal syllable-bigram dependencies into account. We then computed the average IR by multiplying the ID by the SR for each text read by each speaker in our dataset. Individual SR varies in a ratio of more than one to two, with the slowest speaker hovering around 4.3 syllables/s and the fastest one reaching 9.1 syllables/s on average. *ID*, computed at the language level, varies in a more limited but still substantial way (from 4.8 bits per syllable for Basque to 8.0 bits per syllable for Vietnamese).

RESULTS

As a preliminary analysis, we checked whether our definition of *ID* provides a relevant measure of linguistic ID, using the syntagmatic density of information ratio (SDIR), defined in (*9*), as a control. *SDIR* quantifies the relative informational density of language *L* compared to a reference language, based on the semantic information expressed in the context of a limited oral corpus (see Materials and Methods below for more details). It thus provides the ground truth on the semantic information conveyed by the sentences in the spoken corpus. Following (*9*), we used Vietnamese as a reference, such that a language *L* with a ratio bigger than one (or, respectively, less than one) is denser (respectively, less dense) than Vietnamese in terms of semantic information. By contrast, being estimated from a very large written lexical database, *ID* subsumes an overall syllable usage disregarding any semantic consideration. The preliminary analysis nevertheless shows that the two information quantification approaches are connected; we obtain, for our data, a very high correlation between *ID* and *SDIR* (Pearson's $r = 0.91$, $P = 3.4 \times 10^{-7}$ and Spearman's $\rho = 0.80$, $P = 0.00011$), which suggests that, despite differences in material (heterogeneous and written corpus versus parallel and spoken corpus) and nature (an entropy measured on a large lexicon versus a normalized ratio derived from small texts), our *ID* is a good estimate of the average amount of information per syllable.

We next attempted to model the *SR* and *IR* distributions using linear mixed-effects regression, but we observed heteroscedasticity of the residuals in both cases. Therefore, we decided to use generalized additive models for location, scale, and shape (GAMLSS) (*20*, *21*), as they allowed us to model both the mean (μ) and SD (σ) of Gaussian distributions, considering sex as fixed effect and text, language, and speaker as random effects (with a log link function for σ). This resulted in a better fit to the data [as judged by the Akaike information criterion, with AIC (*SR*, fixed σ) − AIC (*SR*, modeled σ) = 171.2 and AIC (*IR*, fixed σ) − AIC (*IR*, modeled σ) = 167.5], a distribution of residuals very close to normality, and a reduced heteroscedasticity to the point where no additional corrections were necessary.

We found that *IR* is centered on a mean of 39.15 bits/s with an SD of 5.10 bits/s, while *SR* is centered on a mean of 6.63 syllables/s, with an SD of 1.15 syllables/s (see Fig. 1). The fixed effect of sex is significant for both *SR* and *IR*, with females having significantly lower means (*SR*, −0.17; *IR*, −1.01) and SDs (*SR*, −0.06; *IR*, −0.06); this finding extends previous observations on English (*22*) to a larger set of languages from different families and geographic areas. In addition, most of the variation in the random effects for both mean and SD is by language [SR, $\sigma_b(\mu) = 0.87$ and $\sigma_b(\sigma) = 0.15$; *IR*, $\sigma_b(\mu) = 3.10$ and $\sigma_b(\sigma) = 0.16$] and speaker (*SR*, $\sigma_b(\mu) = 0.57$ and $\sigma_b(\sigma) = 0.18$; *IR*, $\sigma_b(\mu) = 3.39$ and $\sigma_b(\sigma) = 0.19$]. The model suggests that the relative impact of the two random factors differs, language having the largest impact on *SR*, while speaker has the largest on *IR*. In other words, while *SR* is mainly clustered by language and relatively less so by speaker, the influence of this language-level clustering is reduced for *IR* [family has a very small effect beyond language, with *SR* $\sigma_b(\mu) = 0.000019$ and $\sigma_b(\sigma) = 0.00019$ for SR and *IR* $\sigma_b(\mu) = 0.00023$ and $\sigma_b(\sigma) = 0.00011$ for IR]. The

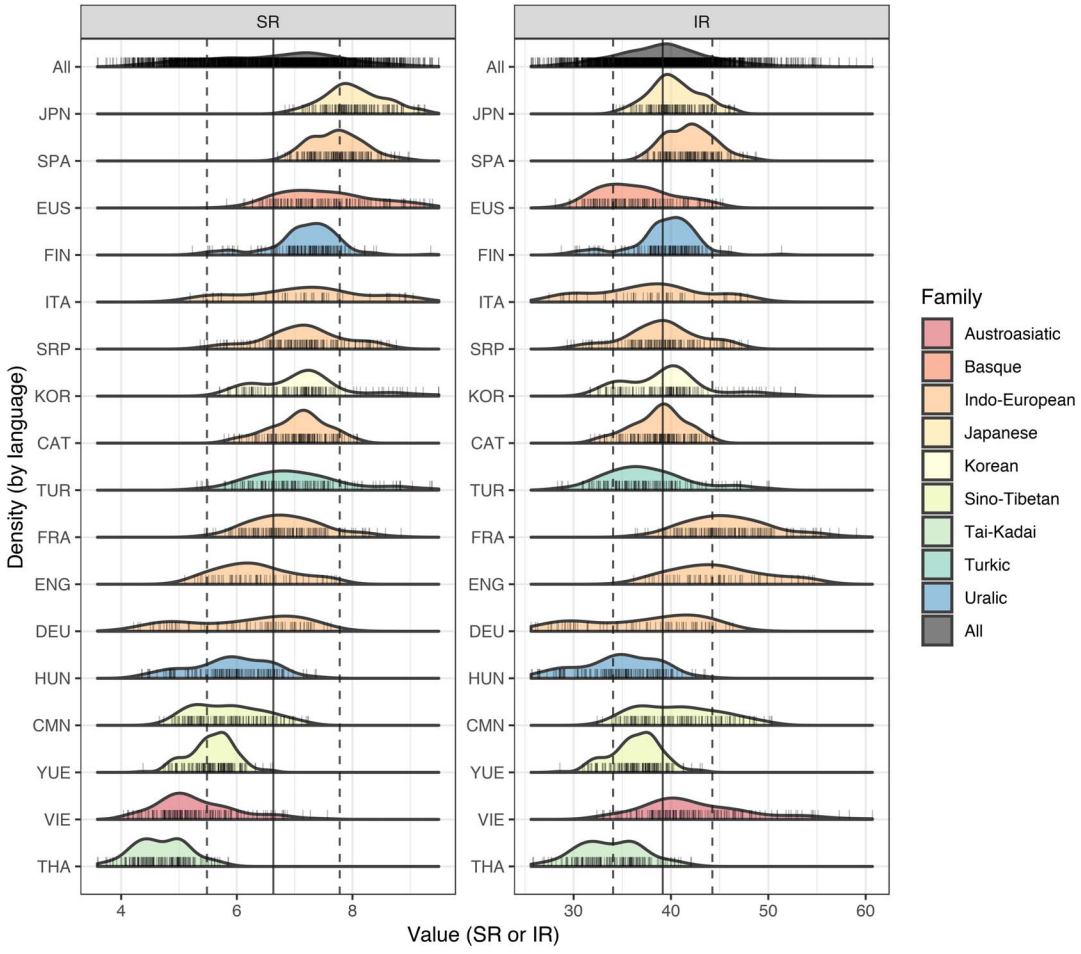

Fig. 1. SR and IR. The distribution of SR (in syllables per second) (left) and IR (in bits per second) (right) within the languages in our database (colored areas; colors represent the language families) and across them (black areas at the top) using a Gaussian kernel density estimate. The black vertical lines spanning the whole plot represent the means (solid lines) ± 1 SD (dashed lines). The short black vertical lines represent the actual data points.

