Efficient coding explains cross-linguistic patterns in Person systems

Mora Maldonado,*[1] Noga Zaslavsky,*[2] and Jennifer Culbertson[3]
[1]Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra; [2]Department of Brain and Cognitive Sciences and Center for Brains Minds and Machines, MIT; [3]Centre for Language Evolution, University of Edinburgh; * Equal contribution.

Person systems, exemplified by pronominal paradigms (e.g., 'I', 'you', 'they'), refer to entities as a function of their conversational role: the speaker, the addressee, and others who play no active role in the conversation. As in other semantic domains (e.g., color and kinship) it has been observed that person systems across languages exhibit variation, and yet not all systems are equally likely (Cysouw, 2003). A number of theories have be formulated to explain person typology—typically by positing featural representations and constraints specific to the person domain (e.g., on feature combination Harley and Ritter, 2002; Harbour, 2016). Here, we test an alternative hypothesis based on a recent information-theoretic approach to semantic systems (Zaslavsky et al., 2018). In this view, languages evolve under functional pressure to efficiently encode meanings into words by optimizing an information-theoretic tradeoff between the complexity and communicative accuracy of the lexicon. This approach is grounded in Rate–Distortion theory (RDT), the branch of information theory that characterizes efficient data compression, and has been shown to explain cross-linguistic patterns in several semantic domains, including colors and containers. It is also closely related to other notions of efficiency (Kemp et al., 2018), which are not grounded in RDT but have been applied to domains qualitatively more similar to person, such as indefinites (Denic et al., 2020).

We assume the space in Fig. 1a, with three conversational roles (a unique speaker (s), a unique addressee (a), an undefined number of others (o), following Harbour, 2016, a.m.o.), and three number distinctions (exactly one, exactly two, and more than two). Languages partition this space in different ways, and some well-documented partitions are shown in Fig. 1b. In addition, we include two unattested partitions, in which the inclusive meaning is homophonous with the second or third person, rather than the first (as in English 'we'). The typological observation that those two systems are unattested, known as Zwicky's Generalization (Zwicky, 1977), has been taken to suggest an asymmetry between speaker and addressee roles, often encoded in terms of hard constraints on possible person systems (Harbour, 2016).

The theoretical framework we use (Zaslavsky et al., 2018) is based on a simple communication model in which a speaker and a listener communicate about a shared domain (here, referents in the person space). An optimal communication system is one in which the interlocutors jointly optimize a tradeoff between the complexity of the system, measured by the mutual information between words and referents, and the accuracy of the system, measured by the similarity between the speaker's and listener's mental representations. This tradeoff, known as the Information Bottleneck principle (Tishby et al., 1999), can be derived as a special case of RDT, and depends on two main components: communicative need and domain representation. We estimate a prior distribution over referents reflecting communicative need (following the maximum-entropy method from Zaslavsky et al., 2019) using word frequency data from the CHILDES and UD Treebank corpora. The speaker's mental representations are grounded in a 5-dimensional feature space, with binary features that correspond to the three conversational roles (s, a, o) and to two number distinctions ('exactly two' and 'more than two'). We define representations in a way that allows us to test the proposal that Zwicky's Generalization stems from an 'egocentric' bias, i.e., a bias for keeping the speaker distinct. Formally, this bias is defined by increasing the weight of s relative to the other feature weights. If this bias shapes person systems, in addition to pressure for efficiency, then an 'egocentric' model including

this bias should provide a better account of attested systems compared to an unbiased model, and it should be able to distinguish between attested and unattested systems in Fig. 1b.

We evaluated the complexity–accuracy tradeoffs for the systems in Fig. 1 as well as for 1,500 hypothetical systems generated by random permutation of the meaning-to-form mapping of the attested systems. Fig. 2a shows the results for the egocentric model. All attested systems lie very close to the theoretical limit, indicating that they are near-optimally efficient. In contrast, most of the hypothetical systems lie further away from the curve, suggesting that it is unlikely that the attested systems arrived at the theoretical limit by chance. Fig. 2b compares the egocentric and unbiased models: under the unbiased model, the unattested second and third inclusive systems are as efficient as attested languages (e.g., English). However, the egocentric model predicts a substantial efficiency gap between these system and attested languages. These findings suggest that person typology, including Zwicky's Generalization, may be explained by pressure for efficiency in the presence of an egocentric bias, supporting the idea that soft biases rather than strong constraints on possible systems shape this semantic domain (Maldonado and Culbertson, 2020). More generally, this provides converging evidence for the idea that fundamental information-theoretic principles shape the lexicon.

