

UC Berkeley

UC Berkeley Previously Published Works

Title

The forms and meanings of grammatical markers support efficient communication

Permalink

<https://escholarship.org/uc/item/9zx351pm>

Journal

Proceedings of the National Academy of Sciences of the United States of America,
118(49)

ISSN

0027-8424

Authors

Mollica, Francis
Bacon, Geoff
Zaslavsky, Noga
et al.

Publication Date

2021-12-07

DOI

10.1073/pnas.2025993118

Peer reviewed



The forms and meanings of grammatical markers support efficient communication

Francis Mollica^{a,1}, Geoff Bacon^b, Noga Zaslavsky^{c,d}, Yang Xu^e, Terry Regier^{b,f}, and Charles Kemp^g

^aInstitute for Language, Cognition and Computation, School of Informatics, University of Edinburgh, Edinburgh EH8 9LE, United Kingdom; ^bDepartment of Linguistics, University of California, Berkeley, CA 94720; ^cDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ^dCenter for Brains, Minds, and Machines, Massachusetts Institute of Technology, Cambridge, MA 02139; ^eDepartment of Computer Science, Cognitive Science Program, University of Toronto, Toronto, ON M5S 3G8, Canada; ^fCognitive Science Program, University of California, Berkeley, CA 94720; and ^gSchool of Psychological Sciences, University of Melbourne, Parkville, VIC 3010, Australia

Edited by William Croft, The University of New Mexico, Albuquerque, NM, and accepted by the Editorial Board October 18, 2021 (received for review March 23, 2021)

Functionalist accounts of language suggest that forms are paired with meanings in ways that support efficient communication. Previous work on grammatical marking suggests that word forms have lengths that enable efficient production, and work on the semantic typology of the lexicon suggests that word meanings represent efficient partitions of semantic space. Here we establish a theoretical link between these two lines of work and present an information-theoretic analysis that captures how communicative pressures influence both form and meaning. We apply our approach to the grammatical features of number, tense, and evidentiality and show that the approach explains both which systems of feature values are attested across languages and the relative lengths of the forms for those feature values. Our approach shows that general information-theoretic principles can capture variation in both form and meaning across languages.

communicative efficiency | grammatical features | linguistic typology | information theory

A primary goal of linguistic typology is to characterize and explain the diversity in extant linguistic systems compared to possible but unattested systems (1). Linguistic typology can be approached from a variety of perspectives (e.g., ref. 2), but here we take a functional approach and build on a large body of work that has explored ways in which language supports efficient communication (3–7). Recent work in this tradition has formalized communicative efficiency in terms of information theory and has used this formalization to demonstrate that linguistic forms and meanings support efficient communication (8), but form and meaning are usually treated separately. On one hand, a substantial body of work has demonstrated that linguistic forms allow communication with a minimum of effort (4, 9–11), but this work typically does not explain the meanings associated with the forms in question. On the other hand, recent work in semantic typology has shown that word meanings within several semantic domains support efficient communication (12, 13) but has not addressed the forms used to express these meanings. Here we show that an existing information-theoretic account of lexical semantics (ref. 13, as formulated in ref. 14) also accounts for classic ideas about coding efficiency from the literature on grammatical marking (9, 11). Connecting these lines of work illustrates how information theory provides a unified account of both the meanings encoded in natural language and the forms used to express them.

Our theoretical framework applies to both grammar and the lexicon, but we focus here on grammatical marking expressed by morphology and in particular on the grammatical features of number, tense, and evidentiality. We chose these features because they primarily convey semantic information and because each encodes a rich semantic dimension instead of a simple binary distinction. Number reflects the number of entities involved in one role of an event (e.g., four lions chasing a giraffe). Tense refers to the location of an event in time (e.g., past, present, and

future). Evidentiality refers to the source of information (e.g., did the speaker see it, or hear someone else describe it?). Grammatical features like these are core components of language, yet there is considerable variation in the size of grammatical feature inventories and the realization of grammatical features across languages. For example, the data analyzed in this paper include 15 distinct morphological systems that languages use to mark grammatical number. Whereas English only distinguishes between singular and plural, Larike distinguishes between singular, dual, trial, and plural (15). Accounting for the diversity of feature inventories and realizations across languages is therefore a significant challenge.

Our work builds on functionalist accounts of grammatical features from several areas of the literature. A long-standing line of work has used corpus analyses to show that the realizations of grammatical feature values are shaped by the principle of least effort (4). Because speakers often need to convey the meanings associated with grammatical features, grammatical markers have short forms that are easy to produce, and the most frequent feature values may receive no overt marking (9, 11, 16). A second line of work has used artificial language learning experiments and evolutionary models to demonstrate that learners restructure their input to produce systems that are simpler, easier to produce, and more informative (17–20), often in line with linguistic universals (e.g., ref. 21). Related work has also demonstrated that more easily acquired grammatical systems occur more frequently in the world's languages (22, 23). Our approach is broadly consistent

Significance

Grammatical marking of features such as number, tense, and evidentiality varies widely across languages. Despite this variation, we show that grammatical markers support efficient information transfer from speakers to listeners. We apply a formal model of communication to data from dozens of languages and find that grammatical marking achieves a near-optimal balance between maximizing informativeness and minimizing code lengths. Our approach shows how general information-theoretic principles can capture variation in both form and meaning across languages.

Author contributions: F.M., G.B., Y.X., T.R., and C.K. designed research; F.M., G.B., Y.X., T.R., and C.K. performed research; N.Z. contributed new reagents/analytic tools; F.M., G.B., and C.K. analyzed data; and F.M., G.B., N.Z., Y.X., T.R., and C.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. W.C. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: f.mollica@ed.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2025993118/-DCSupplemental>.

Published December 3, 2021.

with all of these research strands but formally bridges them by providing an integrated account of both grammatical feature values and the forms used to express them.

The information-theoretic framework that we use formalizes the trade-off between informativeness and simplicity that languages must negotiate. Consider a speaker who wishes to convey some meaning (e.g., the number of empty coffee cups still on their desk) to a hearer (Fig. 1, *Top*). A highly informative system allows the speaker to discriminate between many different meanings (e.g., many different numbers of cups), but this communicative precision can only be achieved if the system is far from simple. The trade-off between informativeness and simplicity has been discussed for many years in the literature on competing motivations (3, 24, 25), and several measures of morphological simplicity have been proposed (see *SI Appendix* for a discussion). Here we build on a recent account of lexical semantics (14, 26) that is grounded in rate–distortion theory (27, 28) (the branch of information theory characterizing efficient data compression) and that formalizes both informativeness and simplicity in information-theoretic terms. Within this framework, the simplicity dimension connects naturally with the notion of coding efficiency from the literature on grammatical marking (6, 16). We will therefore argue that the trade-off between informativeness and simplicity helps to explain both which feature values are attested across languages and the relative lengths of the linguistic realizations of these feature values.

The next section introduces our theoretical framework and provides formal definitions of information loss (the inverse of

informativeness) and complexity (the inverse of simplicity). We then provide an overview of number, tense, and evidentiality across languages and introduce the typological data that we analyze. The first set of analyses focuses on meaning and demonstrates that grammatical feature inventories achieve near-optimal trade-offs between informativeness and simplicity. The second set of analyses focuses on form and demonstrates that the realizations of grammatical features enable concise communication.