style differences among the 15 texts have a much smaller effect, as revealed by the small variation associated with text for both *SR* and *IR* [*SR*, $\sigma_b(\mu) = 0.11$ and $\sigma_b(\sigma) = 0.0005$; *IR*, $\sigma_b(\mu) = 0.66$ and $\sigma_b(\sigma) = 0.00028$]. We included the speaker's age (and its interactions with the other factors) in the models, and we found that while, on the one hand, its effects are as expected (i.e., a negative impact on *SR*), on the other, its inclusion does not improve the model fit [according to the AIC and BIC (Bayesian information criterion)]. Therefore, we adopt the simpler models without age for the main analysis reported here but the models including it are available in the analysis report file S1.

To explore the relationship between *SR* and *ID*, we included *ID* as a fixed effect in the GAMLSS modeling of *SR* (here, we dropped language as a random effect, since there is, by definition, a single *ID* value per language, but we did include family): We found a significant negative effect of *ID* not only on the mean of *SR* ($\beta = -0.89$, $P < 2.2 \times 10^{-16}$) but also on its SD ($\beta = -0.09$, $P = 6.4 \times 10^{-7}$). This negative relationship (see Fig. 2)—between two parameters derived from independent written and oral corpora—indicates that there is a trade-off between *SR* and *ID*, the languages with lower IDs being spoken faster, as also illustrated by "classic" correlation estimates (Pearson's $r = -0.71$ and Spearman's $\rho = -0.70$, in both cases with $P < 2.2 \times 10^{-16}$).

The visual inspection of the distributions of *SR* and *IR* (Fig. 1, black areas) suggests that *IR* and *SR* differ in terms of the compactness of their overall distribution and that languages are more similar in terms of *IR* than *SR*. To assess this difference, we computed several pairwise divergence metrics between languages (Kolmogorov-Smirnov, Kullback-Leibler, Jensen-Shannon, Hellinger, and chi-square divergences; Fig. 3 and analysis report file S1) to quantify their dispersion in the distribution of number of syllables per text (*NS*), *SR*, and *IR*. *NS* is considered here as a proxy for information dilution, since the texts are semantically similar across the languages. Using randomization paired *t* tests (with 1000 permutations), we found that, for all measures, languages are significantly more similar to each other in *IR* than in *NS* and *SR* (all randomization $P < 10^{-4}$). Last, *IR*s are less dispersed around their mean than *SR*, as shown by their coefficients of variation (17.3% for *SR* versus 13.0% for *IR*) and also by several unimodality tests with permutation (see analysis report file S1).

DISCUSSION

In this study, we investigated the relationship between ID (estimated from written data) and SR (computed from parallel spoken data) across 17 languages. By recording read parallel data rather than more

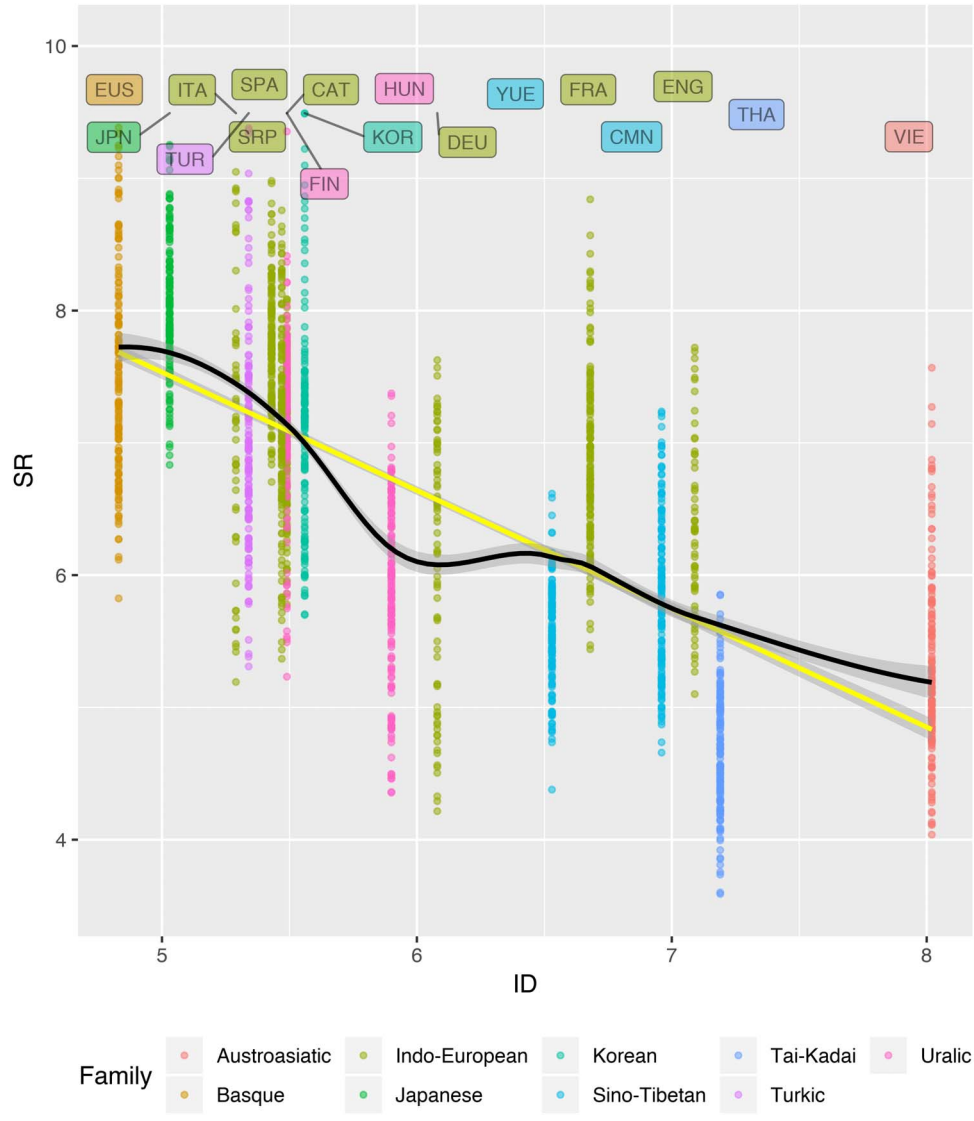

Fig. 2. Relationship between SR and ID across languages Colors represent the language families, and individual languages are identified by the labels on top (to avoid overlapping labels, short black lines might show their actual position). While there is only one value of *ID* per language, there are as many values of SR per language as texts read by individual speakers. The straight yellow line represents the linear regression [with 95% confidence interval (CI)], and the black curve represents the locally estimated scatterplot smoothing regression (with 95% CI) of SR on *ID*.

casual or spontaneous speech, we deliberately increased the comparability across languages and speakers and constrained the degrees of freedom available to each speaker [in line with (*23*)]. Having the same texts read by all speakers controls for one of the main effects reported in (*24*), namely the fact that, for a given language, "fast speakers are likely to produce less informative content." Therefore, while this corpus is not appropriate for studying pragmatic and cognitive planning, it does allow robust results in what concerns the differences across languages and speakers given a controlled linguistic content. In addition, it does not require any preliminary data curation that could raise methodological concerns and potentially induce biases.

