The Integrated Information Bottleneck: A theoretically-grounded model unifying category- and system-level structure in communicative efficiency

Jai Aggarwal (jai@cs.toronto.edu)[1], Jaehyeon Park (jpark@cs.toronto.edu)[1], Barend Beekhuizen (barend.beekhuizen@utoronto.ca)[1,2] & Suzanne Stevenson (suzanne@cs.toronto.edu)[1]

[1]Department of Computer Science, University of Toronto, Toronto, Canada

[2]Department of Language Studies, University of Toronto Mississauga, Mississauga, Canada

Abstract

Much research has shown that lexical semantic systems are shaped by a drive for efficient communication, which yields desirable structure at the level of individual lexical categories (i.e., the semantic coherence of words). Less attention has been given to what we call *semantic systematicity*: structural regularity that exists across lexical categories, where different words make the same semantic distinctions. We develop the first model of the efficiency trade-off that captures both category- and system-level regularity using the same domain-general information-theoretic underpinning. We also propose a stricter evaluation method for comparing alternative formalizations of the efficiency trade-off. Our model that integrates the pressures for category-level structure and semantic systematicity explains crosslinguistic variation better than previous approaches in two semantic domains, shedding new light on how a pressure for efficiency shapes the structure of language.

Keywords: efficiency; information theory; semantic typology

Introduction

A drive for **efficient communication** has been shown to constrain crosslinguistic variation in how languages use words to carve up semantic domains into lexical semantic categories (e.g., Gibson et al., 2019; Kemp et al., 2018). In this view, languages optimally trade off between the two competing pressures of simplicity and informativity. For example, Spanish has separate pronouns for referring to the 2nd-person singular (*tú*) and plural (*vosotros*), while English uses a single word (*you*) to cover both. Here, English is simpler, by minimizing the cognitive load on speakers (with a single word), but Spanish is more informative, by enabling more precise communication to listeners. This efficiency trade-off has been shown to constrain crosslinguistic variation across a range of domains (e.g., Chen et al., 2023; Kemp & Regier, 2012; Regier et al., 2015; Xu et al., 2020).

Work on efficient communication has emphasized the role of *category-internal structure* – how each word refers to a contiguous (i.e., convex) part of the semantic space – in the efficient representation and interpretation of the set of words for a domain. Only recently has work also begun to explicitly address the role of higher-level structure *across categories* (Chen et al., 2023; Hallam, Jordan, et al., 2025), where different words consistently encode similar high-level semantic distinctions – a property we call **semantic systematicity**.

Figure 1 illustrates semantic systematicity in the domain of spatial deictic demonstratives. English is maximally semantically systematic: its words consistently distinguish the Source orientation from Goal and Place, and the Near distance from Medial and Far (see the green lines in the figure). In contrast, Unsystematic English makes different semantic distinctions between Near and Medial/Far locations depending on the orientation (compare the red lines); this leads to a kind of misalignment between the terms *here/there* and *from here/from there* that is generally not seen in attested languages.

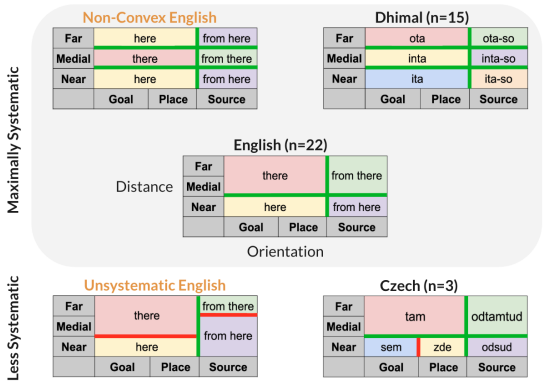

Figure 1: Variation in **semantic systematicity** in the spatial deixis domain. Green lines indicate consistent semantic distinctions, red lines indicate inconsistent distinctions that violate systematicity. Languages in orange are hypothetical.

While many languages exhibit semantic systematicity, such a pressure for higher-level structure cannot alone explain variation. First, not all real languages are maximally semantically systematic; for example, see Czech in Figure 1, where the red line indicates finer-grained distinctions for just one distance value. Moreover, if languages were *only* optimizing for higher-level structure, then systems like those of Dhimal and Non-Convex English in the figure should both be commonly attested. However, Non-Convex English is highly unlikely, as it lacks category-level structure: the words *here* and *from here* refer to non-convex regions of space. This suggests that pressures for both category-level structure *and* semantic systematicity constrain linguistic variation.