Theoretical Framework

We build on the theoretical framework in ref. 14, which has been previously used to account for word meanings across languages, and show that the same framework can also be linked to aspects of linguistic form. The framework, illustrated in Fig. 1, *Top*, assumes a speaker and a listener who wish to communicate about states of the world u drawn from the universe \mathcal{U} . The speaker is uncertain about the true state of the world, and their mental state is captured by a speaker distribution s over states in \mathcal{U} . To summarize this mental state the speaker generates a linguistic form f according to an encoder $q(f|s)$, which maps speaker distributions into forms. Upon receiving this form, the listener computes a distribution \hat{s} that is intended to approximate the speaker distribution s . We assume that this distribution \hat{s} is computed by carrying out Bayesian inference based on the encoder $q(f|s)$ and a prior $p(s)$ over speaker distributions, which gives the optimal \hat{s} (14). The prior reflects communicative need

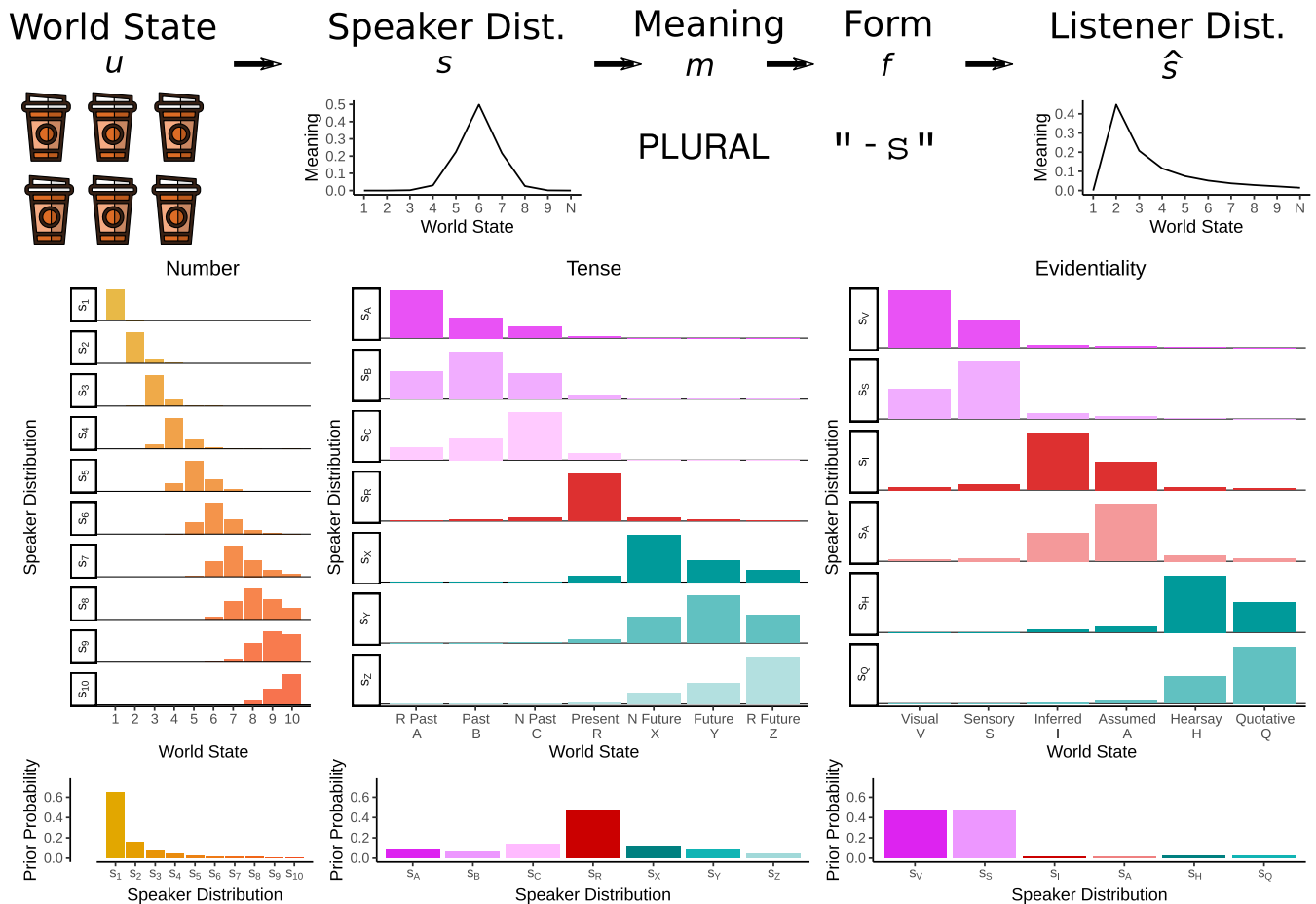


Fig. 1. Communicative scenario along with speaker distributions and priors for number, tense, and evidentiality. (*Top*) Communicative scenario illustrating how a speaker generates a form which is then used by a listener to reconstruct the speaker distribution s over world states. In reality the form would not be uttered in isolation but rather combined with the noun “cup” to generate the utterance “cups.” (*Center*) Speaker distributions s_u for number, tense, and evidentiality. (*Bottom*) Priors $p(s)$ on the three sets of speaker distributions.

or the relative frequency with which speakers communicate about different states of the world (8, 29, 30).

An optimal encoder $q(f|s)$ should satisfy two criteria: it should allow the listener to accurately reconstruct the speaker's mental state, and it should minimize production effort by ensuring that frequently used forms are short. To formalize these criteria it will be convenient to represent a grammatical marker as a pair (m, f) that includes both a meaning (or feature value) m and a form (or realization) f . For number in English, there are two such pairs: (SINGULAR, \emptyset) and (PLURAL, “-s”), where the empty set \emptyset indicates that the singular is zero-marked, or realized without an overt form. Given this representation we can decompose the encoder $q(f|s)$ into a meaning encoder $q_m(m|s)$ that maps speaker distributions into meanings and a form encoder $q_f(f|m)$ that maps meanings into forms. This two-stage encoding process is illustrated in Fig. 1 and is used in the following sections to develop analyses that focus on efficiency of meaning and analyses that focus on efficiency of form.

The meaning encoder q_m is lossy, but for simplicity we assume that the form encoder q_f is lossless, which means that the listener is able to reconstruct the meaning m without error given the form f . In reality this assumption does not hold. Languages permit ambiguity in the linguistic signal, and ambiguity (including ambiguity arising from reanalysis) may have implications for the historical emergence of grammatical forms (31). Assuming that q_f is lossless, however, is a natural starting point given the cross-linguistic data available to us.

Efficiency of Meaning. An efficient encoder q_m achieves an optimal trade-off between complexity and information loss (the inverse of informativeness). Following ref. 14, the formal definitions of complexity and information loss are grounded in the information bottleneck (IB) principle (32), which is a special type of a rate–distortion trade-off. The complexity of an encoder measures how much information about the speaker's mental state is preserved in the meaning of a grammatical marker and is defined as the mutual information between meanings and speaker distributions:

$$I(M; S) = H(M) - H(M|S), \quad [1]$$

which, as shown, can be formulated as the difference of the entropy over meanings $H(M)$ and the conditional entropy $H(M|S)$.