We argue that our results, based on a controlled linguistic material and consisting of read speech, do reflect actual phenomena found in more natural settings. As such, we found that the effects of text in the *SR* and *IR* models are much smaller than those of speaker and language, despite the stylistic differences among the 15 texts (some corresponding to phone information request scenarios, e.g., P0, while others are narratives, such as P8). This aspect rules out the existence of text-related systematic bias across languages and suggests that when talking at a normal rate, each speaker's average SR is quite robust to variation in linguistic content (lexical frequency, syntactic structure, phrase length, etc.). This is still fully compatible with the local changes in SR that have been extensively demonstrated, in a few languages at least [see (*18*, *24*) among many others]. The limited effect of text also suggests that the results reported here should hold for interactions involving styles similar to the normal, clearly spoken style in Koreman's terminology (*19*). In other words, as long as the communication situation requires a correct decoding of the linguistic information encoded by the speaker, we suggest that the trade-off presented above will be observed. We can expect that its strength would gradually decrease along a

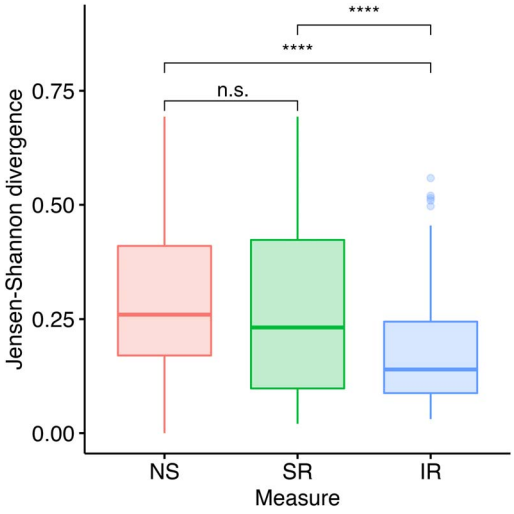

Fig. 3. Pairwise divergence between languages. The distribution of the Jensen-Shannon divergence between pairs of languages for the NS, SR, and IR, also showing the significant differences using a randomization paired *t* test (1000 permutations). The *IR-SR* and *IR-NS P* values are $<10^{-4}$, while the *NS-SR P* value is 0.30. All other divergence measures produce essentially identical results. n.s., not significant. ****$P \leq 0.0001$.

continuum ranging from very carefully pronounced content to very informal interactions where understanding is heavily reliant on contextual and pragmatic factors rather than on the linguistic information itself.

Together, our findings show that while there is wide interspeaker variation in speech and IRs, this variation is also structured by language. This means that an individual's speech behavior is not entirely due to individual characteristics but is further constrained by the language being spoken. The effect of sex we found here is analogous to its effect in other phenomena, such as, for example, body height or the fundamental voice frequency. While both have universally constrained ranges in humans (*25*, *26*) and result from complex interactions between genetics and environment (*27*, *28*), they differ between languages/groups (*26*, *29*) and sexes (*25*, *26*). Of relevance here, the statistically significant difference between males and females does not preclude universal tendencies or between-group patterns of variation (*30*).

However, languages seem to stably inhabit an optimal range of IRs, away from the extremes that can still be available to individual speakers. Languages achieve this balance through a trade-off between ID and SR, resulting in a narrower distribution of IRs compared to SRs. In the introduction, we rhetorically asked whether too low or too high an IR would impede communicative and/or cognitive efficiency. Our results here suggest that the answer to both questions is positive and that human communication seems to avoid two extreme sociolinguistic profiles: on the one hand, high *ID* languages spoken fast by their speakers ("high-fast"), and, on the other, low *ID* languages spoken slowly by their speakers ("low-slow"). Both speakers and listeners have an interest in avoiding high-fast languages: For the speaker, production comes at higher costs both in terms of articulation (more complex and infrequent, less routinized syllables) and planning (since they are less predictable from the context), while for the listener, the resulting speech flow may exceed channel capacity or at least be challenging in terms of lexical access and syntactic parsing. Avoidance of high-fast languages may thus result from a convergence of production- and perception-oriented pressures, with similar factors being suggested in (*24*) to explain that in American English corpora of conversational speech, fast speakers produced less informative content (both in terms of content words and syntactic structure). On the other hand, low-slow languages, if they existed, would present a twofold challenge: First, in terms of general communicative efficiency, they would lead to longer turns in interaction [in human interactions, turn duration is 2 s on average (*1*)]. A second—and probably related—factor is that they would require their speakers to keep longer chunks in working memory, for a given informational content. One can thus hypothesize that speakers from this language would swiftly accelerate their articulation rate to compensate for their language's low *ID*.

This study provides the most extensive estimation of spoken IR to date, whether in terms of numbers of speakers, languages, or language families. Such an IR centered on 39 bits/s (with an SD of about 5 bits/s) is certainly compatible with the rare estimates available for English, Mandarin Chinese, and Spanish (*31*, *32*). The most notable result is that languages are much closer in terms of IR than in SR. Despite the across-language dispersion observed for ID and SR, their regulatory interaction seems to give rise to a universal attractor. This result is far from trivial, especially considering that the substantial freedom speakers have to depart from their average SR without any apparent effort (*33*). Despite this essential capacity enabling each speaker to adapt to specific situations of communication, a convergence is observed, and the deviation from a flat distribution shown here could be explained by a soft constraint toward an average IR of around 39 bits/s.

Metaphorically, our data suggest that languages tend to inhabit a valley of possible IRs with gradual slopes, which allow some speakers to occupy peripheral positions farther away from the attractor in both directions. We suggest that this valley in a fitness landscape illustrates the concept of "good-enough" control proposed as an alternative to optimal control for biological systems (*34*) and that its existence is due to functional and cognitive factors. Several recent proposals highlight that the neural capacity to track speech dynamics through cortical oscillations is crucial for speech processing and understanding, especially in the so-called θ range (*13*), for which an optimal speech-brain alignment would be essential for syllabic sampling (*14*). This line of research led not only to an estimation of an auditory channel capacity of about 9 syllables/s for American English participants (*13*) but also to a much more restricted "optimal" range around 4.5 syllables/s (*16*). This notion of optimality is still to be refined, and it is not yet established whether the auditory channel capacity is a matter of information or of acoustic duration: These experiments manipulate SR within a given language (generally English), and cross-linguistic assessments will be necessary to estimate whether the boundaries of the θ range depend on each individual's mother tongue and to revisit the notion of optimal rate. More specifically, we show here that between-language differences in IR are much smaller than those in SRs. Consequently, given that all humans are fully cognitively equipped regardless of their mother language, IR provides a better candidate than SR for investigating invariance in cognitive capacity. This line of investigation can shed new light on the long-lasting difficulty faced in attempting to detect temporal regularities and predictability in the acoustic signal. The apparent discrepancy observed between the arguable existence of linguistic rhythmic classes and the lack of temporal regularities is beyond the scope of this paper, but interested readers can refer to (*35*) for a recent cross-linguistic approach that found limited evidence for temporal predictability.

