When Efficient Communication Explains Convexity

Ashvin Ranjan[1] **& Shane Steinert-Threlkeld**[2]
[1]Paul G. Allen School of Computer Science & Engineering, University of Washington
[2]Department of Linguistics, University of Washington
{ar31, shanest}@uw.edu

Abstract

Much recent work has argued that the variation in the languages of the world can be explained from the perspective of efficient communication; in particular, languages can be seen as optimally balancing competing pressures to be simple and to be informative. Focusing on the expression of meaning—semantic typology—the present paper asks what factors are responsible for successful explanations in terms of efficient communication. Using the Information Bottleneck (IB) approach to formalizing this trade-off, we first demonstrate and analyze a correlation between optimality in the IB sense and a novel generalization of *convexity* to this setting. In a second experiment, we manipulate various modeling parameters in the IB framework to determine which factors drive the correlation between convexity and optimality. We find that the convexity of the communicative need distribution plays an especially important role. These results move beyond showing that efficient communication can explain aspects of semantic typology into explanations for why that is the case by identifying which underlying factors are responsible.

Keywords: efficient communication; convexity; semantic typology; information bottleneck

Introduction

A central goal in the cognitive science of language centers on explaining semantic typology, the range and limits of variation in how the languages of the world express meaning. Two prominent questions arise in the pursuit of this goal. Descriptively: what are the similarities and differences in how languages encode meaning? Explanatorily: which aspects of this typology are idiosyncratic historical accidents, and which follow from more general cognitive and social factors?

To the descriptive question: it has been argued that word meanings in the world's languages denote *convex* regions in relevant geometric spaces (Chemla et al., 2019; Gärdenfors, 2000, 2014; Jäger, 2010). Convex regions are geometrically well-behaved: for any two points in such a region, any third point between them must also belong to the region. This creates 'smooth' or 'natural' borders. This has been argued to be a kind of 'meta-universal', applying to word meanings across many different semantic domains.

To the explanatory question: multiple factors, especially including ease of learning (Douven, 2025; Maldonado et al., 2022; Steinert-Threlkeld & Szymanik, 2019, 2020) and efficient communication (Denić et al., 2022; Imel et al., 2026; Kemp et al., 2018; Mollica et al., 2021; Steinert-Threlkeld, 2021; Uegaki, 2023; Zaslavsky et al., 2018, 2021) have been argued to explain why semantic systems are structured the way that they are. While the efficient communication approach has been successfully applied in many empirical domains, it has largely not been used to explain convexity. Furthermore, the wide success of the approach calls for a deeper understanding of why it works when it does.

To this end, this paper asks a question connecting the two dimensions: under what conditions can efficient communication explain the convexity of word meanings? In particular, we approach efficient communication from the perspective of the information bottleneck (IB) framework (Tishby et al., 1999; Zaslavsky et al., 2018) and report two experiments.

First, we focus on color naming systems, which have been well-studied both from the perspective of convexity (Jäger, 2010; Steinert-Threlkeld & Szymanik, 2020) and from the IB framework (Zaslavsky et al., 2018). First, we introduce a graded measure of *quasi-convexity* which generalizes the degree of convexity used elsewhere (Carlsson et al., 2024; Koshevoy & Szymanik, 2025; Steinert-Threlkeld & Szymanik, 2020) to the scenario (as in IB) where meanings are probabilistic and show that this measure correlates with IB optimality. Second, to address our primary question, we introduce smaller, 'toy' semantic spaces where we can systematically vary key components of the IB framework, in order to analyze which aspects are essential for producing a correlation between convexity and efficiency. We find that the relative convexity of the prior distribution over meanings—often called the communicative need distribution—plays an especially pivotal role.

The paper is structured as follows. In the next section, we introduce both the IB framework and our measure of the degree of convexity. We then report our results on color naming systems, before turning to the experiment that manipulates IB parameters in a controlled setting. We conclude by discussing the relation to existing work and ramifications of these results.

Methodology

We first present the components of the methodology shared between our two experiments before the results of each of them separately. Code and data for reproducibility may be found at https://github.com/CLMBRs/efficiency-convexity.

Information Bottleneck Encoders

The Information Bottleneck models communication between a speaker and a listener via efficient compression (Zaslavsky

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

et al., 2018). The speaker observes a meaning $m \in M$, which is a probability distribution over the given referents in an environment, U (we use u for an item $u \in U$). The speaker then aims to express the meaning m by selecting a word w in their vocabulary W. Meanings are conditional probability distributions $p(u|m)$ and speakers are another one, $q(w|m)$, referred to as an encoder q.