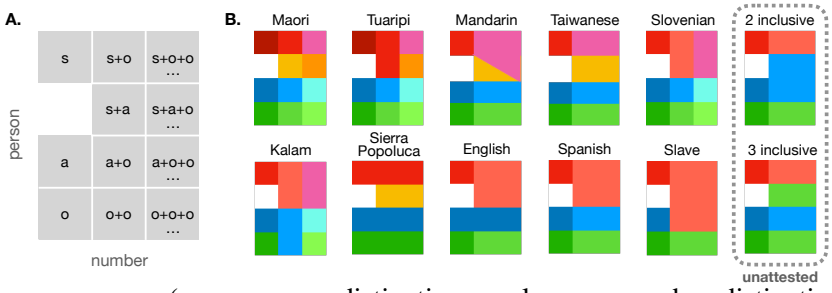

Figure 1: **A.** The person space (rows=person distinctions; columns =number distinctions). **B.** 10 attested Person systems, and two that are believed to be unattested. Colors correspond to distinct pronominal forms.

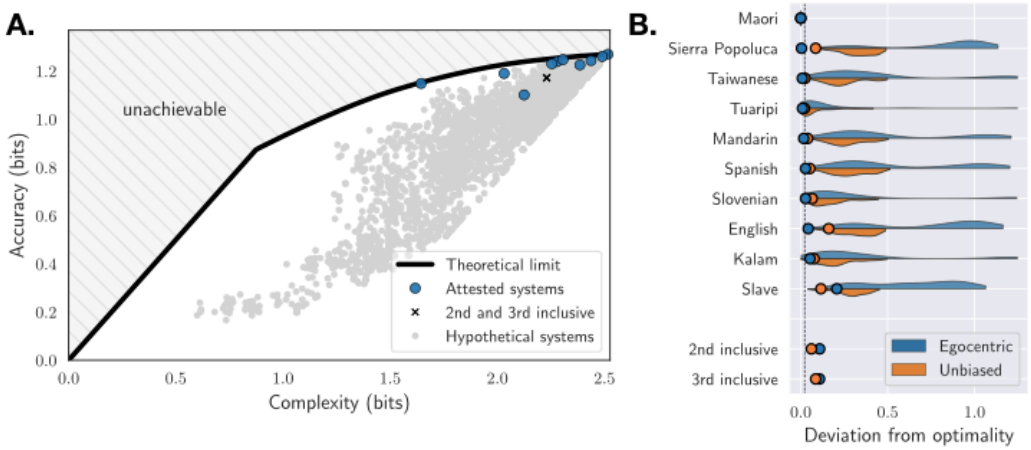

Figure 2: **A.** Theoretical limit of efficiency given the egocentric model (black curve). Blue dots correspond to the attested systems in Fig. 1B. Gray points are hypothetical variants. Black crosses are the two unattested systems. **B.** Colored dots show model efficiency predictions for attested systems. Color regions show density estimation for the efficiency score of hypothetical variants given the egocentric (blue) or unbiased (orange) model. Dashed black line shows median efficiency of attested systems given egocentric model.

References. Cysouw, M. (2003). *The Paradigmatic Structure of Person Marking*. OUP Oxford. || Denic, M., Steinert-Threlkeld, S., and Szymanik, J. (2020). Complexity/informativeness trade-off in the domain ofindefinite pronouns. In *Proceedings of the 30th SALT*. || Harbour, D. (2016). *Impossible Persons*. Linguistic Inquiry Monographs. MIT Press. || Harley, H. and Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, 78(3). || Kemp, C., Xu, Y., and Regier, T. (2018). Semantic Typology and Efficient Communication. *Annual Review of Linguistics*,4(1). || Maldonado, M. and Culbertson, J. (2020). Person of interest: Experimental investigations into the learnability

of person systems. *Linguistic Inquiry*. || Tishby, N., Pereira, F. C., and Bialek, W. (1999). The Information Bottleneck. In 37th Annual Allerton Conference. || Zaslavsky, N., Kemp, C., Regier, T., and Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31).|| Zaslavsky, N., Kemp, C., Tishby, N., and Regier, T. (2019). Communicative need in color naming. *Cognitive Neuropsychology*. || Zwicky, A. M. (1977). Hierarchies of person. In *Proceedings from the Chicago Linguistic Society*, 13.