The informativeness of an encoder is negatively related to the expected information loss associated with each communicative interaction. Following refs. 12 and 14, we define this information loss as the expected Kullback–Leibler divergence $\text{KL}(s||\hat{s})$ between the speaker distribution s and the listener's reconstruction \hat{s} of that distribution:

$$E[\text{KL}[S||\hat{S}]] = \sum_{s, m} p(s) q_m(m|s) \text{KL}[s||\hat{s}_m], \quad [2]$$

where \hat{s}_m is the reconstructed distribution for occasions on which the speaker chooses meaning m .

Every possible mapping from speaker distributions to meanings corresponds to a point in a two-dimensional space where the dimensions represent complexity and information loss. Some points in this space cannot be achieved by any possible language—for example, in any realistic setting it is impossible for an encoder to achieve both zero complexity and zero information loss. The boundary separating achievable points from unachievable points is a special case of a Pareto frontier known as the IB theoretical limit, and encoders along this continuous frontier achieve optimal trade-offs between complexity and information loss. These encoders are optimal in the sense that complexity cannot be reduced without increasing information loss, and information loss cannot be reduced without increasing complexity.

Given this theoretical framework, we can ask whether attested grammatical feature inventories achieve near-optimal trade-offs between complexity and information loss. For any given feature, applying the framework requires three components to be specified: the universe of world states \mathcal{U} , the speaker distributions for each objective world state s_u , and the prior on speaker distributions $p(s)$. Given these components we can compute the complexity and information loss of both attested and hypothetical systems and trace out the IB Pareto frontier of systems that achieve optimal trade-offs between complexity and information loss (14, 32).

Efficiency of Form. We now consider the mapping q_f from meanings to forms, or strings of phonemes. Because this mapping is assumed to be lossless, efficiency is purely a matter of minimizing expected form length. The entropy $H(M)$ gives a lower bound on expected form length (27), and an efficient mapping q_f is expected to assign a form to meaning m that has length close to

$$h(m) = -\log p(m) = \log \sum_s q_m(m|s) p(s). \quad [3]$$

In reality, natural language mappings q_f are likely to yield expected code lengths that do not come especially close to the entropy $H(M)$ (33–35). These mappings, however, may nevertheless reflect a pressure toward brevity. Eq. 3 suggests that shorter forms should be used for more frequent meanings, and we will examine whether this inverse relationship between form length and frequency holds in our data.

Previous work on grammatical marking (11, 36) and the lexicon (4, 10, 37) has emphasized the notion of coding efficiency and has demonstrated that forms tend to be paired with meanings in ways that allow utterances to be relatively concise. Similar results have emerged from studies of phonetic realization (38), online word choice (39), and word ordering (40). Our theoretical approach is consistent with all of these results but goes beyond them by considering efficiency of form within a framework that also captures efficiency of meaning.

Considering efficiency of form and meaning within a single framework is important because the two are connected via the entropy $H(M)$. Minimizing the expected code length for an efficient code (i.e., minimizing $H(M)$) can only be achieved if the encoder q_m generates the same meaning for every speaker distribution. A system of this kind is maximally simple but also maximally uninformative (i.e., the information loss in Eq. 2 is maximized). The pressure toward minimizing code lengths must therefore trade off against a pressure toward informative communication (41). This trade-off is especially clear for deterministic encoders q_m , for which the complexity measure in Eq. 1 is equivalent to the entropy $H(M)$. Most (but not all) of the grammatical feature systems that we analyze are deterministic to a good first approximation: for example, in the English number system the singular is consistently used for individual items, and the plural is consistently used for multiple items.

Framework Instantiations

Now that we have introduced our general theoretical framework, we show how it can be applied to the grammatical features of number, tense, and evidentiality. To evaluate our theory we will make several simplifying assumptions (42, 43). First, while number, tense, and evidentiality can reflect multiple semantic dimensions (e.g., numerosity vs. individuality and absolute vs. relative time), we focus on a single grammaticalized semantic dimension for each. Each of these semantic dimensions can be expressed using a variety of strategies (e.g., numerals, adverbs, and verbal constructions), but for tractability we focus on systems that use obligatory morphological markers. As a result of these simplifications, the set of unique feature inventories in our

Table 1. Example inventories for grammatical number, tense, and evidentiality

	Language	System	Complexity	Information loss	Frontier distance	Form correlation
Number	Pirahã	(GENERAL, \emptyset)	0.00	1.44	0.00	NA
	Russian	(SG, \emptyset)(PL, "би")	0.94	0.55	0.01	1.00
	Larike	(SG, "mane") (DU _O , "matua") (TR _O , "matidu") (PL, "mati")	1.13	0.40	0.03	1.00
	Murrinh-Patha	(SG, "nukunu") (DU, "penintha") (PAUC, "peneme") (PL, "pigunu")	1.43	0.16	0.01	0.16
	Sursurunga	(SG, "i") (DU, "diar") (PAUC, "ditul") (GPAUC, "dihat") (PL, "di")	1.47	0.14	0.02	0.44
Tense	West Greenlandic	(ABCR, \emptyset) (XYZ, "ssa")	0.81	0.50	0.04	1.0
	Japanese	(ABC, "た") (RXYZ, \emptyset)	0.85	0.47	0.04	1.00
	Wolof	(ABC, "naa") (R, "nge") (XYZ, "dinaa")	1.52	0.08	0.00	1.00
	Hixkaryana	(A, "ye") (B, "yako") (C, "no") (RXYZ, "yaha")	1.26	0.42	0.21	0.32
	Zulu	(A, "a") (BC, "ile") (R, \emptyset) (X, "za") (YZ, "yaku")	2.01	0.02	0.00	0.88
Evidentiality	Sissala	(VSIA, \emptyset) (HQ, "ε")	0.28	0.15	0.00	1.0
	Abkhaz	(vs, \emptyset) (IAHQ, "заарен")	0.36	0.12	0.01	1.0
	Quechua	(vs, "mi") (IA, "chi") (HQ, "shi")	0.42	0.08	0.00	0.99
	Turkish	(V, \emptyset) (SIAHQ, "ims")	1.00	0.26	0.23	1.0
	Barasano	(V, "ka") (s, "ruyu") (IA, "ra") (HQ, "yu")	1.35	0.01	0.00	-0.09

Each system includes a single representative form for each meaning, and orthographic forms are shown instead of phonemic forms. The meanings for number are as follows: GENERAL, "the noun can be expressed without reference to number" (ref. 46, p. 10); SG, singular; PL, plural; DU, dual; TR, trial; PAUC, paucal (a few); GPAUC, greater paucal (a bunch); GPL, greater plural. Optional values are shown using the subscript "O." For tense, A, B, and C denote distant, near, and immediate past; R denotes present; and X, Y, and Z denote immediate, near, and remote future. For evidentiality, V and S denote visual and sensory, I and A denote inferred and assumed, and H and Q denote hearsay and quotative. Frontier distance shows the Euclidean distance between a system and the corresponding Pareto frontier in Fig. 2, and small values indicate efficiency of meaning. Form correlations show correlations between optimal and observed form lengths (Fig. 4), and large values indicate efficiency of form.

sample is relatively small, and many languages are coded using a single feature value that spans the entire dimension. This kind of inventory is maximally simple (and therefore technically optimal) but also maximally uninformative. The languages coded in this way may not make use of obligatory grammatical markers but typically rely on other linguistic constructions or contextual information for conveying information about the semantic dimensions in question. Evaluating the efficiency of these alternative communicative strategies is an important challenge for future work, and we return to this issue in the *Discussion*.