To calculate the efficiency of a given encoder, we can use two metrics $I_q(W;U)$ and $I_q(M;W)$ (Tishby et al., 1999; Zaslavsky et al., 2018) which can be thought of as the accuracy and complexity, respectively, of q.

Optimal Encoders The accuracy and complexity of an encoder are related to each other: it is impossible to minimize complexity and simultaneously maximize accuracy. To optimally balance the two measures, the IB framework seeks encoders which minimize

$$\mathcal{F}_\beta[q(w|m)] = I_q(M;W) - \beta I_q(W;U)$$

where $\beta \geq 0$ is the trade-off parameter.

To calculate the optimal frontier $\mathcal{F}_\beta[q(w|m)]$ for various values of β, we used reverse deterministic annealing as described in Zaslavsky et al. (2018). Specific implementation details may be found in an Appendix.

Suboptimal Encoders In addition to the optimal encoders (as well as natural language encoders, when available), it is important to sample a wide range of other, suboptimal encoders. We employ a technique of re-sampling adapted from Skinner (2025) to derive suboptimal encoders while still retaining aspects of the structure of optimal ones. For a given encoder q, we create a re-sampled encoder $\hat{q}$ by selecting a random subset of meanings $\hat{M}$ based on the percentage for re-sampling (ranging from 10% to 100% in 10% increments). For each $\hat{m}_i \in \hat{M}$, we randomly draw with replacement $\hat{m}_j$ from $\hat{M}$ and define $\hat{q}(w|\hat{m}_i) = q(w|\hat{m}_j)$. And for all meanings $m_i \notin \hat{M}$, $\hat{q}(w|m_i) = q(w|m_i)$. In addition to controlling for aspects of the structure of the original encoder, we can adjust how different a re-sampled encoder is from the original by adjusting the percentage of meanings re-sampled.

We have avoided using rotation to define suboptimal encoders (Koshevoy & Szymanik, 2025; Regier et al., 2007; Zaslavsky et al., 2018). Many of our evaluations depend on the geometry of the CIELab color space; the afore-cited papers rotate along the Hue axis of specified Munsell chips. This rotation does not commute with the mappings to and from the two color spaces, and so rotating Hue can have deleterious effects on our metrics.

A primary benefit of re-sampling over shuffling or standard rotation (one that is applied within the space that meanings are being evaluated in) is that re-sampling applies a non-invertible function to the set of words W for a given language. If the function applied is instead invertible, as it is with shuffling or standard rotation, the complexity $I(M;W)$ of the encoder would remain constant due to the data-processing inequality.

Quasi-Convexity

Convexity has been generalized from a binary property to a scalar metric (i.e. a degree of convexity) for a given subset of points in a space in other work (Koshevoy & Szymanik, 2025; Steinert-Threlkeld & Szymanik, 2020):

$$\text{convexity}(X) = \frac{||X||}{||\text{ConvexHull}(X)||}$$

where $\text{ConvexHull}(X)$ is the smallest convex set extending X. This definition applies to sets with hard boundaries. However, in the case of IB encoders we have to instead deal with probability distributions, which can be thought of as sets with 'soft' boundaries.

We can generalize from here to a notion of quasi-convexity as follows. Let $p : \Omega \to [0,1]$ be a p.m.f. or p.d.f. and let $\text{ls}(f,t) := \{x \mid f(x) \geq t\}$ for $f : X \to \mathbb{R}$, $x \in X$, $t \in \mathbb{R}$ (i.e. 'ls' for level-set). Then:

$$\mathsf{dcon}(p) := \int_0^{\max(p(x))} \text{convexity}(\text{ls}(p,t))dt \qquad (1)$$

A finite-sum approximation of $\mathsf{dcon}(p)$ is defined in Algorithm 1, which is lightly modified from Skinner (2025).

