



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Let's talk (efficiently) about us

Person systems achieve near-optimal compression

Citation for published version:

Zaslavsky, N, Maldonado, M & Culbertson, J 2021, Let's talk (efficiently) about us: Person systems achieve near-optimal compression. in *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society: Comparative Cognition: Animal Minds, CogSci 2021*. vol. 43, Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 43, The Cognitive Science Society, pp. 938-944, 43rd Annual Meeting of the Cognitive Science Society: Comparative Cognition: Animal Minds, CogSci 2021, Virtual, Online, Austria, 26/07/21. <<https://escholarship.org/uc/item/2sj4t8m3>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Proceedings of the 43rd Annual Meeting of the Cognitive Science Society

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Let's talk (efficiently) about us: Person systems achieve near-optimal compression

Noga Zaslavsky* (nogazs@mit.edu)
Department of Brain and Cognitive Sciences
Center for Brains Minds and Machines
Massachusetts Institute of Technology

Mora Maldonado* (mora.maldonado@upf.edu)
Departament de Traducció i Ciències del Llenguatge
Universitat Pompeu Fabra

Jennifer Culbertson (jennifer.culbertson@ed.ac.uk)
Centre for Language Evolution
University of Edinburgh

Abstract

Systems of personal pronouns (e.g., ‘you’ and ‘I’) vary widely across languages, but at the same time not all possible systems are attested. Linguistic theories have generally accounted for this in terms of strong grammatical constraints, but recent experimental work challenges this view. Here, we take a novel approach to understanding personal pronoun systems by invoking a recent information-theoretic framework for semantic systems that predicts that languages efficiently compress meanings into forms. We find that a test set of cross-linguistically attested personal pronoun systems achieves near-optimal compression, supporting the hypothesis that efficient compression shapes semantic systems. Further, our best-fitting model includes an egocentric bias that favors a salient speaker representation, accounting for a well-known typological generalization of person systems (‘Zwicky’s Generalization’) without the need for a hard grammatical constraint.

Keywords: person systems; pronouns; semantic typology; information theory; efficient coding

Introduction

Systems of personal pronouns, or person systems for short, categorize entities based on their conversational role: speaker (s), addressee (a), or non-participant others (o, i.e., entities without an active role in the conversation). Personal pronoun paradigms are a universal device by which languages express these categories (Cysouw, 2003). The person category of a given personal pronoun specifies its focal referent. For example, the first person pronouns ‘I’ and ‘we’ have the same focal referent (i.e., the speaker). Person distinctions also interact with number; while the pronoun ‘I’ refers only to the speaker, ‘we’ refers to a group which includes the speaker.

As in other semantic domains, like color and kinship, pronominal paradigms vary widely across languages, but at the same time not all attested systems are equally likely. To see this, assume (as we will throughout) that there are three basic person roles, s, a, and o, which interact with three number categories, exactly one, exactly two, and more than two. Assuming a single speaker, a single addressee, but potentially any number of others,¹ the basic person roles can be combined to form person referents with one or more entities (e.g., the inclusive meaning s+a). The resulting semantic space,

shown in Figure 1A, has 11 referents. Figure 1B shows that languages can express all referents using unique forms (as in Māori); or they can feature homophony such that some meanings are expressed by the same form (as in English); or they can feature redundancy (as in Mandarin, where the inclusive meaning can be expressed by ‘zánmen’ or ‘wǒmen’). Given this space and the possibility of homophony, there are a whopping 678,570 logically possible paradigms (B_{11}). And yet very few seem to be attested across the world’s languages (Cysouw, 2003).

What characterizes the unattested systems? One possibility is that some homophony patterns are more common than others. For example, Zwicky (1977) noted that homophony in the expression of the inclusive meaning *you and us* appears to be restricted. In many languages, this meaning is homophonous with the first person plural form (e.g., ‘we’ in English). By contrast, in no known language is the inclusive meaning homophonous with the second or third person form. We refer to the common system as first-inclusive and to the unattested systems as second- and third-inclusive respectively (see Figure 1B).

Linguistic theories tend to account for typological regularities like these by positing inviolable grammatical constraints on possible person systems (Harbour, 2016; Harley & Ritter, 2002, *inter alia*). These constraints, often conceived of as reflecting characteristics of our language faculty, are designed to account for existing person systems, while preventing the derivation of systems which are believed to be unattested. For example, Harbour (2016) accounts for Zwicky’s observation by positing a “hard-coded” asymmetry in the linguistic representation of the speaker and addressee. However, recent experimental results challenge the strong nature of these constraints, pointing instead to the possibility that soft biases shape person typology (Maldonado & Culbertson, 2020).