We suggest that this cross-linguistic tendency stems from the interaction between social and neurocognitive pressures that define an optimal range for IR, around which the complex adaptive system

(consisting of each language and its speakers) hovers. While ID is mainly a property derived from the language itself (its grammar, lexicon, and long-term usage), SR reflects individual speakers' behavior instantiating language-level norms and constraints generated in their own biocognitive apparatus. The interplay of this long-term collective property with this individual short-term behavior, we propose, leads all individuals to continuously monitor (consciously or not) and adapt their SR to the specific linguistic and communicative context. A prediction is that when a community implements linguistic changes that may cause the IR to drift away from the optimal range, compensatory mechanisms that affect SR (e.g., coarticulation) may bring the average IR back toward optimal regions.

Speakers are obviously not limited to manipulating the phonological level of their languages to achieve an efficient communication, and an obvious extension of this study will be to take grammatical information into account. This can be achieved by considering a longer context when estimating the syllable ID (thus, going beyond the previous syllable within the same word) or by moving to units longer than the syllable (such as words). Doing this will, however, force one to depart from the "mere encoding" considered here and get closer to higher levels of language-specific strategies, generating new challenges in terms of (semantic and pragmatic) information quantification [see (*36*) for a discussion from the angle of overt and hidden linguistic complexity], challenges that require different methods and data and that we leave for future research.

To conclude from a broad evolutionary perspective, we thus see human language as inhabiting a biocultural niche spanning two scales. At a local scale, each system consisting of a given language and its speakers represents one instantiation of a cultural niche construction process (*5*, *37*) in a specific context involving the ecological (*38*), biological (*39*), social (*40*), and cultural (*41*) environments. At a global scale, all of these language speakers' local systems are subjected to universal communicative pressures characterizing the human-specific communication niche and consequently fulfilling universal functions of communication essential for the human species.

MATERIALS AND METHODS

Data and code availability

The data are contained in two TAB-separated CSV files, the R code for the analysis and plotting is contained in an RMarkdown script, and all the results and plots (obtained by compiling this RMarkdown script) are in an HTML analysis report; all these files are freely available under an open-source license in the Supplementary Materials and in the GitHub repository https://github.com/keruiduo/SupplMatInfoRate.

Data

We analyzed here the data from 17 languages from 9 language families, listed below as family [language name (ISO 639-3 code)]: Austroasiatic [Vietnamese (VIE)], Basque [Basque (EUS)], Indo-European [Catalan (CAT), German (DEU), English (ENG), French (FRA), Italian (ITA), Spanish (SPA), and Serbian (SRP)], Japanese [Japanese (JPN)], Korean [Korean (KOR)], Sino-Tibetan [Mandarin Chinese (CMN) and Yue Chinese/Cantonese (YUE)], Tai-Kadai [Thai (THA)], Turkic [Turkish (TUR)], and Uralic [Finnish (FIN) and Hungarian (HUN)]. For each language, we collected existing or new oral and text corpora and performed analyses considering syllable as the reference unit. Syllable is both regarded as "a unit in the organization of the sounds of an utterance" (*42*) and as a salient unit for cognitive processing by psycholinguists and neuroscientists. We opted for the syllabic level to focus on the direct mapping between semantic information and speech signal (seen as a sequence of syllabic "bricks"), bypassing the mental lexicon, which is highly dependent on the language morphological characteristics. The linguistic implications go beyond the scope of this paper and are not further discussed.

The oral corpus (see text S3 for details) is initially derived from the MULTEXT (Multilingual Text Tools and Corpora, ID: ELRA-S0060) parallel corpus (*43*). For British English, German, and Italian, we considered 15 short texts from this corpus, all composed of three to five semantically connected sentences carefully translated by a native speaker from the British English original. For the other 14 languages, two of the authors (C.C. or Y.O.) supervised the translation and recording of new datasets by native speakers of the target language, preferably of a specific variety whenever possible (e.g., Mandarin spoken in Beijing, Serbian in Belgrade, and Korean in Seoul). No strict control of age or other sociolinguistic variables was imposed, but speakers (170 in total, 85 females) were mainly students or members of academic institutions. This data collection complied with ethical regulations at the Université de Lyon, and given its nature, it did not require a formal approval by an Ethics Committee. After providing informed consent, speakers were asked to read each text first silently once and then aloud at least two times, allowing familiarization and reducing reading errors. The ROCme! software (*44*) was used for presentation of the experiment instructions and texts, as well as for recordings. The texts were presented one by one on the screen in random order, one sentence at a time following a self-paced reading paradigm, with the second or the third aloud recording being analyzed here. Thus, in total, we have 2265 data points (i.e., each data point consists of a text t read by a speaker s in language L; some speakers did not read all texts). For each such data point, we measured the total speech duration (D; in seconds) and the total NS of the text's canonical transcription (denoted NS). Pauses longer than 150 ms were identified and discarded through visual inspection of the waveforms and spectrograms. We computed the SR for text t in language L read by speaker s ($\mathrm{SR}_{t,s}^{L}$) as

$$\mathrm{SR}_{t,s}^{L} = \frac{\mathrm{NS}_{t}^{L}}{D_{t,s}^{L}} \tag{1}$$

and the syntagmatic density of information ratio [SDIR^{L}; defined in (*9*)] for language L as

$$\mathrm{SDIR}^{L} = \frac{1}{N_T}\sum_{t=1}^{N_T}\frac{\mathrm{NS}_{t}^{\mathrm{VIE}}}{\mathrm{NS}_{t}^{L}} \tag{2}$$

with $N_T = 15$ (number of distinct texts in the oral corpus). It can be seen that, by definition, an $SDIR^{L} > 1$ represents a language L denser than Vietnamese in terms of semantic information (since it requires less syllables than Vietnamese to encode a similar semantic content), while an $SDIR^{L} < 1$ represents a language L less dense than Vietnamese.

The text corpora were acquired from various sources containing large amounts of written text (see table S2 for details). After an initial data curation, each corpus was phonetically transcribed and automatically syllabified using a rule-based program written by one of the authors (Y.O.), except (i) when syllabification was already provided with the dataset (for English, French, German, and Vietnamese for the multisyllabic words) and (ii) when the corpus was syllabified by

an automatic grapheme-to-phoneme converter (for Catalan, Spanish, and Thai). In addition, no syllabification was required for Sino-Tibetan languages (Cantonese and Mandarin Chinese) since each ideogram corresponds to a single syllable. When applicable, syllables bearing specific tone or accent were considered as distinct in the inventory. For more information on the data and its processing, see (*45*).