Despite its key role in supporting learnability, generalization, and communication (Ackerman & Malouf, 2013; Hallam, Jordan, et al., 2025; Kirby et al., 2015), semantic systematicity is rarely addressed in existing formalizations of efficient communication. The influential Information Bottleneck (IB) framework – as originally described in Tishby and 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*

(2015) and Zaslavsky et al. (2018) – has not been shown to account for structure above the level of individual categories. Other approaches are better able to capture semantic systematicity, but are either difficult to generalize across semantic domains (e.g., MDL-based approaches require a tailored description language for each domain; Kemp and Regier, 2012; Xu et al., 2020), or lack a unified approach for explaining both category- and system-level structure (Chen et al., 2023 [an extension to IB]; Hallam, Jordan, et al., 2025).

Building on preliminary work (Aggarwal et al., 2026), we present a novel information-theoretic account of crosslinguistic variation that models both category- and system-level structure in a mathematically principled, domain-general manner, couched in a generalized version of the IB framework (Slonim et al., 2006). We also propose a novel evaluation framework for comparing different formalizations of communicative efficiency that, unlike previous approaches, penalizes overly permissive models. Using this evaluation framework, we find that our integrated approach to the efficiency trade-off explains linguistic variation better than existing methods in two domains: spatial deictic demonstratives (Chen et al., 2023) and personal pronouns (Zaslavsky et al., 2021). Our findings shed new light on how pressures for cognitive efficiency interact to shape the structure of language.

Related Work

The Information Bottleneck (IB) framework (Tishby & Zaslavsky, 2015; Zaslavsky et al., 2018) is a domain-general formalization of the efficiency trade-off which encourages appropriate category-level structure, in the form of **convex categories** that group together maximally similar meanings (Gärdenfors, 2000; Koshevoy & Szymanik, 2025).[1] However, the standard formulation of the IB trade-off has not yet been shown to account for the cross-category regularities of semantic systematicity. This is to be expected; while the standard IB trade-off encourages grouping similar meanings together, it does not seem to enforce grouping similar *kinds* of meanings together *consistently* across the semantic space.

Minimum Description Length (MDL) approaches to efficiency have the potential to account for category-level structure as well as semantic systematicity: in some description languages, semantically systematic languages have a shorter (simpler) encoding, achieving greater efficiency (Kemp & Regier, 2012; Rissanen, 1989; Xu et al., 2020). However, not all description languages capture semantic systematicity; thus, the explanatory power of MDL approaches rests in the choice of description language, which is often tailored to a specific domain. A desirable goal is to reliably capture semantic systematicity in a more domain-general model component.

Some recent work on communicative efficiency has devised alternative measures of the kind of system-level regularity we call semantic systematicity. Both approaches tap into the notion of consistent splitting patterns across words illustrated in Figure 1, in the domains of spatial deictic demonstratives (Chen et al., 2023) and kinship (Hallam, Jordan, et al., 2025). However, both measures lack a common theoretical grounding with existing formalizations of the efficiency trade-off. Here, we propose a domain-general, mathematically-principled formulation of efficiency that captures both category- and system-level structure in a uniform way.

[1] IB computes meaning similarity in the semantic space of referents; we measure convexity in this space as well. Recent work has shown that IB may not encourage convexity if convexity is instead measured in a meaning space that differs from the referent space (Imel & Zaslavasky, 2026; Ranjan & Steinert-Threlkeld, 2026).

Computational Framework

All relevant code and data can be found here: https://github.com/jaikaggarwal/cogsci_2026_integrated_ib.

Background on Standard IB

To build the intuition for our approach, we first review the IB framework, which models the trade-off between simplicity and informativity as an information-theoretic trade-off between compression and prediction, respectively.

In IB, each meaning $m \in M$, the set of all meanings, is a noisy mental representation of an intended referent $u \in U$, the universe of all referents (i.e., $|M| = |U|$). The aim is to compress the meanings M into a set of words W (where maximum compression requires one word that refers to all meanings), while ensuring the words are predictive of the referents U (where maximum predictiveness requires one word per meaning). Compression is measured as $I(M;W)$ – the mutual information between meanings and words – which decreases with greater compression. The goal is to minimize $I(M;W)$; the more compressed the meanings are into words, the simpler the system, as speakers need to remember fewer bits of information. Prediction is measured as $I(W;U)$ – the mutual information between words and referents – which increases with greater predictiveness. The goal is to maximize $I(W;U)$; the more predictive words are of referents, the more informative they are to the listener. Languages are optimal if they are as informative as possible at a given level of simplicity (or vice versa). The goal then in IB is to minimize:

$$J_{IB} = I(M;W) - \beta I(W;U) \tag{1}$$

The trade-off is parameterized by β, a language-specific bias for prioritizing informativity over simplicity in the domain. Higher β values may arise due to increased cultural importance for precise communication in a domain, leading to finer-grained (less compressed) categories (Zaslavsky et al., 2020).