A second assumption is that the prior on speaker distributions $P(s)$ and the speaker distributions themselves are invariant across cultures. Previous studies have made similar assumptions (14, 44), and in all cases these assumptions should be viewed as rough first approximations that can be subsequently relaxed using data from studies that directly estimate culture-specific priors (45). Third, our operationalization of production effort is relatively coarse, and we treat this quantity as a binary variable (marker present vs. marker absent) or define it as the length of a marker's orthographic representation. Considering phonetic structure would allow for a more satisfying operationalization but is not possible given the data available to us. Finally, the grammatical systems considered in our analyses (including the examples in Table 1) are idealizations that are best treated as high-level summaries of a more complex reality. Within any individual language, there may be departures from our idealizations, and these differences may be irregular (e.g., the English plural is not marked for some nouns like "deer") or context-dependent (e.g., in Hunzib, evidentiality is marked only in the past tense).

The following sections introduce additional assumptions made when analyzing each of the three grammatical features and describe our samples of attested languages. These samples are drawn from a diverse set of language families and geographic regions but are convenience samples that aim for breadth of coverage rather than tight control over genealogical or geographic relationships. Given the nature of our analyses, controlling for historical descent and geographic region does not seem essential, and the more important question is the extent to which our

samples cover the space of attested feature inventories. Some extant inventories are almost certainly missing from our samples, and it will be valuable to revisit our analyses if and when larger datasets become available.

Number. Although the underlying semantic dimension for number is probably the natural numbers,* we consider only natural numbers less than or equal to 10 for simplicity. The universe \mathcal{U} therefore includes 10 world states, one for each number considered. Number marking in English distinguishes between singular (1) and plural (>1), but some languages have more precise systems. For example, Murrinh-Patha distinguishes between singular (1), dual (2), paucal (3 to 6), and plural (>5) (46). While English and Murrinh-Patha require a speaker to always use the most specific marker, some languages allow speakers a choice between specific and less specific markers. For example, Larike distinguishes between singular (1), optional dual (2), optional trial (3), and plural (>2) which means that the plural is always an alternative to the dual and to the trial (15).

When communicating about a state including n items, the speaker distribution s_n is intended to capture the speaker's uncertainty about the precise number of items present. Speaker distributions associated with the 10 possible world states are shown in Fig. 1, *Left Center*, and these distributions are based on data from a timed, high-contrast estimation task (48). The prior distribution $p(s)$ captures the relative frequencies with which speakers attempt to convey the 10 different meanings. Usage frequencies for number have been extensively studied using cross-linguistic corpora (49, 50), and these studies suggest that $p(s)$ can be roughly approximated as an inverse square law (Fig. 1, *Bottom Left*).

For our analysis, we compiled and coded the number marking inventories from 37 languages in ref. 46, representing 15 language families and 15 unique encoding systems. Five of these inventories are shown in Table 1. We adopt the standard linguistic conventions for number distinctions as labels (glossed in the

*We only consider numerical amounts in our analysis, leaving the dimension of individualization (47) for future work.

legend). The main challenge in coding number systems is that indeterminate meanings (paucal, 3 to 6; plural, >2; greater paucal, 6 to 8; greater plural, >9) vary slightly across languages. For example, we code plural in Murrinh-Patha as greater than 5 because there are nonoptional meanings for all lower numbers, whereas plural in English is greater than 2 because there is only one other meaning. When a distinction is optional, we assume that the speaker chooses the two possible meanings equally often (see *SI Appendix* for further details).

To study how meanings are realized as forms, we compiled number forms for a subset of 33 languages. Number is marked in a variety of ways across languages, and we considered only nominal and pronominal marking of grammatical number. For brevity, the systems in Table 1 show a single form for each meaning, but our dataset and analyses allow for multiple forms per meaning. For example, the Russian data include number forms for different combinations of case and gender.

Tense. Tense is analogous to number, but the underlying dimension is time rather than quantity. Tense marking in English distinguishes between past, present, and future, but some languages have more elaborate tense inventories that specify not only whether an event is in the past or future but also how far in the past or future it is. For example, Hixkaryana distinguishes between events in the immediate past (same day or previous night), near past (past few months), and remote past (51). Researchers in formal semantics and artificial intelligence have developed precise representations of tense that could potentially be used in frameworks like ours (52–55), but we take a simpler approach that can be readily applied across languages and is similar to that of ref. 56. We formulate \mathcal{U} as a set of seven temporal intervals: remote past (A), near past (B), immediate past (C), present (R), immediate future (X), near future (Y), and remote future (Z). These intervals are not sufficient to capture the tense inventory of every language in full: for example, Comrie (57) reports that Kiksht, a language of the US Pacific Northwest, distinguishes between six or seven past tense categories. Our seven-interval timeline is therefore a pragmatic choice that allows us to represent the tense inventories of many but not all of the languages of the world.[†]

As in our number analysis, we pair each element of \mathcal{U} with a speaker distribution s , and the seven meaning distributions are shown in Fig. 1, *Center*. These distributions are intended to capture the uncertainty that speakers maintain over the exact time of an event: for example, s_a captures uncertainty about an event that actually took place in the remote past. To formulate these distributions we postulate major boundaries between past, present, and future and minor boundaries between the three pasts (remote, near, and immediate) and between the three futures. The distributions are defined in terms of two parameters κ and λ that specify how sharply probability mass decreases across minor and major boundaries. We set $\kappa = 0.5$ and $\lambda = 0.1$, which means that distributions drop by factors of 2 and 10 across minor and major boundaries, respectively. Our results are qualitatively robust to variation in the speaker distributions as long as there is an appreciable decrease across minor boundaries ($\kappa \leq 0.75$) and a reasonable distinction between major and minor boundaries (κ and λ are not equivalent or near equivalent).

We estimated the prior $p(s)$ using a two-step process. In the first step we used estimates of past, present, and future from an analysis of social media (58). The resulting counts yield a distribution of [0.274, 0.475, 0.251] over the coarse categories of past, present, and future. Second, we used frequencies of temporal adverbs such as “yesterday,” “last week,” and “last month”

(see *SI Appendix, Table S1*, for the complete list) to distribute probability mass among the three levels of remoteness within both past and future categories. All frequencies were derived from the Google Ngram English corpus (59) for 1985, the year of publication for the source of many of our tense systems (60). The prior distribution resulting from the two-step process is shown in Fig. 1, *Bottom Center*.

For our analysis, we compiled tense inventories for 157 languages, representing 73 language families and 16 unique inventories. Our sample was largely taken from ref. 60. To study how meanings are realized as forms, we compiled forms for a subset of 33 languages. Languages were selected with a bias toward languages with more linguistic forms and toward grammars that made the relevant information especially clear.

The major challenge encountered in assembling the data is that tense is often hard to separate from aspect and modality (61). For example, in some languages the primary distinction is between perfective and imperfective (roughly whether an action is complete or incomplete) rather than between past and future. Some languages include markers for categories (e.g., past perfective) that combine tense and aspect. When consulting our primary sources, we made our best judgment about whether a language could be represented in our coding scheme without distorting it too greatly and excluded two languages (Hawaiian and Ewe) for which our scheme seemed especially inadequate. A second and less fundamental challenge is that languages which make remoteness distinctions do not include categories that are precisely equivalent. Our coding scheme distinguishes between remote (more than 7 d distant), near (between 1 and 7 d), and immediate (on the same day), and we fitted each language into this scheme as best we could.

