Bridging semantics and pragmatics in information-theoretic emergent communication

Eleonora Gualdoni
Universitat Pompeu Fabra
eleonora.gualdoni@upf.edu

Mycal Tucker
MIT
mycal@mit.edu

Roger P. Levy
MIT
rplevy@mit.edu

Noga Zaslavsky
UC Irvine
nogaz@uci.edu

Languages evolve through repeated interactions in rich contexts, where various communicative and non-communicative goals co-exist. The conveyed meaning is often shaped by the local conversational context of utterances (Figure 1), as captured by the *pragmatic* behavior of interlocutors, and at the same time, words are associated with non-contextualized meanings, as captured by *lexical semantics*. While semantics and pragmatics are widely studied, their interface and co-evolution is largely under-explored and not well understood. In this work we begin to address this major gap in our understanding by asking: How can a shared lexicon emerge from local pragmatic interactions?

To this end, we build on a framework for information-theoretic emergent communication in artificial agents (Tucker et al., 2022). This framework is particularly relevant to our question because it integrates utility maximization, which is a central component in well-established models of pragmatics (Goodman and Frank, 2016; Benz and Stevens, 2018), with general communicative constraints that are believed to shape human semantic systems (Zaslavsky et al., 2018) as well as pragmatic reasoning (Zaslavsky et al., 2020). We adjust this framework to explicitly model the interface between semantics and pragmatics, such that agents learn to communicate in a pragmatic setting, i.e., in the presence of a shared conversational context, and then we evaluate their emergent lexicon. We test our model in a rich visual domain of naturalistic images, and find that human-like properties of the lexicon can emerge when agents are guided by both context-specific utility and general communicative pressures.

Modeling the co-evolution of semantics and pragmatics. Our model builds on the VQ-VIB architecture (Tucker et al., 2022), which includes a speaker and a listener (Figure 2). The speaker is defined by (i) a representation module (VAE) that maps a referent x to a 'mental' representation m, and (ii) an encoder module S that generates a communication signal w given the speaker's mental state. The listener is defined by (i) a decoder D that observes w and generates a reconstruction $\hat{m}$, and (ii) a policy L for solving a downstream task.

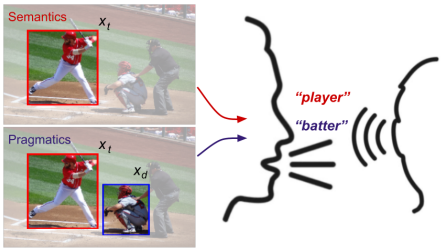

Figure 1: An example of an image from the ManyNames dataset annotated with bounding boxes, illustrating how languages support both semantic categorization and pragmatic behavior.

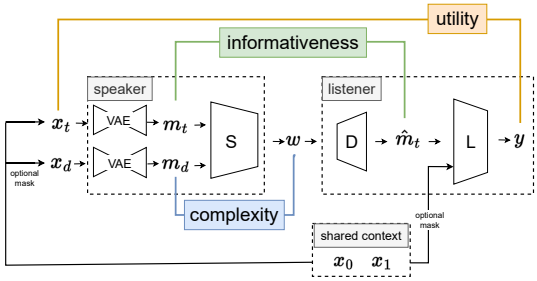

Figure 2: Emergent communication model for the co-evolution of semantics and pragmatics (see main text). Pragmatics setting: both agents observe inputs x_0, x_1 as shared context; one input is randomly selected as target x_t for the speaker. Semantics setting: there is no shared context; the speaker observes x_t while x_d is masked.

In our **pragmatics setting**, which we used for training, both agents observe a shared context (x_0, x_1), while the speaker also observes which referent is the target t and which is a distractor d. The speaker then aims to communicate the target x_t. The listener's task is to guess the target based on $y = L(\hat{m}, (x_0, x_1))$. Agents are

Proceedings of the Society for Computation in Linguistics (SCiL) 2024, pages 333-335.
Irvine, California, June 27-29, 2024

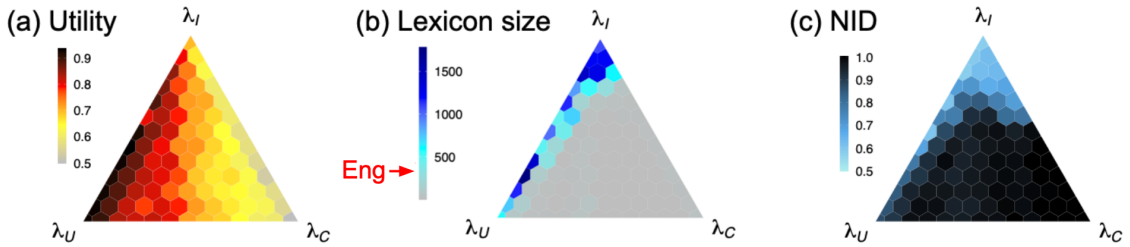

Figure 3: Evaluation of the emergent communication systems in the ManyNames domain. (a) Average utility, reflecting the agents' pragmatic competence; (b) lexicon size; and (c) Normalized Information Distance (NID; Vinh et al., 2010) between the emergent lexicon and the English naming data, measuring their (mis)-alignment.