Capturing Semantic Systematicity

As previously mentioned, the IB framework is unable to capture semantic systematicity: higher-level regularity in how words map to meanings across the system. We propose accounting for such regularity by considering **higher-level semantic features** that capture semantic distinctions relevant to a language. For example, Figure 1 shows that English consistently distinguishes nearby locations (using *here* and *from here*) from others (using *there* and *from there*). Using higher-level features NEAR = {Near} and NOT NEAR =

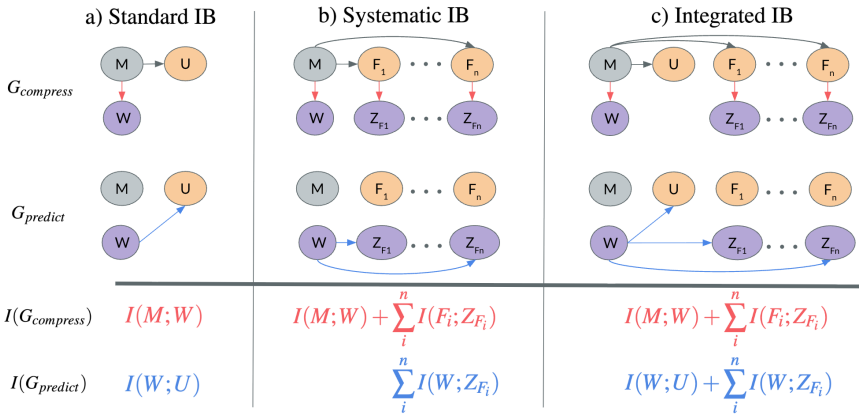

Figure 2: Generalized IB: Bayes graphs and their mutual information terms for each model we consider. Orange variables have language-independent relationships with meanings M (in grey). Purple variables represent language-specific compressions of another variable. Grey arrows represent relationships that are constant with respect to the compression problem, red arrows (and text) represent compression relationships, and blue arrows (and text) represent predictive relationships.

{Medial, Far}, we can see that English is systematic because its words are informative about these higher-level semantic distinctions. However, the words in a less systematic language, like Czech or Unsystematic English in Figure 1, cannot be consistently mapped to any such higher-level features.

As in Standard IB, this intuition can also explain linguistic variation as a simplicity/informativity trade-off, but one that operates over higher-level features instead of referents. For example, languages like Dhimal in Figure 1 may prioritize being informative about distance more than English, leading to finer-grained semantic distinctions along this dimension (this effect is analogous to having a higher β value in the IB framework). However, this gain in informativity requires Dhimal speakers to trade off simplicity, as they need to remember a larger set of higher-level semantic features (one per distance level). While this intuition broadly mirrors the structure of the Standard IB trade-off, accounting for the relationships between meanings, words, features, and higher-level features requires a more general modelling framework.

Formalizing Systematicity with Generalized IB

Here, we present a theoretically-grounded model that draws on the Generalized IB (GIB) framework (Slonim et al., 2006), which can account for probabilistic dependencies between larger sets of variables than Standard IB. GIB formalizes the problems of compression and prediction using two Bayes graphs: $G_{compress}$ specifies the compression problem, and $G_{predict}$ specifies the predictive information to be maintained after compression. In generalizing Standard IB, the efficiency trade-off here requires minimizing $I(G_{compress})$ while maximizing $I(G_{predict})$, where $I(G)$ is simply the sum of the mutual information terms between each variable and its immediate parent in G (Slonim et al., 2006). This is modelled as the following minimization problem, which generalizes the Standard IB objective in Eq. 1, with constraint β again capturing the language-specific importance of prediction (informativity) over compression (simplicity):

$$J_{GIB} = I(G_{compress}) - \beta I(G_{predict}) \quad (2)$$

Slonim et al. (2006) show how Standard IB can be derived in this framework. As seen in Figure 2a, $G_{compress}$ specifies the Standard IB compression problem: each language compresses meanings M into words W (see red arrow), where meanings M have a universal mapping to referents U (see grey arrow), which is constant with respect to compression. The total information $I(G_{compress})$ is thus $I(M;W)$. $G_{predict}$ specifies that after compression, words W should retain as much information as possible about the referents U. The total information $I(G_{predict})$ is $I(W;U)$. Substituting these values into Eq. 2 yields the Standard IB objective function in Eq. 1.

