Pragmatics in referential communication: An investigation of concept communication and the role of pragmatics with an emergent communication game

Kristina Kobrock[1], Xenia Ohmer[1], Elia Bruni[1] and Nicole Gotzner[1]

[1]Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

1 Introduction

We investigate the role of pragmatics in the communication of concepts at different levels of abstraction with an emergent communication paradigm using a reference game. Reference games, where a speaker describes a target and a listener has to identify the correct target among a set of distractors, are ideal to study referential expressions at different levels of abstraction because they allow for systematic manipulation of the context (e.g., Degen et al., 2020; Franke & Degen, 2016; Hawkins et al., 2018). More recently, this game setup has been adapted to computational studies of emergent communication between deep neural-network agents (e.g., Lazaridou et al., 2017; Mu & Goodman, 2021; Ohmer et al., 2022). Such computational methods allow for rigorous manipulations, and for simulating language on various time scales from evolution to situational use. They are therefore increasingly used to answer questions in the field of pragmatics (e.g., Fang et al., 2022; Hu et al., 2022; White et al., 2020). Adding to this work, we systematically study the influence of concept and context type on the choice of referring expressions during emergent communication.

2 Methods

A speaker and a listener agent, modeled with neural networks, develop a communication system while playing a concept-level reference game (see Figure 1). Other than in a classical reference game, the speaker has to communicate not a single but multiple targets belonging to the same concept (Mu & Goodman, 2021). The agents are trained via Reinforcement Learning and are rewarded when the listener picks the correct target objects after having decoded a message generated by the speaker. We train the agents on a novel symbolic dataset that disentangles *concept type*, ranging from specific to generic, from *context type*, ranging from fine to coarse. The most specific concept is defined by target objects where all attributes have a fixed value (e.g., *'blue circle'*). Objects that define the most generic concept have only one fixed attribute (e.g., *'circle'*). Distractors in a fine context share more attributes with the target concept, whereas distractors in a coarser context condition share fewer attributes with the target concept. This setup allows us to systematically study the role of the concept's level of abstraction and of the context granularity in the referential communication of concepts.

We implement three different games to investigate the agents' communicative strategies, and specifically whether they develop and use pragmatic behavior in the sense of context-based pragmatics (Sedivy, 2003) or recursive pragmatic reasoning about communicative intentions (Goodman & Frank, 2016; Grice, 1989). The **basic setup** involves *context-aware* speaker and listener agents. They learn to communicate about concepts ranging from specific to generic in all context conditions. By careful analysis of the emerging communication protocol, we will gain insight into the communicative strategies used by the speakers driven by the game setup. One key question is whether the agents learn to reason about the context when deciding on the referring expression's appropriate level of abstraction. Another key question is whether agents are able to use abstraction to generalize to previously unseen concepts. The second game introduces **external pressures**. A *cost on the message length* has been employed in previous studies to incentivize a least-effort pressure towards brevity, which might help the agents to focus on communicating only what is relevant (e.g., Chaabouni et al., 2019). Again, the key question is whether agents trained with an additional pressure develop the pragmatic behavior of reasoning about the context. The third game implements speakers with a **recursive pragmatic reasoning module** modeled with the *Rational Speech Act (RSA) framework* (Goodman & Frank, 2016) as in Fang et al. (2022). They are trained as context-unaware literal agents, but are able to make use of context information and recursive pragmatic reasoning at test time. We compare the performance and generalization abilities of speakers in all three games to the context-unaware literal agents as a baseline. This helps to identify which communicative strategies are beneficial in communication about concepts by disentangling the role of context information, utterance cost and recursive pragmatic reasoning.

3 Hypotheses and approach for analyzing the emergent communication

We expect the speakers to use different (pragmatic) strategies depending on the game, as well as the concept and context type. Possible strategies of the speakers are: 1) Only communicating relevant information *(implicit abstraction)*, 2) over- or underspecifying, 3) communicating all information and then indicating what of it is relevant *(explicit abstraction)*.

We formulate the following hypotheses mapping to the three games described above:

- **Baseline: Context-unaware literal agents (L)** have to communicate all relevant attributes to be successful, thus may be overinformative (non-pragmatic baseline).
- **H1: Context-aware literal agents (L-aware)** can communicate fewer than all attributes and let uncertainty be resolved by context (context-based pragmatics).
- **H2: L-aware + utterance cost** will further reduce overinformation because communicating fewer attributes becomes beneficial (context-based pragmatics + implicit abstraction).
- **H3: L + RSA** will increase the agents' performance through additional recursive reasoning of the speaker (reasoning about intentions).

In order to see whether our hypotheses hold, we need a thorough analysis of how the agents communicate in all settings. First, we evaluate loss and accuracy on train and validation datasets to measure the agents' performance. Next, we test the agents' generalization abilities on previously unseen data. Whether literal and pragmatic agents are able to generalize to more generic or to more specific concepts shows how well the agents match human abilities, and whether this requires pragmatic reasoning. We also employ a number of metrics to assess informativity of the generated messages (e.g., *normalized mutual information*), compositionality of the emerging communication system (e.g., *topographic similarity*) and how the concepts are represented in the agents' hidden representations.