Here, we take a different approach to understanding pronominal paradigms, using a recent information-theoretic framework for semantic systems (Zaslavsky, Kemp, Regier, & Tishby, 2018; Zaslavsky, 2020). This approach hypothesizes that languages evolve under pressure to efficiently compress meanings into forms by optimizing an information-theoretic tradeoff between the complexity and communicative accuracy of the lexicon, known as the Information Bottleneck principle (IB; Tishby, Pereira, & Bialek, 1999). This principle is grounded in Rate–Distortion theory (RDT; Shannon,

* Equal contribution.

¹Following Harbour (2016), this assumption is motivated primarily by the observation that no known language has a unique form for multiple speakers or for multiple addressees (but see Bobaljik, 2008; Sonnaert, 2018).

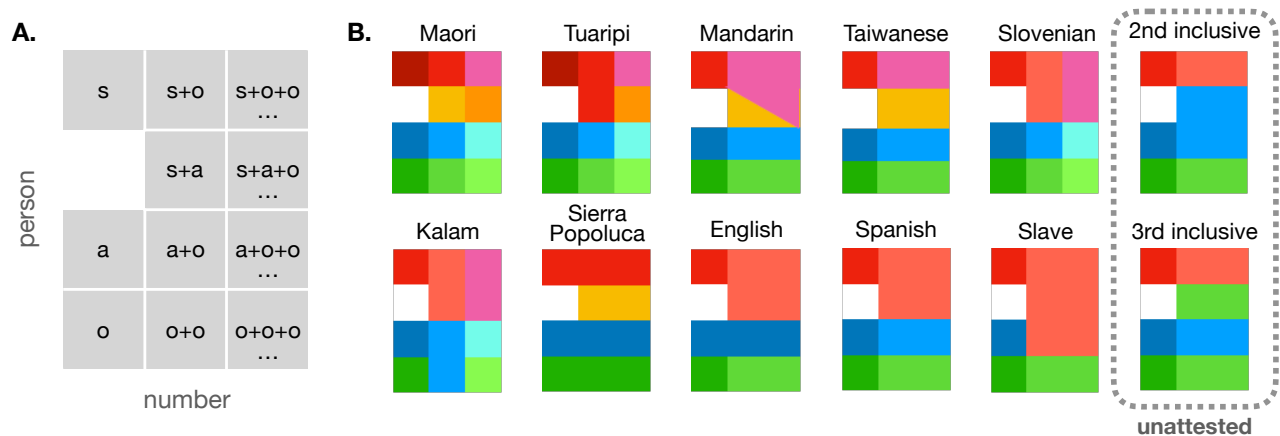


Figure 1: **A.** The person space consists of the 11 possible referents shown in this table. Rows correspond to person distinctions: speaker (s), addressee (a), and other (o). Columns correspond to number distinctions: exactly one, exactly two, and more than 2. **B.** Person systems derived from the pronominal systems of 10 well-documented languages, and two systems that are believed to be unattested, plotted against the person space shown in A. Colors correspond to distinct pronominal forms.

1948), the branch of information theory that characterizes efficient data compression, and has been shown to explain cross-linguistic patterns in several semantic domains, such as color and container naming (Zaslavsky et al., 2018; Zaslavsky, Regier, Tishby, & Kemp, 2019a). It is also closely related to a prominent notion of communicative efficiency in semantic typology (Kemp, Xu, & Regier, 2018), which is not grounded in RDT but has been applied to domains that are qualitatively more similar to person, such as kinship (Kemp & Regier, 2012), numeral systems (Xu, Liu, & Regier, 2020), quantifiers (Steinert-Threlkeld, 2020), and indefinite pronouns (Denic, Steinert-Threlkeld, & Szymanik, 2020). Therefore, the person domain poses an important test case for the broader applicability of RDT to the lexicon.

To summarize: we find that a cross-linguistic test set of personal pronoun systems achieves near-optimal compression, suggesting that these systems have indeed evolved to be efficient in the RDT sense. Importantly, our model is also able to account for Zwicky’s observation about inclusive systems, by including a (soft) egocentric bias, i.e. a bias towards keeping the speaker distinct, similar in spirit to inviolable constraints proposed in the linguistics literature.