Algorithm 1 Quasi-convexity calculation

Require: $steps > 0$
Ensure: $0 \leq p(x)\ \forall x \in X, \sum_{x \in X} p(x) = 1$
  $mesh \leftarrow \frac{1}{steps}$
  $level \leftarrow \frac{\max(p(x))}{steps}$
  $qc \leftarrow 0$
  for $i \in 1, \ldots, steps$ **do**
    $X_{level} \leftarrow \{x \mid p(x) \geq level \times i\}$
    if $|X_{level}| = 0$ **then**
      $qc \leftarrow qc + mesh$
    else
      $X_{hull} \leftarrow ConvHull(X_{level})$
      $qc \leftarrow qc + mesh \times \frac{|X_{level}|}{|X_{hull}|}$
    end if
  end for

In order to apply this to a set of conditional probabilities, we take the weighted sum of the convexity of $p(x|y)$ for all $y \in Y$ utilizing Algorithm 1. Let $p : X \times Y \to [0,1]$ be a conditional p.m.f. or p.d.f. (e.g. an IB encoder). Let $\mathsf{dcon}(p)$ be defined by Equation 1. Then:

$$\mathsf{Convexity}(p) = \sum_{y \in Y} p(y)\mathsf{dcon}(p(\cdot|y)) \qquad (2)$$

In the case of the IB, we will focus primarily on the $\text{dcon}(q(m|w))$ and the weighted average across $q(w)$ in (2).

Experiment 1: Color Naming Systems

Setup Our first experiment focuses on color naming systems, using the IB environment detailed in Zaslavsky et al. (2018). This environment utilizes data from the World Color Survey (WCS), which displayed 330 Munsell color chips to

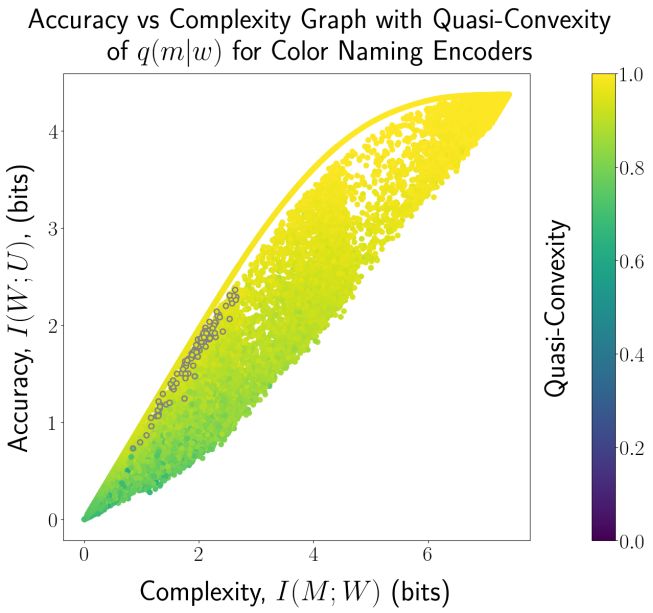

Figure 1: Accuracy vs complexity trade-off for color naming. Color represents quasi-convexity, ranging from purple (0.0 convexity) to yellow (1.0 convexity). Points with a gray outline represent a natural language.

participants in order to gather natural language data for color naming (Berlin & Kay, 1969; Cook et al., 2005). These chips are the referents U and meanings are Gaussian distributions over the color chips in CIELab color space[1], with a standard deviation of 64. Furthermore, we use the least-informative prior over meanings described in Zaslavsky et al. (2018).

Procedure We calculated the optimal encoders using an implementation of IB and the same 1501 values of β and reverse annealing described in Zaslavsky et al. (2018).[2] Natural language data from WCS is turned into encoders by using frequency information. Specific details can be found in an Appendix. We re-sample the optimal encoders to generate suboptimal encoders sampling at percentages ranging from 10% to 100%, with 10% increments.

Because convexity is a property of word meanings, we measure quasi-convexity by applying Equation 2 to the distribution $q(m|w)$, which is defined by Bayes' rule (Zaslavsky et al., 2018). We measure the optimality of an encoder as the negative Epsilon measure from Zaslavsky et al. (2018).

Results Figure 1 visualizes the quasi-convexity of both distributions for optimal, suboptimal, and natural encoders. Pearson correlations with optimality, complexity, and accuracy are reported in Table 1.

[1]The conversion from the WCS color space to CIELab utilizes the data from linguistics.berkeley.edu/wcs/data.html

[2]The values of β, along with additional data which we have used for this experiment, are available at github.com/nogazs/ib-color-naming.

Distribution	Optimality	$I(M;W)$	$I(W;U)$
$q(m\|w)$	0.221	0.797	0.885

Table 1: Pearson's correlation coefficients between the quasi-convexity of $q(m|w)$ and Optimality, Complexity ($I(M;W)$), and Accuracy ($I(W;U)$) in color naming. The p values for all correlation coefficients are less than 0.00.

We observe a small but non-trivial positive correlation between optimality and quasi-convexity, suggesting that as languages become more optimal, they also become more convex. We observe much larger correlations, however, between quasi-convexity and the complexity and accuracy metrics.