For each language *L*, we computed information-theoretical metrics derived from Shannon's seminal theory to estimate the average amount of information transmitted per syllable. More precisely, we estimated the first- and second-order entropies of the syllable distribution. The first-order entropy is the standard Shannon entropy (ShE)

$$\mathrm{ShE} = -\sum_{x} p(x).\log_2(p(x)) \quad (3)$$

where $p(x)$ is the maximum likelihood estimates of the syllable unigram probabilities observed in the corpus.

The second-order entropy is the main information index used here, and it is thus denoted by *ID*. It refers to conditional entropy where the context in which each syllable occurs is taken into account. We characterized this context as the identity of the previous syllable or a null marker for syllables occurring word initially (thus, no bigrams span across word boundaries)

$$\mathrm{ID} = -\sum_{x,y} p(x,y).\log_2\left(\frac{p(x,y)}{p(x)}\right) \quad (4)$$

where $p(x, y)$ is the maximum likelihood estimates of the syllable bigram probabilities observed in the corpus.

The numerical difference between the first- and second-order entropies differs among languages, being larger for languages where across-syllable binding is tighter because of morphology. Given that, for several languages, the text corpora only provide word frequencies (and not raw texts), we considered within-word context only. In future studies, a broader and across-word context could be considered so that we can refine the entropy estimations, but the very strong correlation observed between *ID* and the syntagmatic density of semantic information previously used in (*9*) suggests that the second-order entropy is a relevant proxy of ID.

Last, for each data point (i.e., one text read by one speaker), we computed the Shannon IR (ShIR) and conditional IR (for short, IR) as

$$\mathrm{ShIR} = \mathrm{ShE} \cdot \mathrm{SR} \quad (5)$$

$$\mathrm{IR} = ID \cdot SR \quad (6)$$

ShE and ShIR offer approximations of upper boundaries for each language since they do not take any context into account, despite the large amount of redundancy and dependency induced by morphology in human languages. For this reason, we reported here only the results from conditional entropy *ID* and conditional IR, since they provide much better estimations of the actual information transmitted during speech communication.

Statistical analysis

We describe below the various statistical analyses we performed, each under a dedicated subheading.

Intraclass correlations for text, language, and speaker

A priori, we expected that productions (respectively) of the same text, in the same language, or by the same speaker are not independent, which means that we should model them as random effects (*46*). As a preliminary analysis, we first estimated how much of the variation is explained by each of these factors (i.e., how similar their productions are) using linear mixed models (LMMs; as implemented by R's lmer() function in the lme4 package) to compute the linear intraclass correlations for the random effects text (representing the 15 short texts that were read by the speakers), language (representing the 17 languages) embedded within family (there are 9 language families, each language belonging to a unique family), and speaker (representing the 170 speakers), separately for the dependent variables NS, SR, and IR. For SR and IR, we used the AIC and the BIC to select the best fitting LMM, which, in both cases, included sex as a fixed effect and all three random effects; using R's lmer notation:

lmer(SR ~ 1 + Sex + (1 | Text) + (1 | Family/Language) + (1 | Speaker))

lmer(IR ~ 1 + Sex + (1 | Text) + (1 | Family/Language) + (1 | Speaker))

Generalized additive models for location, scale, and shape

While the fit obtained using LMMs is relatively good, the inspection of the diagnostic plots revealed slightly but potentially relevant deviations from the assumptions of this class of models (*46*), prompting us to use GAMLSS (*20*), as implemented by R's gamlss() function in package gamlss. This class of models is much more flexible than LMMs, allowing us to model not only the mean (location) but also the variance (scale) and the shape of the distribution, and, while relatively recent, it has already been successfully applied to problems in the language sciences (*21*).

More precisely, to preserve simplicity and interpretability, we compared Gaussian distributions with fixed versus modeled SD σ (both model the mean μ). To select among alternative models, we used AIC. For both SR and IR, modeling SD (with all three random effects and sex as fixed effect) results in a better fit to the data. For these models, the link function of the mean is the identity, but for the SD, it is the natural logarithm to ensure that the predicted values are always positive. In gamlss notation, the models are as follows:

gamlss(formula = SR ~ 1 + Sex + random(Text) + random (Language) + random(Family) + random(Speaker), sigma.formula = ~1 + Sex + random(Text) + random(Language) + random(Family) + random(Speaker), family = NO)

gamlss(formula = IR ~ 1 + Sex + random(Text) + random (Language) + random(Family) + random(Speaker), sigma.formula = ~1 + Sex + random(Text) + random(Family) + random(Language) + random(Speaker), family = NO)

We also modeled the relationship between *SR* and *ID* using gamlss as follows (here, language is meaningless as a random effect as there is only one *ID* value per language, but family is still potentially meaningful):

gamlss(formula = SR ~ 1 + ID + Sex + random(Text) + random (Speaker) + random(Family), sigma.formula = ~1 + ID + Sex + random(Text) + random(Speaker) + random(Family), family = NO).

Unimodality

Although *SR* and *IR* distributions are composite, the differences induced by the underlying groups (sex, language, text, and speaker) are compatible with overall unimodal distributions (*30*). To judge this unimodality, we also used two quantitative approaches, in addition to the visual inspection of the histograms. In the first approach, we fitted

Gaussian mixtures to the actual distributions, assessing between one and five Gaussian components, using the gamlssMX() function in the gamlss.mx package [for example, for *SR* with one component: gamlssMX(formula = SR ~ 1, family = NO, K = 1)], and we used AIC to select the best fitting mixtures. However, despite being simple, modeling our distributions with a mixture of Gaussians might not sufficiently capture how "unimodal" these distributions are because the model might need more than one component to fit, for example, a leptokurtic distribution. Mixtures of distributions other than Gaussian could be considered here, but we lack relevant arguments for choosing one distribution over another.

The second approach uses three unimodality tests: the Silverman test, the dip test, and the bimodality coefficient (BC). The Silverman test (*47*) tests the null hypothesis that an underlying density has at most *k* modes; its null hypothesis is that the underlying density has at most *k* modes (H0: number of modes $\leq k$), and the result is the bootstrapped *P* value for rejecting a unimodal distribution (our implementation is based on the code available at www.mathematik.uni-marburg.de/~stochastik/R_packages). The dip test computes Hartigans' dip statistic D (*48*) and the associated (interpolated) *P* value for rejecting a unimodal distribution (as implemented in the R package diptest). The BC is based on an empirical relationship between bimodality and the third and fourth statistical moments of a distribution (skewness and kurtosis) and is proportional to the division of squared skewness by uncorrected kurtosis. The underlying logic is that a bimodal distribution will have very low kurtosis, an asymmetric character, or both, all of which increase BC, with values exceeding 0.555 (the value representing a uniform distribution) suggesting bimodality. We implemented it as $BC = (s2 + 1)/(k + 3 \cdot ((n - 1)2 / ((n - 2) \cdot (n - 3))))$, following (*49*). Unfortunately, these tests tend to disagree (see analysis report file S1), and the problem of unimodality testing is far from settled (*49*). For all SR and IR and for each of these tests, we performed four randomization procedures to obtain an estimate of the "specialness" of the observed unimodality estimate: permutation model 1 (PM1), randomly permute the *SR* values freely among speakers, texts, and languages; PM2, randomly permute the *ID* values among languages; PM3, randomly permute the speaker average *SR* values among speakers (irrespective of language); and PM4, randomly permute the language average *SR* values among languages. Each of these procedures results in a distribution of expected test (or *P*) values that can be compared with the observed estimates actually obtained.