Typological sample of personal pronoun systems

To investigate personal pronoun systems in the RDT framework outlined above, we first need to establish a reliable typological sample. The largest available sample was collected by Cysouw (2003), and consists of 265 individual person marking paradigms from the world’s languages, including both pronominal and agreement paradigms. At first glance, this would be the obvious sample to use, however Cysouw’s dataset is problematic for us in a number of respects. First, Cysouw codes for a smaller domain than the one in Figure 1A, one which consists of only 8 possible referents obtained by combining four core person categories (s, s+a, a, o) with two number categories (singular and augmented). More importantly, Cysouw counts individual paradigms, not

languages, which means that two paradigms from the same language can be counted as two different person systems. Moreover, the sample was compiled with the goal of including as many examples of rare paradigms as possible, but not necessarily every case of commonly occurring patterns (see Cysouw, 2003, pp.19-20), which means that the dataset is skewed in favor of rare paradigms. For these reasons, we use Cysouw’s dataset as a secondary source, and compile our own dataset to use as our primary source.

The dataset we use consists of ten different person systems, derived from the pronominal systems of Māori (Austronesian), Tuaripi (Eleman), Mandarin (Sino-Tibetan), Taiwanese (Sino-Tibetan), Slovenian (Indoeuropean-Slavic), Spanish (Indoeuropean-Romance), English (Indoeuropean-Germanic), Kalam (Trans-New Guinea), Sierra Popoluca (Zoquean), and Slave (Athabaskan). Seven of these systems were drawn from (Cysouw, 2003), and the rest from a range of different sources². These languages were selected on the basis of being both fairly well-documented and belonging to a varied range of language families. When possible, we additionally carried out interviews with informants (native speakers) to corroborate the information provided by the grammars. An illustration of the resulting personal pronoun systems is given in Figure 1B. Following an analogous procedure, we also generated the two unattested systems ruled out by Zwicky’s generalization.

Theoretical framework and predictions

In what follows, we first review the theoretical framework of Zaslavsky et al. (2018) and show how it can be applied to the domain of personal pronouns. We then derive two compression models that embody different cognitive representations of the domain, and test their ability to account for our

²Detailed systems for each language and source list appear in the Appendix: https://osf.io/3vsa5/?view_only=542fc7f8a3bb4ec8aa9b4f2755cd6910.

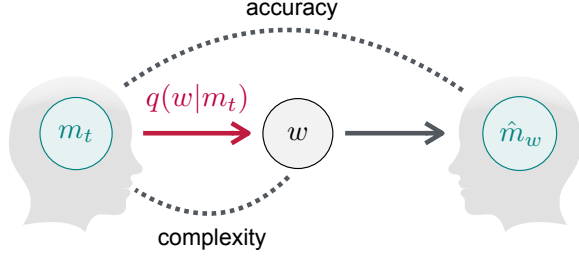


Figure 2: Communication model (adapted from Zaslavsky et al., 2018). See main text for details.

dataset of personal pronoun systems as well as Zwicky’s observation.

Communication model and theoretical bounds

The framework of Zaslavsky et al. (2018) is based on a basic communication model (Shannon, 1948), in which a speaker and a listener communicate about a shared domain \mathcal{U} (see Figure 2). In our case, \mathcal{U} corresponds to the person space shown in Figure 1A. Given a target $t \in \mathcal{U}$, sampled from a prior distribution $p(t)$, the speaker mentally represents t with a probability distribution $m_t(u)$ over the domain, which reflects uncertainty that the speaker may have, and then communicates it to the listener by producing a signal w using a probabilistic encoder $q(w|m_t)$. The listener interprets w by reconstructing from it a mental representation $\hat{m}_w(u)$, reflecting their inference about the speaker’s intended meaning. In our setting, the encoder $q(w|m_t)$ corresponds to a person system.