In order to explore the relationships between quasi-convexity and all of these metrics in more detail, we fit a mixed-effects regression model predicting convexity as a function of encoder type (natural, optimal, suboptimal) and "base type" (natural or optimal; describing which type of encoder a suboptimal one is derived from), with a random effect for the base encoder. A likelihood ratio test with a standard linear regression omitting the random effect showed that including the random effect significantly improves model fit ($\chi^2(1) = 16911.75, p \approx 0$).

Across 17,721 observations from 1,611 item groups (all groups sized 11, for base encoder and 10 variants of it), both predictors showed significant effects. Compared to natural language encoders (the baseline for the encoder type variable), optimal encoders showed slightly lower convexity scores ($\beta = -0.032, \text{SE} = 0.004, z = -7.49, p < .001$)—likely due to the lower-convexity optimal encoders in the low-accuracy and low-complexity region of the plane—and suboptimal items showed substantially lower scores ($\beta = -0.106, \text{SE} = 0.004, z = -25.30, p < .001$). Items derived from optimal base encoders showed significantly higher convexity values than those derived from natural base encoders ($\beta = 0.066, \text{SE} = 0.007, z = 9.91, p < .001$). The random intercept exhibited a small but non-zero variance component (0.004), indicating meaningful between-item variability. These results show that suboptimal encoders do in fact have lower convexity scores, but that those derived from optimal encoders are slightly more convex than those derived from natural language encoders.

Finally, a multiple linear regression was conducted to examine how optimality, complexity, and accuracy jointly predict convexity. This analysis focused on $q(m|w)$ only, since it has the higher correlations with these factors. The model included all two-way and three-way interactions among these continuous predictors, as well as fixed effects for item type and base type. The overall model explained a large amount of variance in convexity scores ($R^2 = .93$, $F(10; 17{,}710) = 23{,}890$, $p < .001$).

All three continuous predictors showed strong main effects. Convexity increased with greater optimality ($\beta = 0.225$, $SE = 0.003$, $t = 80.72$, $p < .001$), complexity ($\beta = 0.16$,

$SE = 0.002$, $t = 87.36$, $p < .001$), and (to a smaller degree) accuracy ($\beta = 0.014$, $SE = 0.001$, $t = 11.30$, $p < .001$). These effects, however, were moderated by their interactions. Small negative interactions between optimality and both complexity ($\beta = -0.017$, $p < .001$) and accuracy ($\beta = -0.006$, $p < .001$) show that the positive effect of optimality gets attenuated at higher levels of the latter two factors. The effect of complexity similarly decreased as accuracy increased ($\beta = -0.030$, $p < .001$). The three-way interaction among optimality, complexity, and accuracy ($\beta = 0.003$, $p = 0.092$) partially offsets these negative interactions but only approaches significance.

By contrast, categorical predictors exhibited minor effects. Neither the optimal type nor the base type had significant effects once the continuous predictors and their interactions were included, although suboptimal items showed slightly lower convexity scores than the natural languages ($\beta = -0.006$, $p = .015$). Similarly, encoders derived from optimal languages were slightly less convex than ones derived from natural languages ($\beta = -0.002$, $p = 0.026$). Overall, the results indicate variance in convexity can be explained by a complex interaction between optimality, complexity, and (to a lesser extent) accuracy, but natural languages are more convex than suboptimal ones even when accounting for those factors.

Experiment 2: Manipulating IB

Having demonstrated and analyzed a correlation between convexity and optimality in color naming, the second experiment aims to understand the factors that explain this correlation by manipulating various components of the IB framework in a simplified setting, creating various environments.

Base Environments We begin with four base environments, focusing on convexity of the priors and uniqueness of meanings. These all share the set of referents $U = \{0, \ldots, 10\}$ and meaning distributions:

$$p(u|m) \propto \phi\left(u, |m|, \frac{9}{4}\right)$$

where $\phi(u, \mu, \sigma^2)$ is the p.d.f. of the normal distribution.

The four environments, with explicit definitions of the meaning space and prior given in Table 2, are: CPUM (convex priors, unique meanings), NPUM (non-convex priors, unique meanings), CPDM (convex prior, duplicate meanings), and NPDM (non-convex prior, duplicate meanings). Convex prior environments have uniform priors, while non-convex ones are skewed heavily towards the 'edge' of the environment. DM environments have two identical meanings (thanks to the $|m|$ in the definition of $p(u|m)$) for each referent u.

We additionally explore the following additional environments, to more thoroughly explore the IB parameter space.