Pairwise distances between languages

NS, *SR*, and *IR* each have a distribution of values within a given language; thus, differences in the distribution of any of these three variables can be computed for any language pair. For each of the three variables, we computed all the possible $17 \cdot (17 - 1)/2 = 136$ differences between the pairs of languages, and we compared these distributions using paired permutation *t* tests (with 1000 permutations). We used five methods for computing these differences [all implemented by R's function distance() in package philentropy]: Hellinger distance, Jensen-Shannon divergence, Kolmogorov-Smirnov distance, Kullback-Leibler divergence, and chi-square divergence, and we found that all strongly agree.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/9/eaaw2594/DC1

Text S1. Relationship between semantic and encoding levels of information.
Text S2. Relationship between canonical SR and automatically estimated SR.
Text S3. Multilingual parallel corpus.
Fig. S1. Two different encoding strategies that convey the same semantic information.
Fig. S2. Automatic versus canonical SRs per language.
Table S1. The 17 languages used in this study.
Table S2. The written corpora.
Data file S1. Raw data (as a TAB-separated CSV file) used for the analyses.
Data file S2. Raw data (as a TAB-separated CSV file) used for the analyses.
Analysis report file S1. Full analysis report and details about the data in HTML format.
Analysis script file S1. The RMarkdown script continuing all the R code needed to reproduce the results and plots reported here.
References (*50*–*59*)

REFERENCES AND NOTES

1. S. C. Levinson, Turn-taking in human communication – Origins and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14 (2016).
2. M. Dingemanse, S. G. Roberts, J. Baranova, J. Blythe, P. Drew, S. Floyd, R. S. Gisladottir, K. H. Kendrick, S. C. Levinson, E. Manrique, G. Rossi, N. J. Enfield, Universal principles in the repair of communication problems. *PLOS ONE* **10**, e0136100 (2015).
3. N. Evans, S. C. Levinson, The myth of language universals: Language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492 (2009).
4. D. Dediu, S. C. Levinson, Neanderthal language revisited: Not only us. *Curr. Opin. Behav. Sci.* **21**, 49–55 (2018).
5. J. Odling-Smee, K. N. Laland, Cultural niche construction: Evolution's cradle of language, in *The Prehistory of Language* (Oxford Univ. Press, 2009), pp. 99–121.
6. T. F. Jaeger, Redundancy and reduction: Speakers manage syntactic information density. *Cognit. Psychol.* **61**, 23–62 (2010).
7. M. Aylett, A. Turk, The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech.* **47**, 31–56 (2004).
8. A. Fenk, G. Fenk-Oczlon, Menzerath's law and the constant flow of linguistic information, in *Contributions to Quantitative Linguistics: Proceedings of the First International Conference on Quantitative Linguistics, QUALICO, Trier, 1991*, R. Köhler, B. B. Rieger, Eds. (Springer, 1993), pp. 11–31.
9. F. Pellegrino, C. Coupé, E. Marsico, A cross-language perspective on speech information rate. *Language* **87**, 539–558 (2011).
10. E. D. Casserly, D. B. Pisoni, Speech perception and production. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 629–647 (2010).
11. L. M. Hyman, Does Gokana really have no syllables? Or: What's so great about being universal? *Phonology* **28**, 55–85 (2011).
12. S. Greenberg, Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Commun.* **29**, 159–176 (1999).
13. J. E. Peelle, M. H. Davis, Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* **3**, 320 (2012).
14. A.-L. Giraud, D. Poeppel, Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).
15. O. Ghitza, Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Front. Psychol.* **5**, 652 (2014).
16. M. F. Assaneo, D. Poeppel, The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Sci. Adv.* **4**, eaao3842 (2018).
17. Y.-C. Tsao, G. Weismer, Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *J. Speech Lang. Hear. Res.* **40**, 858–866 (1997).
18. T. F. Jaeger, E. Buz, Signal reduction and linguistic encoding, in *The Handbook of Psycholinguistics*, E. M. Fernández, H. S. Cairns, Eds. (John Wiley & Sons Ltd., 2017), pp. 38–81.
19. J. Koreman, Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *J. Acoust. Soc. Am.* **119**, 582–596 (2006).
20. M. D. Stasinopoulos, R. A. Rigby, G. Z. Heller, V. Voudouris, F. De Bastiani, *Flexible Regression and Smoothing: Using GAMLSS in R* (CRC Press, 2017).
21. C. Coupé, Modeling linguistic variables with regression models: Addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Front. Psychol.* **9**, 513 (2018).
22. E. Jacewicz, R. A. Fox, C. O'Neill, J. Salmons, Articulation rate across dialect, age, and gender. *Lang. Var. Change.* **21**, 233–256 (2009).
23. P. Wagner, J. Trouvain, F. Zimmerer, In defense of stylistic diversity in speech research. *J. Phon.* **48**, 1–12 (2015).
24. U. C. Priva, Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* **160**, 27–34 (2017).