Given these assumptions, optimal communication systems are those in which the speaker and listener jointly satisfy the IB tradeoff between minimizing complexity and maximizing accuracy. In IB, as in RDT more generally, complexity roughly corresponds to the number of bits required for communication and it is measured by the mutual information between target meanings and signals,

$$I_q(M_t; W) = \sum_{t,w} p(t)q(w|m_t) \log \frac{q(w|m_t)}{q(w)}, \quad (1)$$

where $q(w) = \sum_t p(t)q(w|m_t)$. The accuracy of the system intuitively corresponds to the similarity between the speaker’s and listener’s mental representations. It is measured by $I_q(W; U)$, and maximizing this informational term amounts to minimizing the expected Kullback-Leibler divergence between m_t and \hat{m}_w ,³

$$\mathbb{E}_q[D[m_t||\hat{m}_w]] = \mathbb{E}_{p(t)} \left[\sum_u m_t(u) \log \frac{m_t(u)}{\hat{m}_w(u)} \right]. \quad (2)$$

Minimizing complexity alone yields a non-informative system that compresses the entire domain into a single form.

³Equation (2) is negatively related to accuracy, i.e., $I_q(W; U) = -\mathbb{E}_q[D[m_t||\hat{m}_w]] + \text{constant}$, and we take \hat{m}_w to be the optimal Bayesian estimator (see Zaslavsky et al., 2018, for derivation).

Maximizing accuracy alone yields a highly complex system that requires encoding the domain with a unique form for each referent. In between, there is a range of optimal systems that balance a tradeoff between these two competing goals by optimizing the IB objective function:

$$\mathcal{F}_\beta[q] = I_q(M_t; W) - \beta I_q(W; U), \quad (3)$$

where the parameter $\beta \geq 0$ controls the tradeoff. Every value of β provides an optimal encoder $q_\beta(w|m_t)$, and the IB theoretical limit is defined by the set of optimal systems as a function of β . Zaslavsky et al. (2018) hypothesized that semantic systems evolved under pressure to remain near this theoretical limit. If this hypothesis is true in our case, then we expect that attested systems will achieve near-optimal tradeoffs, and will tend to be more efficient compared to unattested systems.

To test this prediction for person systems, we need to specify two components that are required for instantiating the communication model: the prior distribution $p(t)$, also referred to as need distribution (Regier, Kemp, & Kay, 2015) as it reflects the frequency with which referents are communicated; and the structure of the speaker’s mental representations, m_t . Next, we discuss these two components.

Inferring patterns of communicative need

Previous work has often assumed a uniform need (e.g., Regier et al., 2015; Gibson et al., 2017) or a non-uniform but universal need distribution (e.g., Kemp & Regier, 2012; Zaslavsky et al., 2018). However, it seems unlikely that all the elements of the domain are equally needed, and it has been noted that communicative need may vary across languages (Kemp et al., 2018). To address these challenges, we infer language-specific need distributions from corpus frequencies using a recent domain-general method (Zaslavsky, Kemp, Tishby, & Regier, 2019b) based on the maximum entropy (MaxEnt) principle. This method infers a need distribution under minimal assumptions by finding the MaxEnt prior $p(t)$ that is consistent with the system and the normalized corpus frequencies $p(w)$ associated with its forms. Formally, this means that the prior must satisfy $\sum_t p(t)q(w|m_t) = p(w)$ for all w .

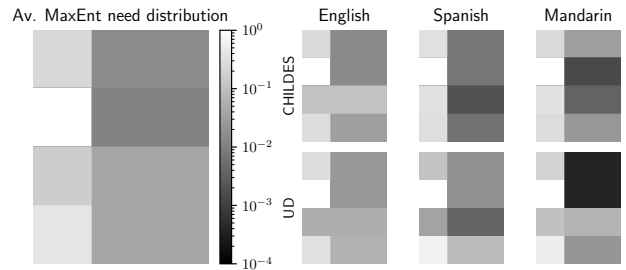


Figure 3: Inferred maximum-entropy (MaxEnt) communicative need distributions for the person domain. Left: Average MaxEnt need distribution. Right: Language-specific need distributions inferred from the CHILDES (top) and UD (bottom) corpora for each language.

We inferred corpus-based MaxEnt need distributions for English, Spanish, and Mandarin, the three languages in our dataset for which we have word frequency data both from the UD Treebank corpus (Nivre et al., 2017), that is based on written language, and from the CHILDES corpus (MacWhinney, 2000), that is based on spoken language (see Appendix). Figure 3 shows the inferred need distributions. For each language, the differences across corpora are relatively small. In addition, while this analysis shows interesting cross-language variation in the inferred needs, the main differences appear to be in very low probabilities, which are typically difficult to estimate, and the higher need probabilities actually reveal interesting patterns that are shared across languages. For example, in all three languages, the singular concepts tend to have greater need probabilities, with $p(\mathbf{s}) \geq p(\mathbf{a})$ and $p(\mathbf{o}) \geq p(\mathbf{a})$. This observation supports the simplifying universal need assumption, echoing a similar observation in the case of color (Zaslavsky et al., 2019b). It also suggests that using language-specific need distributions, which adds a huge overhead in model complexity, may not be necessary in order to account for cross-language variation in this domain. Therefore, we consider here only the average need distribution (Figure 3, left panel) and test the extent to which it may account not only for the three languages from which it was inferred, but for all systems in our dataset.