-Dual Variations For each base environment, we introduce a -Dual variant, with $U = M = \{-10, \ldots, 10\}$ and

$$p(u|m) \propto \left(\phi\left(u, m, \frac{3}{2}\right) + \phi\left(u, -m, \frac{3}{2}\right)\right)/2$$

Environment	M	$p(m)$
CPUM	$\{0, \ldots, 10\}$	$\frac{1}{\|\|M\|\|}$
NPUM	$\{0, \ldots, 10\}$	$\begin{cases} 0.455 & m = 0, \\ 0.455 & m = 10, \\ 0.01 & \text{otherwise} \end{cases}$
CPDM	$\{-10, \ldots, 10\}$	$\frac{1}{\|\|M\|\|}$
NPDM	$\{-10, \ldots, 10\}$	$\begin{cases} 0.455 & m = -10, \\ 0.455 & m = 10, \\ 0.01 & \text{otherwise} \end{cases}$

Table 2: The meanings M and prior $p(m)$ for the base environments. U and $p(u|m)$ are shared across these environments.

This distribution is a mixture of two normal distributions with peaks at u and $-u$, which means that $p(u|m) = p(-u|m)$.

CPDM-Convex This resembles CPDM, but the priors are heavily weighted to negative values in M. This means that the priors are still convex, but are not uniform. Specifically:

$$p(m) = \begin{cases} 0.099 & m < 0, \\ \frac{0.01}{11} & \text{otherwise} \end{cases}$$

CPDM-Adj This resembles CPDM, but with duplicate meanings adjacent to each other in the space. Specifically: $M = \{0, \ldots, 19\}$ and $p(u|m) \propto \phi\left(u, \left\lfloor \frac{m}{2} \right\rfloor, \frac{3}{2}\right)$

NPDM-Adj Like CPDM-Adj, but groups of identical meanings are non-convex. Specifically: $M = \{0, \ldots, 17\}$, $p(u|m) \propto \phi\left(u, \left\lfloor \frac{m}{3} \right\rfloor, \frac{3}{2}\right)$, and

$$p(m) = \begin{cases} \frac{0.01}{6} & m - 1 \equiv 0 \mod 3 \\ \frac{0.495}{6} & \text{otherwise} \end{cases}$$

Duplicate meanings come in contiguous blocks of three (thanks to $\lfloor m/3 \rfloor$), but the middle one has substantially lower prior probability than the outer two.

Manhattan-5-5 The referents form a 5×5 grid, with $p(u|m)$ being the normalized Manhattan distance from m to u. Specifically: $U = M = \{0, \ldots, 4\} \times \{0, \ldots, 4\}$, the prior over meanings is uniform, and $p(u|m) \propto \text{dist}(u, m) + 1$, where dist is the Manhattan distance and 1 is added to avoid 0.

Procedures

For each environment, we followed the same reverse deterministic annealing procedure as in the previous experiment to estimate the optimal frontier, and used the same re-sampling procedure and percentages to generate suboptimal encoders. Note that there are no natural language encoders in this experiment, since we are using hypothetical environments to test the correlation between convexity and optimality.