We can now turn to Figure 2b, which illustrates our model that optimizes semantic systematicity. Here, $G_{compress}$ specifies three kinds of relationships. First, analogous to the mapping in Standard IB of meanings M to referents, here meanings have a universal mapping to their n component features; i.e., each $m \in M$ maps to the value it has in semantic dimension F_i. $G_{compress}$ also specifies two language-specific compressions: of meanings M into words W (as in Standard IB), and of each dimension F_i into a set of higher-level features Z_{F_i}. In line with the intuition presented earlier about these high-level features, $G_{predict}$ specifies that words W should be as informative as possible about Z_{F_i}.[2] Substituting $I(G_{compress})$ and $I(G_{predict})$ for these graphs into Eq. 2 yields our Systematic IB model (Aggarwal et al., 2026):

$$J_{SIB} = \underbrace{I(M;W) + \sum_i^n I(F_i;Z_{F_i})}_{I(G_{compress})} - \beta \underbrace{\sum_i^n I(W;Z_{F_i})}_{I(G_{predict})} \quad (3)$$

As a concrete example of how this model captures semantic systematicity, consider the spatial deixis domain, characterized by $F_1 = D$(istance) and $F_2 = O$(rientation); here, we

[2]This use of mutually informative clusters (W and Z_F) can be seen as an instance of **co-clustering** (Kemp et al., 2006).

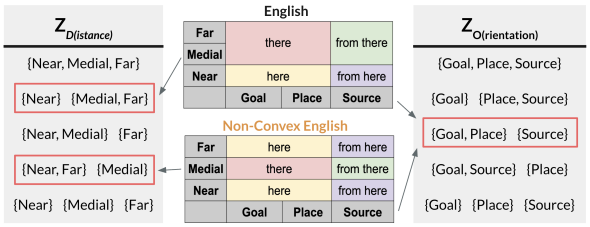

Figure 3: The grey boxes list all possible clusterings of features F into higher-level features Z_F on each feature dimension. The optimal Z_F for English and Non-Convex English on each feature dimension is shown in red boxes.

focus on the distance dimension D. In the universal mapping between M and D, each meaning has the same feature value across languages (e.g., m = Near+Goal always maps to the feature d_m = Near). The main difference across languages is how they compress D into Z_D, as exemplified in Figure 3. The red boxes in the Z_D column show that English and Non-Convex English differ in the higher-level features that best fit their set of words W. As the words in both languages predict their higher-level features perfectly, they both optimize Eq 3, which only captures the pressure for semantic systematicity. Importantly, only English would also optimize the Standard IB equation Eq. 1, as Non-Convex English has poor category-level structure, with words covering disjoint semantics.

Integrating Category Structure and Systematicity

Given that lexical semantic systems have been shown to exhibit both category-level structure and semantic systematicity (Chen et al., 2023; Hallam, Jordan, et al., 2025), we propose a novel model that can account for both pressures simultaneously, by integrating the Bayes graphs for Standard IB and Systematic IB; see Figure 2c. Substituting the relevant $I(G_{compress})$ and $I(G_{predict})$ terms from the graphs in Figure 2c into Eq. 2, and introducing a new parameter α (explained just below), gives rise to our **Integrated IB** model:

$$J_{IIB} = \underbrace{I(M;W) + \sum_{i}^{n} I(F_i; Z_{F_i}) -}_{I(G_{compress})}$$
$$\underbrace{\beta[(1-\alpha)I(W;U) + \alpha[\sum_{i}^{n} I(W; Z_{F_i})]]}_{I(G_{predict})} \quad (4)$$

Here, given a fixed number of bits to compress the system, languages should be maximally predictive about both referents (promoting category-level structure) and higher-level features (promoting semantic systematicity). A question then arises of whether a pressure for one kind of structure plays a greater role in driving crosslinguistic variation. To quantify this, we introduce the parameter α in Eq. 4. While Slonim et al. (2006) only consider β, which captures the importance of prediction over compression, the introduction of α helps isolate the importance of the **kind** of information to be predicted. Specifically, do languages prioritize finer-grained prediction about individual referents (indicated by a lower α) or higher-level semantic distinctions (indicated by a higher α)? We assume that α is constant across languages to investigate the extent to which there is a universal cognitive preference for semantic systematicity. We also explore whether varying α across semantic domains reveals domain-specific biases towards systematicity.