25. H. Traunmüller, A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults" (Department of Linguistics, University of Stockholm, 1994); https://pdfs.semanticscholar.org/aa8b/acb5e7843740fbea24742c3046fbcc009a49.pdf.
26. J. M. Perkins, S. V. Subramanian, G. D. Smith, E. Özaltin, Adult height, nutrition, and population health. *Nutr. Rev.* **74**, 149–165 (2016).
27. F. Debruyne, W. Decoster, A. Van Gijsel, J. Vercammen, Speaking fundamental frequency in monozygotic and dizygotic twins. *J. Voice* **16**, 466–471 (2002).
28. E. Marouli, E. Marouli, M. Graff, C. Medina-Gomez, K. S. Lo, A. R. Wood, T. R. Kjaer, R. S. Fine, Y. Lu, C. Schurmann, H. M. Highland, S. Rüeger, G. Thorleifsson, A. E. Justice, D. Lamparter, K. E. Stirrups, V. Turcot, K. L. Young, T. W. Winkler, T. Esko, T. Karaderi, A. E. Locke, N. G. D. Masca, M. C. Y. Ng, P. Mudgal, M. A. Rivas, S. Vedantam, A. Mahajan, X. Guo, G. Abecasis, K. K. Aben, L. S. Adair, D. S. Alam, E. Albrecht, K. H. Allin, M. Allison, P. Amouyel, E. V. Appel, D. Arveiler, F. W. Asselbergs, P. L. Auer, B. Balkau, B. Banas, L. E. Bang, M. Benn, S. Bergmann, L. F. Bielak, M. Blüher, H. Boeing, E. Boerwinkle, C. A. Böger, L. L. Bonnycastle, J. Bork-Jensen, M. L. Bots, E. P. Bottinger, D. W. Bowden, I. Brandslund, G. Breen, M. H. Brilliant, L. Broer, A. A. Burt, A. S. Butterworth, D. J. Carey, M. J. Caulfield, J. C. Chambers, D. I. Chasman, Y.-D. I. Chen, R. Chowdhury, C. Christensen, A. Y. Chu, M. Cocca, F. S. Collins, J. P. Cook, J. Corley, J. C. Galbany, A. J. Cox, G. Cuellar-Partida, J. Danesh, G. Davies, P. I. W. de Bakker, G. J. de Borst, S. de Denus, M. C. H. de Groot, R. de Mutsert, I. J. Deary, G. Dedoussis, E. W. Demerath, A. I. den Hollander, J. G. Dennis, E. Di Angelantonio, F. Drenos, M. Du, A. M. Dunning, D. F. Easton, T. Ebeling, T. L. Edwards, P. T. Ellinor, P. Elliott, E. Evangelou, A.-E. Farmaki, J. D. Faul, M. F. Feitosa, S. Feng, E. Ferrannini, M. M. Ferrario, J. Ferrieres, J. C. Florez, I. Ford, M. Fornage, P. W. Franks, R. Frikke-Schmidt, T. E. Galesloot, W. Gan, I. Gandin, P. Gasparini, V. Giedraitis, A. Giri, G. Girotto, S. D. Gordon, P. Gordon-Larsen, M. Gorski, N. Grarup, M. L. Grove, V. Gudnason, S. Gustafsson, T. Hansen, K. M. Harris, T. B. Harris, A. T. Hattersley, C. Hayward, L. He, I. M. Heid, K. Heikkilä, Ø. Helgeland, J. Hernesniemi, A. W. Hewitt, L. J. Hocking, M. Hollensted, O. L. Holmen, G. K. Hovingh, J. M. M. Howson, C. B. Hoyng, P. L. Huang, K. Hveem, M. A. Ikram, E. Ingelsson, A. U. Jackson, J.-H. Jansson, G. P. Jarvik, G. B. Jensen, M. A. Jhun, Y. Jia, X. Jiang, S. Johansson, M. E. Jørgensen, T. Jørgensen, P. Jousilahti, J. W. Jukema, B. Kahali, R. S. Kahn, M. Kähönen, P. R. Kamstrup, S. Kanoni, J. Kaprio, M. Karaleftheri, S. L. R. Kardia, F. Karpe, F. Kee, R. Keeman, L. A. Kiemeney, H. Kitajima, K. B. Kluivers, T. Kocher, P. Komulainen, J. Kontto, J. S. Kooner, C. Kooperberg, P. Kovacs, J. Kriebel, H. Kuivaniemi, S. Küry, J. Kuusisto, M. La Bianca, M. Laakso, T. A. Lakka, E. M. Lange, L. A. Lange, C. D. Langefeld, C. Langenberg, E. B. Larson, I.-T. Lee, T. Lehtimäki, C. E. Lewis, H. Li, J. Li, R. Li-Gao, H. Lin, L.-A. Lin, X. Lin, L. Lind, J. Lindström, A. Linneberg, Y. Liu, Y. Liu, A. Lophatananon, J. Luan, S. A. Lubitz, L.-P. Lyytikäinen, D. A. Mackey, P. A. F. Madden, A. K. Manning, S. Männistö, G. Marenne, J. Marten, N. G. Martin, A. L. Mazul, K. Meidtner, A. Metspalu, P. Mitchell, K. L. Mohlke, D. O. Mook-Kanamori, A. Morgan, A. D. Morris, A. P. Morris, M. Müller-Nurasyid, P. B. Munroe, M. A. Nalls, M. Nauck, C. P. Nelson, M. Neville, S. F. Nielsen, K. Nikus, P. R. Njølstad, B. G. Nordestgaard, I. Ntalla, J. R. O'Connel, H. Oksa, L. M. O. Loohuis, R. A. Ophoff, K. R. Owen, C. J. Packard, S. Padmanabhan, C. N. A. Palmer, G. Pasterkamp, A. P. Patel, A. Pattie, O. Pedersen, P. L. Peissig, G. M. Peloso, C. E. Pennell, M. Perola, J. A. Perry, J. R. B. Perry, T. N. Person, A. Pirie, O. Polasek, D. Posthuma, O. T. Raitakari, A. Rasheed, R. Rauramaa, D. F. Reilly, A. P. Reiner, F. Renström, P. M. Ridker, J. D. Rioux, N. Robertson, A. Robino, O. Rolandsson, I. Rudan, K. S. Ruth, D. Saleheen, V. Salomaa, N. J. Samani, K. Sandow, Y. Sapkota, N. Sattar, M. K. Schmidt, P. J. Schreiner, M. B. Schulze, R. A. Scott, M. P. Segura-Lepe, S. Shah, X. Sim, S. Sivapalaratnam, K. S. Small, A. V. Smith, J. A. Smith, L. Southam, T. D. Spector, E. K. Speliotes, J. M. Starr, V. Steinthorsdottir, H. M. Stringham, M. Stumvoll, P. Surendran, L. M. 't Hart, K. E. Tansey, J.-C. Tardif, K. D. Taylor, A. Teumer, D. J. Thompson, U. Thorsteinsdottir, B. H. Thuesen, A. Tönjes, G. Tromp, S. Trompet, E. Tsafantakis, J. Tuomilehto, A. Tybjaerg-Hansen, J. P. Tyrer, R. Uher, A. G. Uitterlinden, S. Ulivi, S. W. van der Laan, A. R. Van Der Leij, C. M. van Duijn, N. M. van Schoor, J. van Setten, A. Varbo, T. V. Varga, R. Varma, D. R. V. Edwards, S. H. Vermeulen, H. Vestergaard, V. Vitart, T. F. Vogt, D. Vozzi, M. Walker, F. Wang, C. A. Wang, S. Wang, Y. Wang, N. J. Wareham, H. R. Warren, J. Wessel, S. M. Willems, J. G. Wilson, D. R. Witte, M. O. Woods, Y. Wu, H. Yaghootkar, J. Yao, P. Yao, L. M. Yerges-Armstrong, R. Young, E. Zeggini, X. Zhan, W. Zhang, J. H. Zhao, W. Zhao, W. Zhao, H. Zheng, W. Zhou, The EPIC-InterAct Consortium, CHD Exome+ Consortium, ExomeBP Consortium, T2D-Genes Consortium, GoT2D Genes Consortium, Global Lipids Genetics Consortium, ReproGen Consortium, MAGIC Investigators, J. I. Rotter, M. Boehnke, S. Kathiresan, M. I. McCarthy, C. J. Willer, K. Stefansson, I. B. Borecki, D. J. Liu, K. E. North, N. L. Heard-Costa, T. H. Pers, C. M. Lindgren, C. Oxvig, Z. Kutalik, F. Rivadeneira, R. J. F. Loos, T. M. Frayling, J. N. Hirschhorn, P. Deloukas, G. Lettre, Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
29. M. Ordin, I. Mennen, Cross-linguistic differences in bilinguals' fundamental frequency ranges. *J. Speech Lang. Hear. Res.* **60**, 1493–1506 (2017).
30. M. F. Schilling, A. E. Watkins, W. Watkins, Is human height bimodal? *Am. Stat.* **56**, 223–229 (2002).