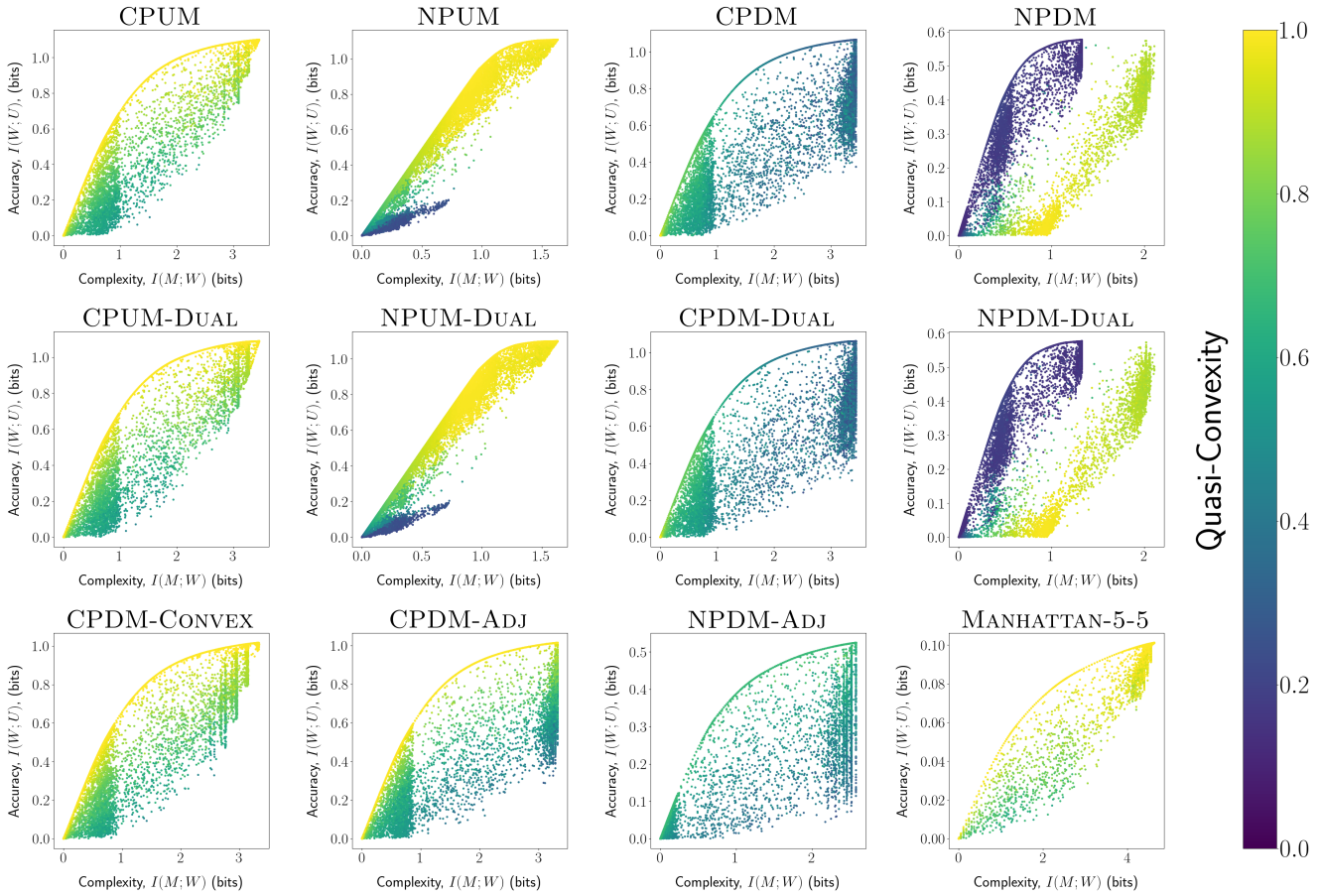

Figure 2: Accuracy vs complexity trade-off graphs for all environments. Color corresponds to quasi-convexity of $q(m|w)$.

Results

Figure 2 visualizes quasi-convexity of $q(m|w)$ for optimal and suboptimal encoders in each environment. Pearson correlations with optimality, complexity, and accuracy are reported in Table 3. Several trends can be observed from this data.

First, convex priors play a crucial role: every environment with convex priors (those starting with CP-) exhibits a strong correlation between optimality and quasi-convexity. These environments have nearly perfectly convex encoders. Unlike the color naming case, these environments also exhibit *negative* correlations between quasi-convexity and the IB accuracy and complexity metrics. Manhattan-5-5, which is roughly a two-dimensional analog of CPUM, exhibits the same pattern.

Crucially, however, the optimal languages in the duplicate meaning (-DM) environments are not, in general, fully convex. The CPDM environment, in particular, has convex priors but non-convex optimal meanings. This echoes concurrent results from Imel and Zaslavasky (2026), whose construction of IB-optimal but non-convex languages resembles the CPDM construction but with a circular instead of a linear symmetry. Although the optimal encoders in this setting are not convex, and they become less convex as one moves along the optimal frontier (also evidenced by strong negative correlations with complexity and accuracy), there is still a correlation between optimality and convexity.

Furthermore, each of the -Dual environments show similar behavior—both visually and in terms of the three correlations—as their base counterpart. This suggests that adding dual meanings to the environment does not dramatically change the behavior of IB, consistent with recent results showing that IB complexity is not sensitive to synonymy (Bruneau–Bongard et al., 2025). Along with Manhattan-5-5, the behavior of the -Dual environments indicates that the convexity of the $p(u|m)$ does not play a dramatic role in the quasi-convexity of $q(m|w)$ for the environment's encoders.

The non-convex prior environments exhibit an important distinction: all NPUM environments exhibit a positive but weaker correlation between optimality and quasi-convexity and, unlike the CP- variants, positive correlations with the two IB metrics. NPDM and its dual variant show a strong negative correlation between optimality and convexity, reflected in the optimal encoders there all having nearly-0.0 quasi-convexity.