trained by optimizing a tradeoff between expected utility, informativeness, and complexity. Utility $U(x_t, y)$ is defined by the (task-specific) accuracy of the listener's predictions. Informativeness and complexity are (task-agnostic) communicative objectives, derived from the Information Bottleneck (IB) framework for semantic systems (Zaslavsky et al., 2018). Informativeness is related to the negative distortion between m_t and $\hat{m}$, which can be approximated by their MSE, and complexity corresponds to $I(m; w)$, which is roughly the number of bits for communication. In practice, we optimize a bound on the mutual information, denoted by $\tilde{I}$ (see Tucker et al., 2022, for details). Overall, the training objective is to maximize $\lambda_U \mathbb{E}\left[U\left(x_t, y\right)\right] - \lambda_I \mathbb{E}\left[\|m_t - \hat{m}_t\|^2\right] - \lambda_C \tilde{I}(w; m_t, f(I))$, where the λs are non-negative tradeoff weights that sum to 1.

In our **semantics setting**, which is used to evaluate the emergent lexicon after training, only the target is shown to the speaker (the distractor is masked) and then the listener reconstructs $\hat{m}_t$ based on the speaker's word w, without any additional context or downstream task.

Results. We test our model on the ManyNames dataset (Silberer et al., 2020), which provides a rich visual domain of 25K naturalistic images (Figure 1), as well as free naming data from English native speakers who were asked to describe with a single word a target object highlighted with a bounding box. We trained 457 agent pairs with different values of λ_U, λ_C and λ_I, by considering the ManyNames targets, plus one distractor per image.

As shown in Figure 3, for each set of λs, we recorded three measures for evaluation, corresponding to (a) the agents' pragmatic competence, (b) the emergent lexicon size, and (c) the alignment between the emergent lexicon and English. As expected, none of the extremes are human-like: $\lambda_U = 1$ yields high pragmatic competence but also high NID, suggesting that the emergent lexicon does not reflect a human-like semantic categorization of the domain. $\lambda_I = 1$ yields lower NID but at a cost of a huge lexicon with over 1500 unique ws, whereas English speakers used less than 400 words. Finally, $\lambda_C = 1$ yields non-informative communication. In between, however, there is a range of λs > 0, where λ_U and λ_I tend to be larger than λ_C, in which the emergent communication systems have similar lexicon sizes as English, relatively low NID, and high pragmatic competence. This suggests that human-like properties of the lexicon may emerge from local pragmatic interactions when agents are guided by all three terms.

Conclusions. We propose a novel approach to studying the interface and co-evolution of semantics and pragmatics, using multi-agent simulations in unsupervised settings, guided by a tradeoff between utility, informativeness, and complexity. Our results suggest that all three terms are crucial for understanding language evolution. An important direction for future research is to further evaluate the structure of the emergent lexicon and explore conditions in which the alignment between our agents and human languages could be further improved.

Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 715154) and Ministerio de Ciencia e Innovación and the Agencia Estatal de Investigación (Spain; ref. PID2020-112602GB-I00/MICIN/AEI/10.13039/501100011033). This paper reflects the authors' view only, and the funding agencies are not responsible for any use that may be made of the information it contains.

References

Anton Benz and Jon Stevens. 2018. Game-theoretic approaches to pragmatics. *Annual Review of Linguistics*, 4.

Noah D. Goodman and Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.

Carina Silberer, Sina Zarrieß, and Gemma Boleda. 2020. Object naming in language and vision: A survey and a new dataset. In *Proceedings of LREC*, pages 5792–5801, Marseille, France. European Language Resources Association.

Mycal Tucker, Roger P. Levy, Julie Shah, and Noga Zaslavsky. 2022. Trading off utility, informativeness, and complexity in emergent communication. In *Advances in Neural Information Processing Systems*.

Nguyen Vinh, Julien Epps, and James Bailey. 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *JMLR*, 11:2837–2854.

Noga Zaslavsky, Jennifer Hu, and Roger Levy. 2020. A Rate–Distortion view of human pragmatic reasoning.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *PNAS*, 115(31):7937–7942.