Note that Integrated IB yields Standard IB and Systematic IB at the extreme values of $\alpha = 0$ and $\alpha = 1$, respectively.[3]

Evaluation Framework

To assess whether the efficiency trade-off constrains linguistic variation, prior work tests whether attested languages are closer to optimal than a set of hypothetical systems on average (e.g., Zaslavsky et al., 2018, 2021). However, a model can meet this standard while being *underconstrained*: it may assess a large number of hypothetical (i.e., unattested) languages as near-optimal. We propose an evaluation in which an apt set of constraints should find *both* that attested systems are near-optimal (i.e., have high recall), *and* that near-optimal systems are truly attested (i.e., have high precision).[4]

Formally, we first compute a language's **deviation from optimality** ϵ_L, where $\epsilon_L = \min_\beta \frac{1}{\beta}[J^L_\beta - J^*_\beta]$, and J^*_β is the objective function value achieved by a maximally optimal language at a particular β (Zaslavsky et al., 2018).[5] To capture precision and recall, we frame the efficiency trade-off as a binary classification task, where systems within a certain deviation from optimality ϵ are classified as positive (expected to exist), and the rest are classified as negative (not expected to exist). Near-optimal attested systems are true positives, while near-optimal hypotheticals are false positives. We then produce precision-recall (PR) curves by computing the precision and recall values for each model at increasing values of ϵ.[6]

To explore whether category-level structure or semantic systematicity plays a more important role in shaping linguistic systems, we produce PR curves for J_{IIB} with α values in the range [0, 1]. We compare model performance both by (1) visually comparing PR curves (better curves have higher precision across a large range of recall values, leading to higher arches), and (2) quantitatively comparing the best F1 score of each model across its PR curve. If models with higher α values achieve better PR results, this suggests that systematicity plays a more important role in explaining variation than category-level structure.

[3] At $\alpha = 0$, J_{IIB} is minimized when each $I(F_i; Z_{F_i}) = 0$ (as there is no longer a pressure to maximize $I(W; Z_{F_i})$), yielding J_{IB}.

[4] Our evaluation may be overly strict, as some near-optimal hypotheticals may correspond to realistic languages, but are simply not attested. In ongoing work, we are analyzing the naturalness of near-optimal hypotheticals as a way to balance this evaluation approach.

[5] We do not compute the theoretically optimal curve; we approximate ϵ_L using the set of most-optimal systems in each case study.

[6] Bruneau–Bongard et al. (2025) recently proposed a similar evaluation framework, using ROC curves instead of PR curves. PR curves are more appropriate in settings like ours, and others (e.g., Chen et al., 2023; Kemp & Regier, 2012), where the number of negative examples far outweighs the number of positive examples.

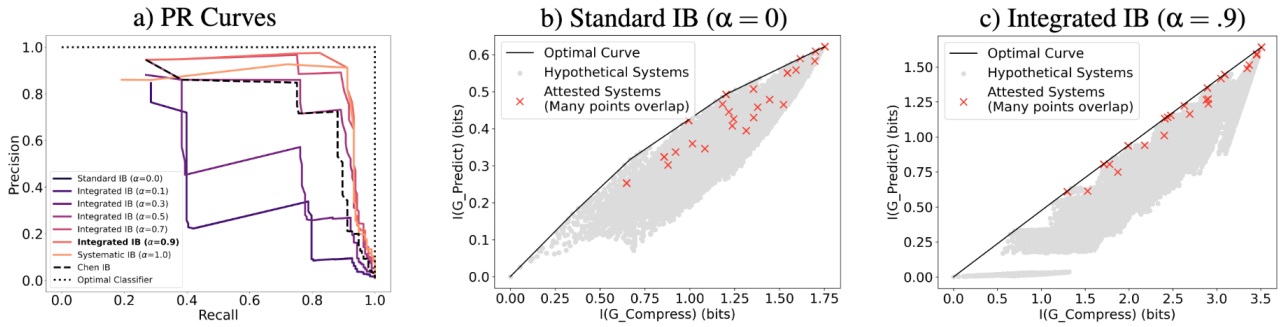

Figure 4: Spatial Deixis. (a) PR curves for Integrated IB at various values of α ($\alpha = 0$ and $\alpha = 0.1$ overlap). The simplicity-informativity trade-offs are visualized for (b) Standard IB ($\alpha = 0$) and (c) the best performing model, Integrated IB at $\alpha = .9$.