Environment	Optimality	$I(M;W)$	$I(W;U)$
CPUM	0.860	-0.169	-0.074
NPUM	0.238	0.791	0.9
CPDM	0.751	-0.907	-0.879
NPDM	-0.935	0.736	0.295
CPUM-Dual	0.855	-0.193	-0.092
NPUM-Dual	0.226	0.791	0.901
CPDM-Dual	0.752	-0.907	-0.872
NPDM-Dual	-0.934	0.736	0.289
CPDM-Convex	0.869	-0.381	-0.257
CPDM-Adj	0.882	-0.619	-0.423
NPDM-Adj	0.889	-0.641	-0.380
Manhattan-5-5	0.811	-0.242	-0.188

Table 3: Pearson's correlation coefficients between the quasi-convexity of $q(m|w)$ and Optimality, Complexity ($I(M;W)$), and Accuracy ($I(W;U)$) for environments in Experiment 2. The p values for all correlation coefficients are less than 0.00.

These results connect to the color naming scenario: the correlation coefficients for the NPUM environments are quite similar to the coefficients for the $q(m|w)$ distributions in the color naming environment in Table 1. This is further substantiated by the fact that all meanings are unique in the color naming environment and that $\mathsf{dcon}(p(m)) = 0.705$, which, while more convex than the prior distribution of NPUM (where $\mathsf{dcon}(p(m)) = 0.206$), is still relatively non-convex.

Discussion

Through two experiments connecting the information bottleneck framework for efficient communication to a generalization of the convexity universal, we have shown the following. In color naming, there is a non-trivial correlation between convexity and optimality, suggesting that efficient communication can partially explain convexity. We also found that natural languages are slightly more convex than other suboptimal encoders, even when controlling for optimality, accuracy, and complexity. By manipulating various components of the IB framework, we found that the convexity of the prior distribution (with unique meanings) is the most significant factor in driving a correlation between convexity and optimality.

Concurrently with this work, Koshevoy and Szymanik (2025) have also explored the connection between convexity and optimality in color naming, finding a strong and convincing correlation. Our work, while consistent with their results, differs in a few important ways. They rely on the partition-generating algorithm and measure of degree of convexity from Steinert-Threlkeld and Szymanik (2020), both of which apply to meanings with 'hard' boundaries. One of our main contributions generalizes the notion of the degree of convexity to probabilistic encoders of the kind used in IB. Similarly, their primary analysis operationalizes complexity and accuracy in slightly different ways, while we focus only on the IB framework. To this end, Bruneau–Bongard et al. (2025) compare multiple operationalizations of complexity (including the earlier degree of convexity) as well as combining more factors than just complexity and accuracy, in the color naming domain. Also currently with this work, Imel and Zaslavasky (2026) analyze the relationship between convexity and optimality using the same notion of convexity as the two works above. Like the present paper, they utilize the IB system for analyzing encoders; they show, among other things, that neither IB-optimality nor convexity entail the other and argue that the former is more fundamental. In addition to the more fitting definition of degree of convexity, the present paper also moves beyond only color naming in using constructed domains to tackle the question of which factors underlie the observed correlation between convexity and optimality.

Carlsson et al. (2024) show that a model of cultural evolution capturing both iterated learning and communication amongst neural agents produces color naming systems that are both IB-optimal and human-like. Implicit in their studies is also the production of systems which are convex but inefficient, suggesting that convexity is necessary but not sufficient for distinguishing natural from unnatural color naming systems. Our work—including cases with dual meanings where convexity and efficiency are anti-correlated—complements this perspective well and extends it beyond color naming in trying to identify *when* it is that efficient systems are also convex.

Future work can expand the results here along several axes. The analyses in Experiment 2 can be made more systematic by manipulating factors like convexity of the priors in a continuous manner and attempting to predict the strength of the correlation between convexity and optimality from them. Similarly, these analyses can be applied to a wider range of semantic spaces, including others that have been analyzed in the IB framework. Finally, it may be possible to prove analytic results in the vicinity of our experimental results. Our CPDM case—as well as the similar construction in Imel and Zaslavasky (2026)—shows that convex priors does not suffice. One candidate, refined conjecture: with convex priors and *unique* meanings, all IB-optimal encoders are convex. More generally, the methods here could be used to help adjudicate between different operationalizations of complexity and accuracy in efficient communication: if some more robustly generate a correlation with semantic universals across other semantic domains and manipulations of other parameters, it suggests that they are more likely to be explanatory factors.