Domain representation

The second component we need to define is the structure of the speaker’s mental representations of the domain. We ground these representations in a 5-dimensional binary feature space, where the first three bits correspond to the conversational roles \mathbf{s} , \mathbf{a} , \mathbf{o} , and the two additional bits encode the three number distinctions. For example, \mathbf{s} corresponds to the feature vector $(1, 0, 0, 0, 0)$ and $\mathbf{a+o}$ to $(0, 1, 1, 0, 1)$. Following Regier et al. (2015), we then define the mental representations of each target referent $t \in \mathcal{U}$ by a similarity-based distribution, $m_t(u) \propto \exp(-\gamma d(u, t))$, where here we take $d(u, t)$ to be a weighted Hamming distance⁴ between the feature vectors of t and u . The free parameter γ reflects the degree of speaker uncertainty. Following Eisape, Levy, Tenenbaum, and Zaslavsky (2020), we set γ by assuming that the memory capacity of the speaker is fixed relative to the entropy of target referents (see Eisape et al., 2020, for more details). By manipulating the weights of each feature, this formulation allows us to derive precise quantitative predictions as to how cognitive biases in the mental representation of the domain, combined with pressure for efficient compression, can shape person systems across languages.

Returning to Zwicky’s observation

One simple way of setting the feature weights in our model is by treating all features equally, where the two bits for the

number distinction are considered as a single feature. However, it seems unlikely that this model instantiation, which we refer to as the **unbiased model**, will account for Zwicky’s observation. Most accounts of this observation posit an inherent asymmetry between the representation of the speaker and addressee (Harbour, 2016; Harley & Ritter, 2002). Based on experimental results suggesting that unattested inclusive systems may nevertheless be learnable, Maldonado and Culbertson (2020) propose that Zwicky’s observation may arise from a soft ‘egocentric’ bias favoring a salient speaker representation. A bias of this sort, perhaps stemming from a tendency to perceive the world as a function of our presence in it, has also been observed in early pronoun acquisition (Charney, 1980; Moyer, Harrigan, Hacquard, & Lidz, 2015).

Here, we aim to test this proposal by formulating an egocentric bias within our theoretical framework. Specifically, we introduce such a bias to our model by increasing the relative feature weight of \mathbf{s} in the mental representation of the domain. We also consider a secondary asymmetry proposed in the literature between participants (speaker and addressee) and non-participants in the conversation (e.g., Harbour, 2016; Harley & Ritter, 2002), which we formulate by decreasing the relative feature weight corresponding to \mathbf{o} . We refer to this type of model instantiation as an **egocentric model**. Grounding empirically the exact values of the feature weights requires substantial experimental work, which we leave for future research. Here, we set the feature weight vector to $(16, 1, 0.1, 1)$, corresponding to weights for \mathbf{s} , \mathbf{a} , \mathbf{o} , and number, while verifying that our results are robust within a range of reasonable weights that satisfy these asymmetries.⁵

If an egocentric bias shapes person systems, in addition to pressure for efficiency, then we expect that the egocentric model would provide a better account of attested systems compared to the unbiased model. Further, the biased model should be able to distinguish between attested and unattested systems, such as the second- and third-inclusive systems. Each model predicts that person systems across languages should lie near its theoretical bound, and therefore, following Zaslavsky et al. (2018), we take as our primary measure for evaluation the deviation from optimality attained by actual person systems (i.e., the minimal distance from the optimal value of (3), across all values of β).

Results

For each model instantiation (unbiased and egocentric), we evaluated the IB theoretical limit for personal pronoun systems, defined by the set of optimal systems derived from that model for different tradeoffs β between complexity and accuracy. Additionally, we evaluated the complexity and accuracy tradeoffs of the attested and unattested systems of Figure 1, by treating them as encoders and plugging them into equations (1) and (2). The results for the egocentric model

⁴The Hamming distance is a natural metric for binary vectors. To evaluate the influence of the choice of metric, we also considered a weighted Euclidean distance and obtained similar results.