Spatial Deictic Demonstratives

Our first case study aims to explain crosslinguistic variation in the domain of spatial deictic demonstratives (cf. Figure 1). Chen et al. (2023) provide preliminary evidence that semantic systematicity constrains variation in this domain. Here, we show that our theoretically-grounded model accounts for attested variation better than both Standard IB and the model used by Chen et al. (2023) (referred to as Chen IB below).

Model Instantiation Standard IB requires two probability distributions: the need probabilities $p(m)$ (the relative frequency of needing to refer to each meaning), and the meaning-referent mapping $p(u|m)$ (the probability that a meaning represents a particular referent). $p(u|m)$ is defined such that it captures the similarity between referents in a semantic space (Zaslavsky et al., 2018). We define $p(m)$ and $p(u|m)$ as in Chen et al. (2023) for all three models.

Chen IB extends Standard IB with the term $S[W : M]$, defined as the number of distinct higher-level patterns a language exhibits in each feature dimension. For example, in Figure 1, Unsystematic English has two distinct patterns in the Distance dimension, while English only has one. Adding $S[W : M]$ to Eq. 1 penalizes less systematic languages, but loses the information-theoretic grounding of Standard IB.

Our Integrated IB model captures semantic systematicity in a more integrated manner by instantiating the J_{IIB} formula in Eq. 4 with $F_1 = D$(istance) and $F_2 = O$(rientation); Figure 3 shows the possible Z_D and Z_O compressions.

Data We consider 193 of the languages used by Chen et al. (2023), drawn from typological data compiled by Nintemann et al. (2020).[7] We compare the efficiency of these systems to the exhaustive set of 21, 147 hypothetically possible partitions of the space of 9 meanings.

Analyses & Results We compare the explanatory power of the Integrated IB model at 11 values of α, evenly spaced across [0, 1]. If semantic systematicity plays a meaningful role in shaping linguistic variation, then the best PR performance should occur when α is substantially higher than 0.

Figure 4a provides evidence in favour of this. We see that as α increases, the PR curves steadily improve (i.e., have higher-arches) suggesting that languages are optimizing for semantic systematicity. However, languages do not seem to purely optimize for semantic systematicity, as the best results are achieved when $\alpha = .9$ in Integrated IB (best F1= .91), indicating that a slight pressure for category-level structure remains. The lowest precision occurs when $0 \leq \alpha \leq .1$ ($\alpha = 0$ is the Standard IB model, best F1= .51). Figure 4b shows that under Standard IB, the attested systems are fairly close to optimal, but a large number of hypotheticals are near-optimal as well. Conversely, Figure 4c shows that when prioritizing semantic systematicity within Integrated IB ($\alpha = .9$), far fewer hypotheticals are considered as near-optimal as the attested systems, thereby increasing precision.

While Chen IB (dashed line in Figure 4a; best F1= .80) exhibits a much better PR curve than Standard IB, it performs worse than our Integrated model when $.5 \leq \alpha \leq 1$. Manual inspection reveals that this occurs because their semantic systematicity term $S[W : M]$ outweighs the Standard IB trade-off. This captures semantic systematicity at the expense of **convexity**, a key property of attested categories.

To show that our model simultaneously captures both convexity and semantic systematicity, we perform a follow-up analysis: for each model, we explore whether attested and hypothetical languages that are closer to optimal are more convex by computing the Spearman correlation between each language's deviation from optimality (ϵ_L) and its convexity (computed using the method of Steinert-Threlkeld and Szymanik, 2020). We find that our best model, Integrated IB at $\alpha = .9$, captures convexity (with $\rho = -.60$) almost as well as Standard IB ($\rho = -.64$), and much better than the Chen IB model ($\rho = -.20$); all correlations are significant at $p \ll .001$.

Taken together, these results demonstrate the explanatory power that our Integrated IB model obtains over previous work by simultaneously accounting for both category-level structure and semantic systematicity.[8]

[7] We removed 27 languages from Chen et al. (2023), as they contained N/A values that can misrepresent a language's structure.