By analyzing which factors underlie a connection between efficient communication and a prominent semantic universal, this paper takes a first step in deepening the kinds of explanations in computational approaches to semantic typology.

References

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press.

Bruneau–Bongard, J., Chemla, E., & Brochhagen, T. (2025). Assessing Pressures Shaping Natural Language Lexica.

Cognitive Science, *49*(12), e70145. https://doi.org/10.1111/cogs.70145

Carlsson, E., Dubhashi, D., & Regier, T. (2024). Cultural evolution via iterated learning and communication explains efficient color naming systems. *Journal of Language Evolution*, *9*(1–2), 49–66. https://doi.org/10.1093/jole/lzae010

Chemla, E., Buccola, B., & Dautriche, I. (2019). Connecting Content and Logical Words. *Journal of Semantics*, *36*(3), 531–547. https://doi.org/10.1093/jos/ffz001

Cook, R. S., Kay, P., & Regier, T. (2005). The world color survey database. https://doi.org/10.1016/b978-008044612-7/50064-0

Denić, M., Steinert-Threlkeld, S., & Szymanik, J. (2022). Indefinite Pronouns Optimize the Simplicity/Informativeness Trade-Off. *Cognitive Science*, *46*(5), e13142. https://doi.org/10.1111/cogs.13142

Douven, I. (2025). The learnability of natural concepts. *Mind & Language*, *40*(1), 120–135. https://doi.org/https://doi.org/10.1111/mila.12523

Gärdenfors, P. (2000). *Conceptual Spaces: The geometry of thought*. The MIT Press.

Gärdenfors, P. (2014). *The Geometry of Meaning*. The MIT Press.

Imel, N., Guo, Q., & Steinert-Threlkeld, S. (2026). An Efficient Communication Analysis of Modal Typology. *Open Mind*, *10*, 1–28. https://doi.org/10.1162/OPMI.a.313

Imel, N., & Zaslavasky, N. (2026). On convexity and efficiency in semantic systems. https://arxiv.org/abs/2602.08238

Jäger, G. (2010). Natural Color Categories Are Convex Sets. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, Language, and Meaning: Amsterdam Colloquium 2009* (pp. 11–20). https://doi.org/10.1007/978-3-642-14287-1_2

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, 1–23. https://doi.org/10.1146/annurev-linguistics-011817-045406

Koshevoy, A., & Szymanik, J. (2025, March). *Convexity (probably) makes languages efficient*. https://osf.io/preprints/psyarxiv/wtpke_v1

Maldonado, M., Culbertson, J., & Uegaki, W. (2022). Learnability and constraints on the semantics of clause-embedding predicates. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). Retrieved January 29, 2025, from https://escholarship.org/uc/item/9h13v9db

Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, *118*(49), e2025993118. https://doi.org/10.1073/pnas.2025993118

Piasini, E., Filipowicz, A. L. S., Levine, J., & Gold, J. I. (2021). Embo: A python package for empirical data analysis using the information bottleneck. *J. Open Res. Softw.*, *9*(1), 10. https://doi.org/10.5334/jors.322

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441. https://doi.org/10.1073/pnas.0610341104

Skinner, L. P. (2025). *Convexity is a Fundamental Feature of Efficient Semantic Compression in Probability Spaces.* [MSc Thesis]. University of Washington. https://hdl.handle.net/1773/53008

Steinert-Threlkeld, S. (2021). Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals. *Entropy*, *23*(10), 1335. https://doi.org/10.3390/e23101335

Steinert-Threlkeld, S., & Szymanik, J. (2019). Learnability and Semantic Universals. *Semantics & Pragmatics*, *12*(4). https://doi.org/10.3765/sp.12.4

Steinert-Threlkeld, S., & Szymanik, J. (2020). Ease of learning explains semantic universals. *Cognition*, *195*, 104076. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104076

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. *Proc. of the 37-th Annual Allerton Conference on Communication, Control and Computing*, 368–377. https://arxiv.org/abs/physics/0004057

Uegaki, W. (2023). The Informativeness/Complexity Trade-Off in the Domain of Boolean Connectives. *Linguistic Inquiry*, *55*(1), 174–196. https://doi.org/10.1162/ling_a_00461

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115

Zaslavsky, N., Maldonado, M., & Culbertson, J. (2021). Let's talk (efficiently) about us: Person systems achieve near-optimal compression. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. https://escholarship.org/uc/item/2sj4t8m3