⁵Specifically, we got qualitatively similar results when varying the \mathbf{s} -weight $\in \{4, 8, 16, 20\}$, the \mathbf{a} -weight $\in \{0.75, 1, 1.25, 1.5\}$, and the \mathbf{o} -weight $\in \{0.05, 0.1, 0.5, 0.75\}$.

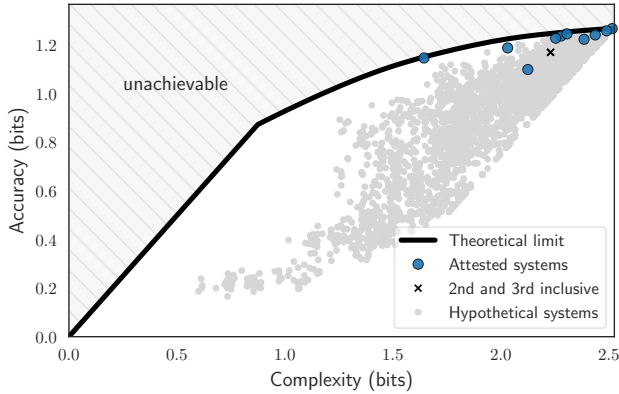


Figure 4: The theoretical limit of efficiency for person systems given the egocentric model (black curve). Blue dots correspond to the attested systems shown in Figure 1B. Most attested systems achieve near-optimal compression, in contrast to most of their hypothetical variants (gray points) and to the two unattested systems (black crosses).

are shown in Figure 4. It can be seen that almost all of the attested systems lie very close the theoretical limit derived from the egocentric model, suggesting that they achieve near-optimal compression. The two unattested systems lie further away from the curve, indicating that they are less efficient, at least when an egocentric bias is at play.

These results are shown quantitatively in Figure 5 (left panel), along with comparison to the unbiased model. All attested systems but one (Slave, which we will discuss later), achieve near-zero deviation from optimality with respect to the egocentric model, and deviate more from the unbiased model. This suggests that attested systems are more communicatively efficient in the presence of an egocentric bias. By contrast, the unattested second-inclusive and third-inclusive systems deviate more from optimality and are thus less efficient under the egocentric model. Under the unbiased model, however, the two unattested systems deviate less from optimality and are actually predicted to be as efficient as attested systems. Thus, as expected, the egocentric model captures Zwicky’s observation better than the unbiased model.

To test whether the attested systems are more likely to be near-optimal than expected by chance, we compared them to a large set of hypothetical systems. For each attested system, we constructed 300 hypothetical variants by randomly permuting the labels they assign to referents. Figure 4 shows that these 3,000 hypothetical systems are widely spread below the curve defining the theoretical limit of the egocentric model. Figure 5 (right panel) shows the distribution of the efficiency scores of these hypothetical systems for the two models. In both models, all attested systems are more efficient than most of their hypothetical variants, and lie at the lower tail of the distribution (which is bounded below by zero). This suggests that it is unlikely that the attested systems arrived at the theoretical limit by chance, without pressure for efficiency. We

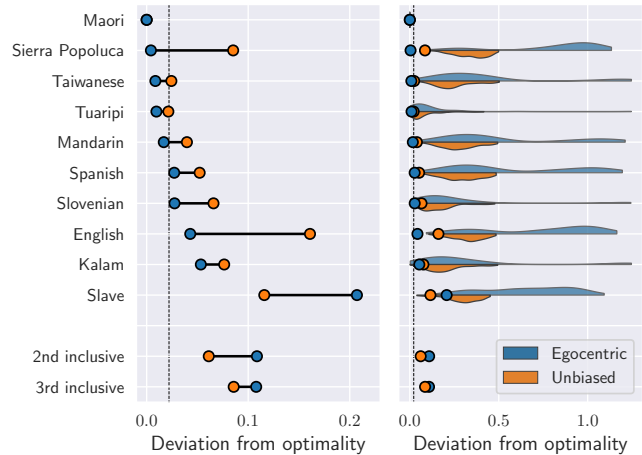


Figure 5: Left: Deviation from optimality of attested and unattested systems with respect to the egocentric model (blue dots) and unbiased model (orange dots). Lower values correspond to higher efficiency (zero is optimal). The dashed black line shows the median efficiency of attested systems given the egocentric model. Right: Density estimation for the efficiency of the hypothetical variants of each language (colored regions), given each model. Colored dots and the dashed black line are the same as in the right panel.

also note that the egocentric model predicts a larger efficiency gap between attested and hypothetical systems, lending further support to the egocentric model.