Evidentiality. Evidentiality is a grammatical feature that conveys the source of a piece of information: for example, whether the speaker saw an event or heard it described by another person. There is no standard characterization of the space of possible sources, and we therefore formulate \mathcal{U} as the set of six sources distinguished by Aikhenvald in her typology of evidential systems (62). In principle, our framework allows these sources to be located within a multidimensional space, but for simplicity we order them along a single dimension that is consistent with Willett’s (63) hierarchy of evidentiality values and roughly captures distance from the speaker. The first source is visual perception, and the second includes all senses other than vision. Next comes inference from visual evidence (e.g., learning there was a fire by seeing smoke), followed by assumption. The penultimate source combines general world knowledge (e.g., “it is known”) and hearsay, and the final source is quoted speech. Languages group these six sources in different ways. For example, Quechua (Table 1) has markers for direct evidence (visual and sensory perception), indirect evidence (inference and assumption), and reported evidence (hearsay and quotation). In contrast, Turkish makes a simple partition between firsthand (visual sources) and nonfirsthand (all other information sources).

Psychological evidence from Western populations suggests that speakers are often uncertain about the source of information retrieved from memory (64), but there are no detailed characterizations of how this uncertainty is distributed across different kinds of sources. We therefore specify the speaker distributions using the same hierarchical approach described for tense. Following Willett’s (63) hierarchy, we assume major boundaries between perception (visual and sensory), reasoning (inference and assumption), and external report (hearsay and quotation speech) and minor boundaries within each of these three pairs. As for our tense analysis we use parameters κ and λ that specify how sharply probability mass decreases across minor and major boundaries. Our results are again qualitatively robust to variation in these parameters, and as before we set $\kappa = 0.5$ and

[†]We focus in this paper on absolute tense, leaving relative tense for future work.

$\lambda = 0.1$. The resulting speaker distributions are shown in Fig. 1, *Right Center*.

Although evidentiality occurs in around a quarter of the world's languages, few corpora are available for languages with fine-grained evidentiality inventories. We therefore estimate the prior $p(s)$ using a single corpus of Cuzco Quechua text (65). Quechua groups the six sources in \mathcal{U} into three pairs, and we therefore divide corpus frequencies evenly within these pairs to produce the prior shown in Fig. 1, *Bottom Right*. For evidentiality in particular, the data available for grounding assumptions about the prior and speaker distributions are relatively limited, and our results should be viewed as tentative conclusions only.

We conducted our analysis on a set of 184 extant languages, representing 61 language families and 16 unique inventories. Descriptions of all languages were taken from ref. 62, and five are represented in Table 1. To study how meanings are realized as forms, we compiled forms for a subset of 31 languages. Similar to tense, there was some difficulty separating evidentiality from other grammatical features including mood, mirativity (grammaticalized surprise), aspect, and tense. For example, in Mansi, Svan, and Turkish, evidentiality correlates with mirativity, and the nonfirsthand marker can be used when the speaker has perfect visual evidence but witnesses something so surprising that they do not believe it. Evidentiality can also interact with genre, register, and person systems. For example, in Meithei, the source of information is not always with respect to the speaker (first person) but can be calculated with respect to the listener (second person). As with tense, we tried our best to encode each system as faithfully as possible but acknowledge that our encoding of evidentiality represents a starting point only and that future work will be required (see *SI Appendix* for further discussion).

Analysis of Meaning

We first analyze the feature values or meanings captured by each language in our dataset, and the next section analyzes the forms that realize these meanings. For each of the three grammatical features, the space of possible encoders q_m is shown in Fig. 2.

Encoders that achieve optimal trade-offs between information loss and complexity lie along the Pareto frontier, shown here as a solid line, and the dark gray region below the curve shows trade-offs that are impossible to achieve. Attested inventories are shown as black points, and the light gray points include all possible inventories that partition \mathcal{U} into nonoverlapping feature values. Attested inventories (black points) are generally closer to the Pareto frontier than are unattested inventories (light gray points), suggesting that attested inventories for number, tense, and evidentiality are near optimal. *SI Appendix* includes a quantitative analysis that supports this conclusion strongly for number and tense and less strongly for evidentiality. It also shows that our model accounts better for attested inventories than an alternative approach previously applied to tense marking (66) that defines the complexity of an inventory as the number of markers that it includes.

Although most attested inventories lie close to the Pareto frontier, there are a handful of notable exceptions. There are no clear outliers for number, and for tense, the single outlier is Hixkaryana. The Hixkaryana tense inventory (Table 1) is unusual because it includes a relatively large number of categories but does not distinguish between present and future. For evidentiality, there are two notable groups of outliers driven by the same principle: our model predicts that distinctions at the level of Willett's (63) clusters should be made before distinctions within these clusters. For a few two-term systems, including Turkish, Mansi, and Meithei, and one three-term system, Siona, a distinction between visual and sensory information is made before distinctions are made between all of Willett's (63) levels. For a more detailed discussion of individual languages and outliers, see *SI Appendix*. The general conclusion from this discussion is that most attested systems are qualitatively similar to optimal systems.

For each of the plots in Fig. 2, traversing the Pareto frontier from top left to bottom right generates a hypothetical evolutionary trajectory that makes predictions about the order in which distinctions are introduced if a system grows more complex over time (14, 26). *SI Appendix* includes a detailed analysis of