31. C. M. Reed, N. I. Durlach, Note on information transfer rates in human communication. *Presence Teleoperators Virtual Environ.* **7**, 509–518 (1998).
32. J. Villasenor, Y. Han, D. Wen, E. Gonzales, J. Chen, J. Wen, The information rate of modern speech and its implications for language evolution, in *Evolution Of Language, The - Proceedings Of The 9th International Conference (Evolang9)*, T. Scott-Phillips, M. Tamariz, E. A. Cartmill, J. R. Hurford, Eds. (World Scientific, 2012), pp. 376–383.
33. H. Quené, Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *J. Acoust. Soc. Am.* **123**, 1104–1113 (2008).
34. G. E. Loeb, Optimal isn't good enough. *Biol. Cybern.* **106**, 757–765 (2012).
35. Y. Jadoul, A. Ravignani, B. Thompson, P. Filippi, B. de Boer, Seeking temporal predictability in speech: Comparing statistical approaches on 18 world languages. *Front. Hum. Neurosci* **10**, 586 (2016).
36. W. Bisang, Overt and hidden complexity – Two types of complexity and their implications. *Poznan Stud. Contemp. Linguist.* **50**, 127–143 (2014).
37. C. Sinha, Language and other artifacts: Socio-cultural dynamics of niche construction. *Front. Psychol.* **6**, 1601 (2015).
38. C. Everett, D. E. Blasí, S. G. Roberts, Language evolution and climate: The case of desiccation and tone. *J. Lang. Evol.* **1**, 33–46 (2016).
39. D. Dediu, R. Janssen, S. R. Moisik, Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Lang. Commun.* **54**, 9–20 (2017).
40. G. Lupyan, R. Dale, Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cogn. Sci.* **20**, 649–660 (2016).
41. D. Dediu, M. Cysouw, S. C. Levinson, A. Baronchelli, M. H. Christiansen, W. A. Croft, N. J. Evans, S. Garrod, R. D. Gray, A. Kandler, E. Leiven, Cultural evolution of language, in *Cultural Evolution: Society, Technology, Language, and Religion*, P. J. Richerson, M. H. Christiansen, Eds. (MIT Press, 2013), vol. 12, pp. 303–332.
42. P. Ladefoged, Articulatory features for describing lexical distinctions. *Language* **83**, 161–180 (2007).
43. E. Campione, J. Véronis, A statistical study of pitch target points in five languages, in *Fifth International Conference on Spoken Language Processing* (International Speech Communicaton Association, 1998), pp. 3163–3166.
44. E. Ferragne, S. Flavier, C. Fressard, in *Proceedings of 14th Interspeech Conference* (International Speech Communicaton Association, 2013), pp. 1864–1865.
45. Y. M. Oh, thesis, Université de Lyon, France (2015).
46. A. Gelman, J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, ed. 1, 2006).
47. P. Hall, M. York, On the calibration of Silverman's test for multimodality. *Stat. Sin.* **11**, 515–536 (2001).
48. J. A. Hartigan, P. M. Hartigan, The dip test of unimodality. *Ann. Stat.* **13**, 70–84 (1985).
49. J. B. Freeman, R. Dale, Assessing bimodality to detect the presence of a dual cognitive process. *Behav. Res. Methods.* **45**, 83–97 (2013).
50. Z. Malisz, E. Brandt, B. Möbius, Y. M. Oh, B. Andreeva, Dimensions of segmental variability: Interaction of prosody and surprisal in six languages. *Front. Commun.* **3**, 25 (2018).
51. N. H. de Jong, T. Wempe, Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. Methods* **41**, 385–390 (2009).
52. K. Maekawa, H. Kikuchi, Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report, in *Voicing in Japanese*, J. van de Weijer, K. Nanjo, T. Nishihara, Eds. (De Gruyter Mouton, 2005), pp. 205–228.
53. I. Maddieson, S. Flavier, E. Marsico, C. Coupé, F. Pellegrino, LAPSyd: Lyon-Albuquerque phonological systems database, in *Proceedings of 14th Interspeech Conference* (International Speech Communicaton Association, 2013), pp. 3022–3026.
54. M. Perea, M. Urkia, C. J. Davis, A. Agirre, E. Laseka, M. Carreiras, E-Hitz: A word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behav. Res. Methods* **38**, 610–615 (2006).
55. A. C. Chin, New resources for cantonese language studies: A linguistic corpus of mid-20th century Hong Kong cantonese. *Curr. Res. Chin. Linguist.* **92**, 7–16 (2013).
56. B. New, C. Pallier, L. Ferrand, R. Matos, Une base de données lexicales du français contemporain sur internet : LEXIQUE™//A lexical database for contemporary french : LEXIQUE™. *Année Psychol.* **101**, 447–462 (2001).
57. V. Lyding, E. Stemle, C. Borghetti, M. Brunello, S. Castagnoli, F. Dell'Orletta, H. Dittmann, A. Lenci, V. Pirrelli, The PAISÁ corpus of Italian web texts, in *Proceedings of the 9th Web as Corpus Workshop (WaC-9)* (Association for Computational Linguistics, 2014), pp. 36–43.
58. S. Sharoff, Open-source corpora: Using the net to fish for linguistic data. *Int. J. Corpus Linguist.* **11.4**, 435–462 (2006).
59. V.-B. Le, D.-D. Tran, E. Castelli, L. Besacier, J.-F. Serignat, Spoken and written language resources for Vietnamese, in *Proceedings of LREC4* (European Language Resources Association, 2004), pp. 599–602.

Acknowledgments: We thank D. E. Blasi for suggestions and feedback on the statistical analysis and on previous versions of this paper. We also thank E. Castelli for help with collecting the Vietnamese data. **Funding:** D.D. was funded by a European Institutes for Advanced Study (EURIAS) Fellowship 2017–2018 and by an IDEXLYON (16-IDEX-0005) Fellowship grant (2018–2021). C.C., Y.M.O., and F.P. were funded by LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon within the French program Investissements d'Avenir

program (ANR-11-IDEX-0007) operated by the National Research Agency (ANR). **Author contributions:** C.C., Y.O., and F.P. jointly conceived the study. C.C. and Y.O. collected the data. C.C., Y.O., D.D., and F.P. analyzed the data, produced and discussed the results, and wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The primary data are also available in the dedicated GitHub repository https://github.com/keruiduo/SupplMatInfoRate.

Submitted 21 December 2018
Accepted 2 August 2019
Published 4 September 2019
10.1126/sciadv.aaw2594

Citation: C. Coupé, Y. M. Oh, D. Dediu, F. Pellegrino, Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Sci. Adv.* **5**, eaaw2594 (2019).