The fact that the Slave-type system is less efficient given the egocentric model is not particularly surprising because this system uses the same pronominal form for groups containing and excluding the speaker. However, this system is relatively rare among the world’s languages, appearing in roughly six languages, all within a single family (Athabaskan). Further, it has been shown to be dispreferred by learners in recent work using artificial language learning (Maldonado & Culbertson, 2020). Having said that, it is worth noting that given both models, this system is more efficient than most of its hypothetical variants, implying that it is still non-trivially efficient. One possibility is that this system is shaped by pressure for efficient compression, but the egocentric bias is less active for some reason.

Because the sample we use above is quite small, we test the extent to which our results generalize to a larger set of attested systems, namely the large typological dataset of Cysouw (2003). As noted earlier, one limitation of this dataset is that it documents a reduced domain with only 8 referents (collapsing the middle and right columns of Figure 1A, except s+a). We therefore adapted our models accordingly and repeated the same analysis as before. As with our smaller sample, we again find that the vast majority of systems in this larger dataset are near-optimal given the egocentric model (Figure 6A), and that the egocentric model outperforms the unbiased model (Figure 6B) while also predicting a larger ef-

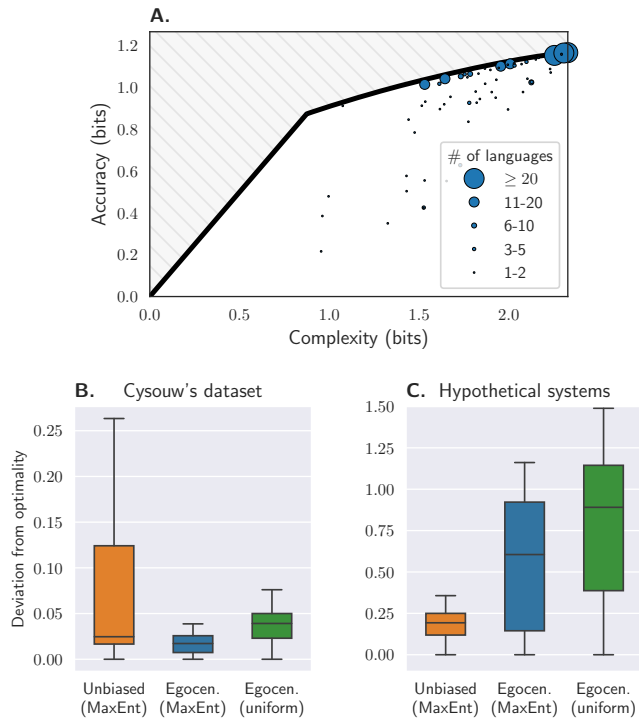


Figure 6: Analysis of the Cysouw (2003) dataset. **A.** The black curve is the theoretical limit of efficiency given the egocentric model, adapted for the smaller domain documented by Cysouw. All frequent (≥ 6) systems are highly efficient, while less efficient systems tend to be rare. **B.** Efficiency of the attested systems from the Cysouw dataset. The egocentric model (blue) achieves the best overall performance compared to the unbiased model (orange), and to an egocentric model with a uniform need distribution (green). **C.** Efficiency of hypothetical variants (note that different scale compared to B).

efficiency gap between attested and hypothetical systems (Figure 6C). However, as in the case of Slave, we observe in Figure 6A a cluster of very rare systems that are inefficient under the egocentric model. Nonetheless, given that this dataset is heavily skewed in favor of rare paradigms, it is remarkable that so many systems are in fact aligned with the egocentric model. Finally, we also tested the influence of our inferred MaxEnt need distribution by considering a variant of the egocentric model that is based on the uniform need distribution. The egocentric model with uniform need performed worse than our original egocentric model (Figure 6B), further supporting the proposal that substantial patterns of communicative need may be shared across unrelated languages.