[8] In general, we find that our model captures convexity (and thus, category-level structure) well across domains.

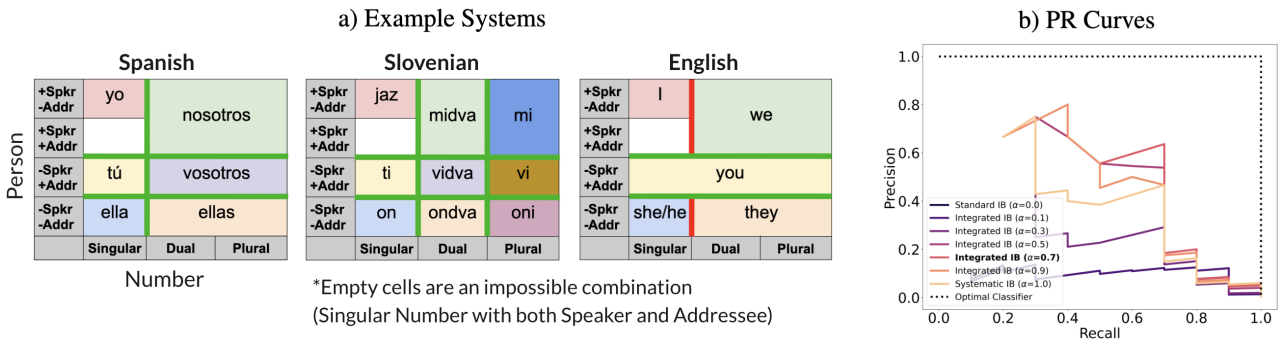

Figure 5: Personal pronouns domain. a) Example systems. Green and red lines can be interpreted as in Figure 1. b) PR curves for Integrated IB at various values of α ($\alpha = 0$ and $\alpha = 0.1$ overlap). See code repository for simplicity-informativity trade-offs.

Personal Pronouns

Here, we turn to the domain of personal pronouns, illustrated in Figure 5a. While Zaslavsky et al. (2021) show that attested pronoun systems exhibit optimal category-level structure using Standard IB, we aim to show that our Integrated IB model (which also captures semantic systematicity) readily generalizes to this domain, and can better account for the data.

Model Instantiation We derive the need probability distribution $p(m)$ from data released by Zaslavsky et al. (2021) (see their paper for details). To define the meaning-referent mapping $p(u|m)$, we adapt the Egocentric model from Zaslavsky et al. (2021), modelling grammatical number as a three-valued ordinal feature (singular, dual, plural), but simplifying the representation of grammatical person to the two features they found to be important: those that indicate the presence/absence of the speaker and addressee (see Figure 5a).[9]

To capture semantic systematicity within Integrated IB, we instantiate the formula in Eq. 4 with $F_1 = S$(peaker), $F_2 = A$(ddressee), and $F_3 = N$(umber).

Data We use the dataset compiled by Zaslavsky et al. (2021), of 10 person systems from well-documented languages across various language families. We also follow the same procedure as in their work to generate our 2101 hypothetical systems.

Analyses & Results We investigate whether models with higher values of α (indicating a greater pressure for semantic systematicity) achieve better PR curves than the Standard IB model ($\alpha = 0$). Figure 5b illustrates that our Integrated IB model again shows better PR results than Standard IB; the PR curve is worst when $0 \leq \alpha \leq .1$ (best F1= .22), and improves as α increases until $.6 \leq \alpha \leq .8$ (Figure 5b shows $\alpha = .7$, best F1= .67). This suggests that semantic systematicity plays an important role in determining the structure of pronoun systems, though to a lesser extent than in the domain of spatial deixis (where $\alpha = .9$); we expand on this in the Discussion.

[9]Defining $p(u|m)$ as in previous work yields similar results.

Discussion

We develop a novel information-theoretic objective function that jointly captures pressures for category-level structure and semantic systematicity in a theoretically-unified framework. In two semantic domains, we find that both pressures are necessary to best explain patterns of crosslinguistic variation, providing a more complete view of how a drive for cognitive efficiency shapes lexical semantic systems.