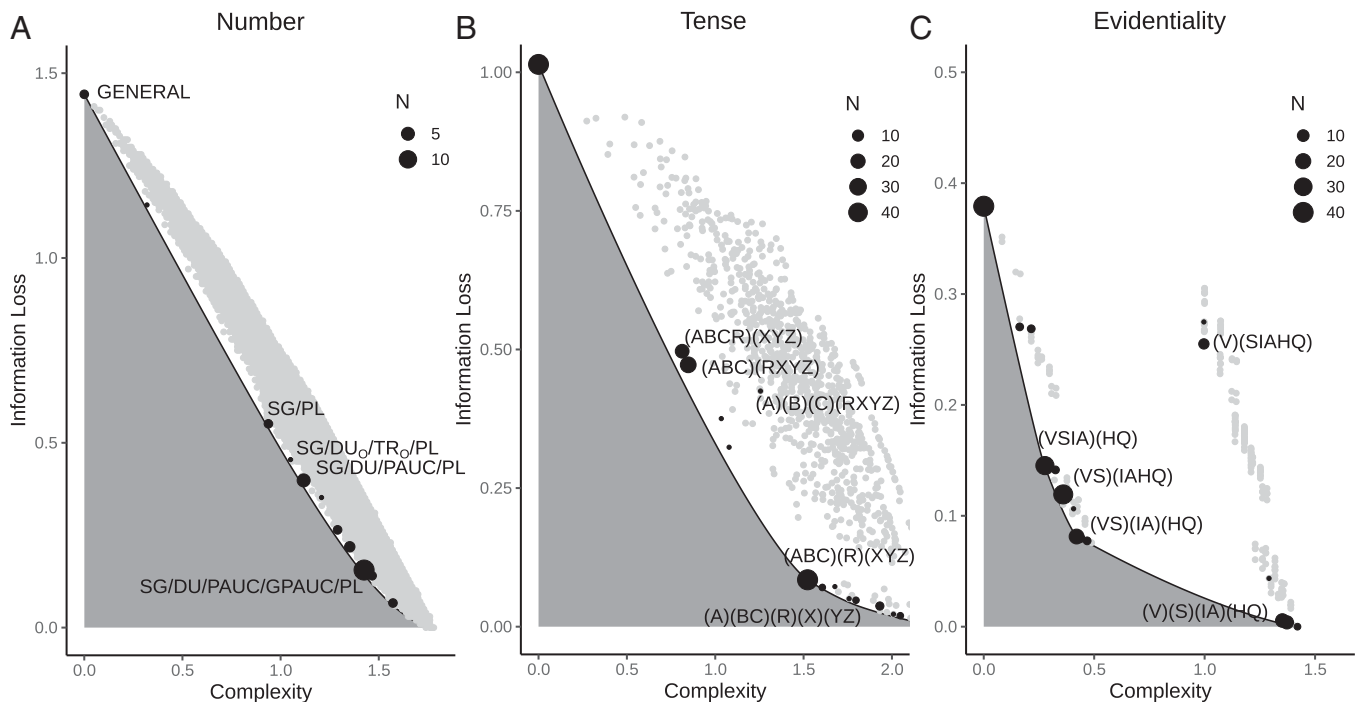


Fig. 2. Analysis of the meaning of grammatical markers. Trade-offs between information loss and complexity for (A) number, (B) tense, and (C) evidentiality. Attested inventories (black points) and unattested systems (gray points) are plotted in the space of all possible grammatical systems. Systems that achieve optimal trade-offs lie along the Pareto frontier (solid line), and the shaded region below the line shows trade-offs that are impossible to achieve.

these trajectories and shows that they recapitulate some patterns previously identified by work in linguistic typology. Following (67), these patterns are often formulated as universal constraints on possible systems: for example, one such universal states that if a number system has a trial, then it also has a dual. Our theory broadly captures this and other known patterns but is most compatible with the view that they are strong regularities that emerge from soft functional constraints instead of strict universals that hold without exception (68, 69).

Analysis of Form

We now analyze form length for number, tense, and evidentiality. Before comparing form lengths across languages, we normalize lengths within each system to allow for the fact that lengths may be systematically longer in some languages (e.g., those with relatively small phoneme inventories) than others. Our first analysis asks whether feature values that are zero-marked (i.e., not overtly expressed, as for the nominal singular marker in English) tend to be more frequent than other feature values belonging to the same system (9, 11). To address this question we use a coarse form of normalization that assigns a length of 0 to any feature value that is zero-marked and lengths of 1 to all other feature values. Fig. 3 plots the information loss from our analysis of meaning against expected length for tense and therefore shows how informativeness of meaning trades off against brevity of form. *SI Appendix* contains an analogous plot for evidentiality but not number because our sample of number markers includes pronominal forms for which zero-marking does not occur. The black dots represent attested systems, and the light blue dots include all permutations of systems that use zero-marking for at most one category in an attested system. The small gray dots show all ways to apply zero-marking to unattested systems. Attested systems with zero-marking overwhelmingly tend to zero-mark the most frequent feature value and therefore lie along the Pareto frontier. The remaining attested systems explicitly mark all grammatical feature values and appear as a column of black dots with expected length equal to 1. *SI Appendix* includes a statistical analysis suggesting that whether or not a tense system uses

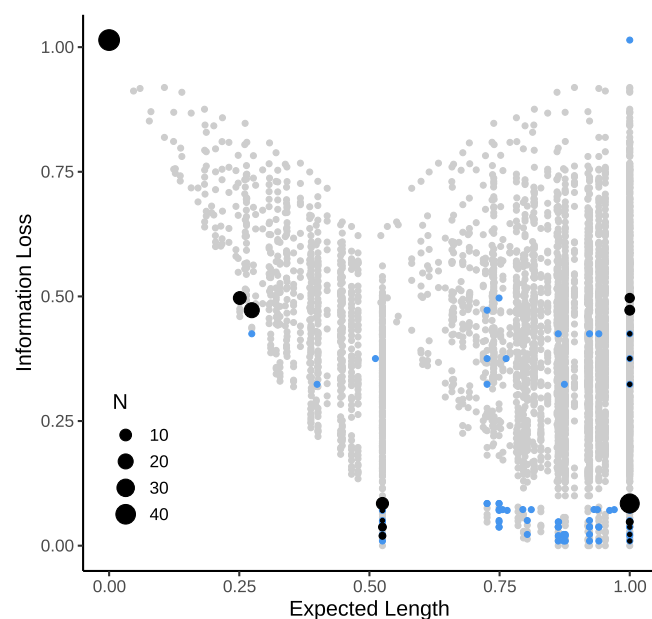


Fig. 3. Zero-marking analysis of tense systems. Trade-off between information loss and expected length when zero-marking is allowed for all tense systems ($N = 157$). Black dots show attested systems (size denotes frequency), blue dots show all ways to zero-mark at most one feature value in an attested system, and gray dots show possible but unattested systems.

zero-marking can be partially predicted by the information loss of the meaning encoding system. When information loss is high, zero-marking provides relatively large reductions in expected length and is relatively likely to be used. In contrast, systems with low information loss have little to gain by zero-marking and are relatively likely to be explicit.

We now ask more generally whether the frequency of a grammatical marker is inversely related to the length of its form. Form length should ideally be measured in phonemes, but we do not have phonemic transcriptions for all languages in our samples and therefore use orthographic length as a rough proxy for phonemic length. Within each system, form lengths are normalized so that the longest form has length 1. We compare these observed lengths to predicted or optimal lengths, where the optimal length for a marker with probability $p(m)$ is the surprisal $-\log(p(m))$. Frequent markers have short optimal lengths, and the optimal length of each marker can be interpreted as the number of bits used to represent the marker given an optimal code. Fig. 4 shows that observed and optimal lengths are positively correlated across our samples of number, tense, and evidentiality systems, and correlations for selected languages are shown in the seventh column of Table 1. The labeled data points in Fig. 4 are based on averages across all systems that share a given feature value, and the gray lines are regression lines based on lengths from individual languages. Some individual languages (gray lines) represent exceptions to the general trend—for example, Tamil and Seneca have slightly shorter forms for future tense than for past (ABC) or past/present (ABCR) tenses. In general, however, languages tend to assign relatively short forms to markers that are high in frequency.

The results in Fig. 4 are highly compatible with previous discussions of coding efficiency and grammatical marking. Most relevant to our approach is the work of Haspelmath (11), who uses a broad range of grammatical patterns to demonstrate that more frequent grammatical feature values tend to have relatively short forms and explains this result using the same functional-adaptive principles invoked by our theory. Relative to this prior work, our main contribution is to suggest that coding efficiency should not be considered in isolation but rather trades off against a pressure for informative communication.