Discussion

We have shown that personal pronoun systems across languages achieve near-optimal compression, replicating findings in qualitatively different semantic domains (Zaslavsky et al., 2018, 2019a), and providing converging evidence for the idea that pressure for efficient compression may be a fun-

damental principle shaping the lexicon. Additionally, our approach provides a formal and principled way of testing the ability of the egocentric bias proposed by Maldonado and Culbertson (2020) to account for Zwicky’s observation (Zwicky, 1977). Our findings support this proposal, and suggest that personal pronoun systems may be shaped by soft, rather than hard constraints. Under this view, pressure to maintain efficient compression schemes drives languages toward the IB theoretical limit of efficiency, which is in turn influenced by underlying cognitive mechanisms that shape our mental representation of the world.

Our work suggests a number of avenues for further research. Most obviously, our models should be tested on a more balanced dataset of personal pronoun systems, documenting the full meaning space. Beyond that, our findings suggest the possibility that languages (or cultures) may differ in the degree to which they are subject to an egocentric bias, or indeed operate under other biases in the representation of this domain. Our approach can be used to address this by formalizing alternative biases within the same information-theoretic framework. The language-specific analysis of communicative need reported above is also noteworthy, as it suggests that substantial patterns of communicative need may be shared across languages. However, this does not rule out the possibility that the cross-linguistic variation we observed in communicative need may be meaningful. An important direction for future work is to further probe these differences.

Acknowledgments

N.Z. was supported by a BCS Fellowship in Computation. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 757643 to J.C.).

References

- Bobaljik, J. D. (2008). Missing persons: A case study in morphological universals. *The Linguistic Review*, 25(1-2), 203-230. doi: 10.1515/TLIR.2008.005
- Charney, R. (1980). Speech roles and the development of personal pronouns. *J. Child Lang*, 7(3), 509–528.
- Cysouw, M. (2003). *The Paradigmatic Structure of Person Marking*. Oxford, UK: OUP Oxford.
- Denic, M., Steinert-Threlkeld, S., & Szymanik, J. (2020). Complexity/informativeness trade-off in the domain of indefinite pronouns. In *Proceedings of the 30th semantics and linguistic theory conference*.
- Eisape, T., Levy, R., Tenenbaum, J. B., & Zaslavsky, N. (2020). Toward human-like object naming in artificial neural systems. In *Bridging Artificial Intelligence and Cognitive Science workshop, ICLR*.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingham, S., ... Conway, B. R. (2017). Color naming across languages reflects color use. *PNAS*, 114(40), 10785-10790.

- Harbour, D. (2016). *Impossible persons*. Cambridge, MA: MIT Press.
- Harley, H., & Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, 78(3), 482–526. doi: 10.1353/lan.2002.0158
- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084), 1049–1054.
- Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, 4(1).
- MacWhinney, B. (2000). *The childes project: The database* (Vol. 2). Psychology Press.
- Maldonado, M., & Culbertson, J. (2020). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, 1–71.
- Moyer, M., Harrigan, K., Hacquard, V., & Lidz, J. (2015). 2-year-olds' comprehension of personal pronouns. In E. Grillo, K. Jepson, & M. LaMendola (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (p. 11). Boston, MA.
- Nivre, J., Agic, Z., Ahrenberg, L., Antonsen, L., Aranzabe, M. J., Asahara, M., ... Zhu, H. (2017). *Universal Dependencies 2.1*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Retrieved from <http://hdl.handle.net/11234/1-2515>
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: Wiley-Blackwell.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Sonnaert, J. (2018). *The Atoms of Person in Pronominal Paradigms*. Netherlands: LOT, Netherlands Graduate School.
- Steinert-Threlkeld, S. (2020). Quantifiers in natural language optimize the simplicity/informativeness trade-off. In *Proceedings of the 22nd amsterdam colloquium*.
- Tishby, N., Pereira, F. C., & Bialek, W. (1999). The Information Bottleneck method. In *37th annual Allerton conference on communication, control and computing*.
- 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. (2020). Information-theoretic principles in the evolution of semantic systems (Doctoral dissertation, The Hebrew University of Jerusalem).
- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *PNAS*, 115(31), 7937–7942.
- Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2019b). Communicative need in colour naming. *Cognitive Neuropsychology*. doi: 10.1080/02643294.2019.1604502
- Zaslavsky, N., Regier, T., Tishby, N., & Kemp, C. (2019a). Semantic categories of artifacts and animals reflect efficient coding. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 1254–1260). Austin, TX: Cognitive Science Society.
- Zwicky, A. M. (1977). Hierarchies of person. In *Proceedings from the Chicago Linguistic Society* (Vol. 13, pp. 714–733). Chicago, MI: Chicago Linguistic Society.