The explicit weighting of the two pressures in our model (using the parameter α) revealed that semantic systematicity played a weaker role in pronouns than in spatial deixis. This raises the question of *why* domains might differ in their reliance on systematicity. For our two domains, we suggest this difference may be due to social forces counteracting systematicity, given that pronouns communicate a speaker's relation to others. For example, English became less systematic over time when it lost the distinction between second person singular and non-singular (as seen in Figure 5a), likely due to social pressure to use the formerly non-singular *you* to convey politeness to a single addressee (Dury, 2011). This illustrates how social factors may need to trade-off against cognitive factors in efficiency analyses (cf. Aggarwal et al., 2024).

Another key next step is to extend the efficiency trade-off to account for compositionality (Gibson et al., 2019), where parts of word forms systematically convey particular semantic distinctions. For example, in Figure 1, Dhimal might be considered simpler if we analyze it at the level of morphemes, as it systematically reuses the suffix *-so* to convey the Source orientation. Though such reuse has been shown to support efficient communication in principle (Hallam, Kirby, et al., 2025; Kirby et al., 2015), compositionality has yet to be integrated directly into the efficiency trade-off. Our model has the potential to capture compositionality by means of an additional layer in the Bayes graphs: analogously to how we imposed additional organization on the semantics side, by compressing features into higher-level semantic distinctions, we can also augment the Bayes graphs to relate word forms to their component pieces. In this way, our model could extend the efficiency trade-off to jointly capture similar processes over both meaning and form.

Acknowledgements

This research was made possible by an NSERC Discovery Grant to BB (RGPIN-2019-06917).

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 429–464.

Aggarwal, J., Park, J., Stevenson, S., & Beekhuizen, B. (2026). Semantic systematicity: A new perspective on compressibility and expressivity in iterated learning. In S. Hartmann, M. Sibierska, M. Fröhlich, Y. Jadoul, M. Josserand, T. Matzinger, K. Mudd, J. Nölle, M. Pleyer, S. Wacewicz, & P. Żywiczyński (Eds.), *The Evolution of Language: Proceedings of the 16th International Conference (EVOLANG XVI)*. https://doi.org/10.17617/2.3696655

Aggarwal, J., Watson, J., Senapati, P., & Stevenson, S. (2024). The (in) efficiency of within-language variation in online communities. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Bruneau–Bongard, J., Chemla, E., & Brochhagen, T. (2025). Assessing pressures shaping natural language lexica. *Cognitive Science*, *49*(12), e70145.

Chen, S., Futrell, R., & Mahowald, K. (2023). An information-theoretic approach to the typology of spatial demonstratives. *Cognition*, *240*, 105505.

Dury, R. (2011). You and thou in early modern english: Cross-linguistic perspectives. *Germanic Language Histories 'from Below'(1700-2000)*, *86*, 129.

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

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.

Hallam, M., Jordan, F. M., Kirby, S., & Smith, K. (2025). Predictive structure emerges during the generalisation of kin terms to new referents. *Open Mind*, *9*, 1431–1466.

Hallam, M., Kirby, S., Jordan, F., & Smith, K. (2025). Efficient communication drives the semantic structure of kinship terminology. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*.

Imel, N., & Zaslavasky, N. (2026). On convexity and efficiency in semantic systems. *arXiv preprint arXiv:2602.08238*.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. *Proceedings of the 21st National Conference on Artificial Intelligence*, *1*.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

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

Koshevoy, A., & Szymanik, J. (2025). Convexity (probably) makes languages efficient.

Nintemann, J., Robbers, M., & Hober, N. (2020). *Here–hither–hence and related categories: A cross-linguistic study* (Vol. 26). Walter de Gruyter GmbH & Co KG.

Ranjan, A., & Steinert-Threlkeld, S. (2026). When efficient communication explains convexity. *arXiv preprint arXiv:2602.02821*.

Regier, T., Kemp, C., & Kay, P. (2015). 11 word meanings across languages support efficient communication. *The Handbook of Language Emergence*, *87*, 237.

Rissanen, J. (1989). Stochastic complexity in statistical inquiry. *World Scientific, Series in Computer Science*, *15*.

Slonim, N., Friedman, N., & Tishby, N. (2006). Multivariate information bottleneck. *Neural Computation*, *18*(8), 1739–1789.

Steinert-Threlkeld, S., & Szymanik, J. (2020). Ease of learning explains semantic universals. *Cognition*, *195*, 104076.

Tishby, N., & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. *2015 IEEE Information Theory Workshop (ITW)*, 1–5.

Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, *4*, 57–70.

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.

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2020). Communicative need in colour naming. *Cognitive Neuropsychology*, *37*(5-6), 312–324.

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*(43).