Discussion

We presented an account of grammatical marking which suggests that number, tense, and evidentiality systems across languages achieve efficient trade-offs between informativeness and simplicity. Our results align with related results previously reported for domains including color naming (14), kin naming (44), quantifiers (70), numeral systems (71), and indefinite pronouns (72) and with a broader literature that characterizes ways in which language supports efficient communication (8). Within this literature, there are studies that focus on meaning (e.g., ref. 13) and studies that focus on form (e.g., ref. 10) but few that address both meaning and form (e.g., refs. 73, 74). Our work suggests how form and meaning can be brought together in an integrated information-theoretic framework.

Our analysis assumed that the function of each grammatical feature is to convey information about a single underlying dimension. Our approach, however, can be directly applied to settings in which the conceptual universe \mathcal{U} combines multiple semantic dimensions, for example, both person and number (75). Applying the framework in this way may provide a fresh perspective on grammatical paradigms and may help to explain attested patterns of syncretism. A further possible extension is to allow for additional functions that grammatical features may serve: for example, some features (e.g., case; ref. 76) may convey information about structural dependencies via indexing, and others (e.g., grammatical gender; ref. 77) may convey information about what forms should be expected next. Frequent linguistic

0.001 increments. For tense, the β schedule was 2^x for x from 5 to 0 by 0.001 increments. For evidentiality, the β schedule was 2^x for x from 5 to 0 by 0.001 increments.

Data Availability. All data and code used in the analyses are available in an Open Science Foundation repository (89) at <https://osf.io/s5b7h/>.

1. W. Croft, *Typology and Universals* (Cambridge University Press, 2002).
2. F. J. Newmeyer, *Possible and Probable Languages: A Generative Perspective on Linguistic Typology* (Oxford University Press, 2005).
3. G. von der Gabelentz, *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse* (Chr. Herm. Tauchnitz, Leipzig, 1901).
4. G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley Press, 1949).
5. J. Bybee, *Language, Usage and Cognition* (Cambridge University Press, 2010).
6. J. A. Hawkins, *Efficiency and Complexity in Grammars* (Oxford University Press, 2004).
7. J. A. Hawkins, *Cross-Linguistic Variation and Efficiency* (Oxford University Press, 2014).
8. E. Gibson *et al.*, How efficiency shapes human language. *Trends Cogn. Sci.* **23**, 389–407 (2019).
9. J. H. Greenberg, *Language Universals with Special Reference to Feature Hierarchies* (Mouton, The Hague, 1966).
10. S. T. Piantadosi, H. Tily, E. Gibson, Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3526–3529 (2011).
11. M. Haspelmath, Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *J. Linguist.* **57**, 605–633 (2021).
12. T. Regier, C. Kemp, P. Kay, “Word meanings across languages support efficient communication” in *The Handbook of Language Emergence*, B. MacWhinney, W. O’Grady, Eds. (Wiley, 2015), pp. 237–263.
13. C. Kemp, Y. Xu, T. Regier, Semantic typology and efficient communication. *Annu. Rev. Linguist.* **4**, 109–128 (2018).
14. N. Zaslavsky, C. Kemp, T. Regier, N. Tishby, Efficient compression in color naming and its evolution. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7937–7942 (2018).
15. W. D. Laidig, C. J. Laidig, Larike pronouns: Duals and trials in a central Moluccan language. *Ocean. Linguist.* **29**, 87–109 (1990).
16. M. Haspelmath, A. Karjus, Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* **55**, 1213–1235 (2017).
17. M. Fedzechkina, T. F. Jaeger, E. L. Newport, Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17897–17902 (2012).
18. J. Kanwal, K. Smith, J. Culbertson, S. Kirby, Zipf’s law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition* **165**, 45–52 (2017).
19. C. Kurumada, S. Grimm, Predictability of meaning in grammatical encoding: Optional plural marking. *Cognition* **191**, 103953 (2019).
20. M. Fedzechkina, T. F. Jaeger, Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition* **196**, 104115 (2020).
21. J. Culbertson, P. Smolensky, G. Legendre, Learning biases predict a word order universal. *Cognition* **122**, 306–329 (2012).
22. D. Gentner, M. Bowerman, “Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis” in *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*, J. Guo *et al.*, Eds. (Psychology Press, 2009), pp. 465–480.
23. D. Saratsli, S. Bartell, A. Papafragou, Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition* **197**, 104194 (2020).
24. J. Haiman, “Competing motivations” in *The Oxford Handbook of Linguistic Typology*, J. J. Song, Ed. (Oxford University Press, 2010), pp. 148–165.
25. J. W. Du Bois, “Competing motivations” in *Iconicity in Syntax*, J. Haiman, Ed. (Benjamins Amsterdam, 1985), pp. 343–365.
26. N. Zaslavsky, “Information-theoretic principles in the evolution of semantic systems,” PhD thesis, The Hebrew University of Jerusalem, Jerusalem, Israel (2020).
27. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
28. C. E. Shannon “Coding theorems for a discrete source with a fidelity criterion” in *Institute of Radio Engineers International Convention Record* (Institute of Radio Engineers, 1959), vol. 7, pp. 325–350.
29. N. Zaslavsky, C. Kemp, N. Tishby, T. Regier, Communicative need in colour naming. *Cogn. Neuropsychol.* **37**, 312–324 (2020).
30. A. Karjus, R. A. Blythe, S. Kirby, K. Smith, Communicative need modulates competition in language change. arXiv [Preprint] (2020). <https://arxiv.org/abs/2006.09277> (Accessed 17 June 2020).
31. E. C. Traugott, “Grammaticalization and mechanisms of change” in *The Oxford Handbook of Grammaticalization*, B. Heine, H. Narrog, Eds. (Oxford University Press, 2011), pp. 19–30.
32. N. Tishby, F. Pereira, W. Bialek, “The information bottleneck method” in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, N. Tishby, F. C. Pereira, B. Hajek, R. S. Screenivas, Eds. (University of Illinois Press, 1999), pp. 368–377.
33. R. L. Futrell, “Memory and locality in natural language,” PhD thesis, Massachusetts Institute of Technology, Cambridge, MA (2017).
34. J. Pate, “Optimization of American English, Spanish, and Mandarin Chinese over time for efficient communication” in *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, G. Gunzelmann, A. Howes, T. Tenbrink, E. J. Davelaar, Eds. (Cognitive Science Society, 2017), pp. 901–906.
35. T. Pimentel, I. Nikkarinen, K. Mahowald, R. Cotterell, D. Blasi, “How (non-)optimal is the lexicon?” in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Association for Computational Linguistics, 2021), pp. 4426–4438.
36. G. Fenk-Oczlon, Familiarity, information flow, and linguistic form. *Typolog. Stud. Lang.* **45**, 431–448 (2001).
37. C. Bentz, R. Ferrer-i Concho, “Zipf’s law of abbreviation as a language universal” in *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics* (University of Tübingen, 2016).
38. 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).
39. K. Mahowald, E. Fedorenko, S. T. Piantadosi, E. Gibson, Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* **126**, 313–318 (2013).
40. M. Hahn, D. Jurafsky, R. Futrell, Universals of word order reflect optimization of grammars for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2347–2353 (2020).
41. R. Ferrer i Cancho, R. V. Solé, Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 788–791 (2003).
42. R. P. Cooper, O. Guest, Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cogn. Syst. Res.* **27**, 42–49 (2014).
43. O. Guest, A. E. Martin, How computational modeling can force theory building in psychological science. *Perspect. Psychol. Sci.* **16**, 789–802 (2021).
44. C. Kemp, T. Regier, Kinship categories across languages reflect general communicative principles. *Science* **336**, 1049–1054 (2012).
45. P. Rácz, S. Passmore, C. Sheard, F. M. Jordan, Usage frequency and lexical class determine the evolution of kinship terms in Indo-European. *R. Soc. Open Sci.* **6**, 191385 (2019).
46. G. G. Corbett, *Number* (Cambridge University Press, 2000).
47. S. Grimm, Grammatical number and the scale of individuation. *Language* **94**, 527–574 (2018).
48. S. J. Cheyette, S. T. Piantadosi, A unified theory of numerosity perception. *Nat. Hum. Behav.* **4**, 1265–1272 (2020).
49. S. Dehaene, J. Mehler, Cross-linguistic regularities in the frequency of number words. *Cognition* **43**, 1–29 (1992).
50. S. T. Piantadosi, A rational analysis of the approximate number system. *Psychon. Bull. Rev.* **23**, 877–886 (2016).
51. D. C. Derbyshire, *Hixkaryana* (North-Holland, 1979).
52. N. Rescher, A. Urquhart, *Temporal Logic* (Springer-Verlag, Wien, 1971).
53. J. McCarthy, P. J. Hayes, “Some philosophical problems from the standpoint of artificial intelligence” in *Readings in Artificial Intelligence*, B. L. Webber, N. J. Nilsson, Eds. (Elsevier, 1981), pp. 431–450.
54. J. F. Allen, Maintaining knowledge about temporal intervals. *Commun. ACM* **26**, 832–843 (1983).
55. H. Reichenbach, “The tenses of verbs” in *The Language of Time: A Reader*, I. Mani, J. Pustejovsky, R. Gaizauskas, Eds. (Oxford University Press, 1947), pp. 71–78.
56. V. Velupillai, Partitioning the timeline: A cross-linguistic survey of tense. *Stud. Lang.* **40**, 93–136 (2016).
57. B. Comrie, *Tense* (Cambridge University Press, 1985).
58. G. Park *et al.*, Living in the past, present, and future: Measuring temporal orientation with language. *J. Pers.* **85**, 270–280 (2017).
59. J. B. Michel *et al.*; Google Books Team, Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
60. Ö. Dahl, *Tense and Aspect Systems* (Basil Blackwell, 1985).
61. J. L. Bybee, R. D. Perkins, W. Pagliuca, *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World* (University of Chicago Press, 1994).
62. A. Aikhenvald, *Evidentiality* (Oxford University Press, 2004).
63. T. Willett, A cross-linguistic survey of the grammaticization of evidentiality. *Stud. Lang.* **12**, 51–97 (1988).
64. M. K. Johnson, S. Hashtroudi, D. S. Lindsay, Source monitoring. *Psychol. Bull.* **114**, 3–28 (1993).
65. A. Rios, A. Göhring, M. Volk, “A Quechua-Spanish parallel treebank” in *LOT Occasional Series* (LOT, Netherlands Graduate School of Linguistics), vol. 12, pp. 53–64 (2008).
66. G. I. Bacon, “Evaluating linguistic knowledge in neural networks,” PhD thesis, University of California, Berkeley, CA (2020).
67. J. H. Greenberg, *Universals of Language* (MIT Press, 1963).
68. M. S. Dryer, “Why statistical universals are better than absolute universals” in *Papers from the 33rd Regional Meeting of the Chicago Linguistic Society* (Chicago Linguistic Society, 1998), pp. 123–145.
69. N. Evans, S. C. Levinson, The myth of language universals: Language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448, discussion 448–494 (2009).

70. S. Steinert-Threlkeld, "Quantifiers in natural language optimize the simplicity/informativeness trade-off" in *Proceedings of the 22nd Amsterdam Colloquium*, J. J. Schöder, D. McHugh, F. Roelofsen, Eds. (Institute for Logic, Language and Computation at the University of Amsterdam, 2020), pp. 513–522.
71. Y. Xu, E. Liu, T. Regier, Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind (Camb.)* **4**, 57–70 (2020).
72. M. Denić, S. Steinert-Threlkeld, J. Szymanik, "Complexity/informativeness trade-off in the domain of indefinite pronouns" in *Semantics and Linguistic Theory 2020*, J. Rhyne, K. Lamp, N. Dreier, C. Kwon, Eds. (Linguistic Society of America, 2020), pp. 166–184.
73. I. Dautriche, K. Mahowald, E. Gibson, S. T. Piantadosi, Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cogn. Sci. (Hauppauge)* **41**, 2149–2169 (2017).
74. M. Tamariz, Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *Ment. Lex.* **3**, 259–278 (2008).
75. N. Zaslavsky, M. Maldonado, J. Culbertson, "Let's talk (efficiently) about us: Person systems achieve near-optimal compression" in *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, T. Fitch, C. Lamm, H. Leder, K. Teßmar-Raible, Eds. (Cognitive Science Society, 2021), pp. 938–944.
76. F. Mollica, C. Kemp, "An efficient communication analysis of morpho-syntactic grammatical features" in *Proceedings of the 42th Annual Conference of the Cognitive Science Society*, S. Denison, M. Mack, Y. Xu, B. C. Armstrong, Eds. (Cognitive Science Society, 2020), pp. 3198–3204.
77. M. Dye, P. Milin, R. Futrell, M. Ramscar, "A functional theory of gender paradigms" in *Perspectives on Morphological Organization*, F. Kiefer, J. Blevins, H. Bartos, Eds. (Brill, 2017), pp. 212–239.
78. M. Haspelmath, "The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison" in *The New Psychology of Language*, M. Tomasello, Ed. (Psychology Press, 2003), vol. 2, pp. 1–30.
79. S. Caplan, J. Kodner, C. Yang, Miller's monkey updated: Communicative efficiency and the statistics of words in natural language. *Cognition* **205**, 104466 (2020).
80. S. Kirby, M. Tamariz, H. Cornish, K. Smith, Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141**, 87–102 (2015).
81. A. Carstensen, J. Xu, C. Smith, T. Regier, "Language evolution in the lab tends toward informative communication" in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, D. C. Noelle et al., Eds. (Cognitive Science Society, 2015), pp. 303–308.
82. J. W. Carr, K. Smith, J. Culbertson, S. Kirby, Simplicity and informativeness in semantic category systems. *Cognition* **202**, 104289 (2020).
83. N. Fay, S. Garrod, L. Roberts, N. Swoboda, The interactive evolution of human communication systems. *Cogn. Sci.* **34**, 351–386 (2010).
84. C. L. Hudson Kam, E. L. Newport, Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang. Learn. Dev.* **1**, 151–195 (2005).
85. J. Culbertson, P. Smolensky, A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cogn. Sci.* **36**, 1468–1498 (2012).
86. F. M. Jordan, A phylogenetic analysis of the evolution of Austronesian sibling terminologies. *Hum. Biol.* **83**, 297–321 (2011).
87. H. J. Haynie, C. Bowern, Phylogenetic approach to the evolution of color term systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13666–13671 (2016).
88. M. Dunn, S. J. Greenhill, S. C. Levinson, R. D. Gray, Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79–82 (2011).
89. F. Mollica et al., The forms and meanings of grammatical markers support efficient communication. Open Science Framework. <https://doi.org/10.17605/OSF.IO/55B7H>. Deposited 29 September 2021.