4 Results

Here, we present results for the first two planned simulations, the baseline context-unaware literal agents and the context-aware literal agents. The agents have access to the target objects (*concept*) in both settings, but only the context-aware trained agents also have access to the distractor objects (*context*). The datasets were implemented as described above and named by the number of attributes (think 'shape', 'color', etc.) and values (think 'square', 'circle' etc.) an object in this dataset can take (e.g., 'D(3,4)' means that objects have three attributes that take four values each). We ran simulations for six datasets that span a range of three to five attributes and four to 16 values. We report means and standard deviations from five simulations per dataset and 300 training epochs.

First, we observe very high training and validation accuracies for all game settings and datasets (mean validation accuracies > 96 for all datasets and both settings). This suggests that the agents learn to successfully communicate about concepts on various levels of abstraction and in various context conditions.

Second, we find compositional structure in the messages as indicated by relatively high topographic similarity (topsim) scores for the baseline setting (topsim scores ranging from 0.287 (*sd=0.04*) for dataset D(3,16) to 0.539 (*sd=0.02*) for D(4,4); 1 indicates perfect compositionality). In the context-aware setting, topsim scores are slightly lower (ranging from 0.185 (*sd=0.01*) for D(3,16) to 0.463 (*sd=0.05*) for D(5,4)). This might be due to the fact that topsim scores are calculated without taking the context conditions into account. Context-aware agents might tailor their utterances more to the context which results in lower overall topsim scores.

Third, we look at entropy scores for both settings. We report normalized mutual information (MI) which indicates how much information about the messages can be obtained by observing the concepts and vice versa, i.e. whether there is a strong one-to-one correspondence between messages and concepts. When agents were trained context-unaware, their messages have high overall entropy scores (MI ranging from 0.911 (*sd=0.019*) for D(3,8) to 0.972 (*sd=0.013*) for D(3,8)) without visible differences between the context conditions. This suggests that concepts and messages tend to have a one-to-one mapping which does not change depending on context. This might reflect the expected speaker's strategy of communicating all relevant attributes and thus being overinformative in coarse context conditions when being trained context-unaware. When agents are trained context-aware, on the other hand, we observe that the entropy scores differ more between context conditions (see Table 1). Specifically, we observe a pattern where the coarser the context, the lower the MI and the finer the context, the higher the MI. This reflects the intuition from above that agents might develop less one-to-one mappings between messages and concepts in the coarse context conditions. The reason for this might be that in coarse contexts, both more and less specific messages can be successful (e.g., "circle" can mean 'red circle', 'blue circle' etc.) because when less specific messages are used, the target concept can still be disambiguated by the context. In fine contexts, on the other hand, the messages need to contain more information on more specific levels of abstraction to be sufficiently discriminative in the context which intuitively results in more one-to-one mappings (e.g., a more specific utterance like "red circle" is only used for the more specific concept 'red circle'). These results are in line with previous work on how an emerging vocabulary depends on the contexts in which the targets are presented. Hawkins et al. (2018) have found the same pattern in an artificial language learning paradigm with human participants: The finer the context, the more one-to-one mappings are established in an emerging language, and the coarser the context, the more synonyms can be found.

In conclusion, the here presented models and analyses contribute to our understanding of referential communication and the role of pragmatics in communicating concepts through a systematic manipulation of communicative pressures. Our first results show that the speaker's access to the context shapes the emerging communication system reproducing a pattern that was observed in humans. Although the differences we observe between the context-aware and context-unaware settings are rather small, they do indicate that the mere presence of context already drives its use in communication. We expect even more clear results from future simulations where we plan to incentivize the use of the context more and equip the agents with recursive pragmatic reasoning abilities.

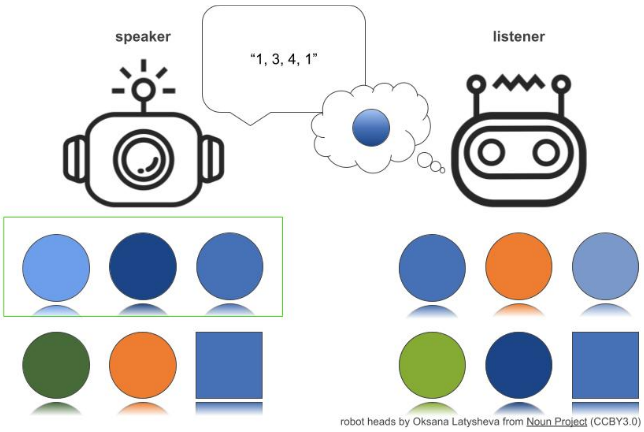

Figure 1: Schematic illustration of the concept-level reference game with the *specific* target concept "blue circle" (fixing both shape and color attributes) and a *fine* context condition (distractors share one attribute, either shape or color, with the target concept). Note that the specific objects that satisfy the concepts can differ between agents.

Datasets	0 shared attributes	1 shared attribute	2 shared attributes	3 shared attributes	4 shared attributes
D(3,4)	0.903	0.932	0.969	-	-
D(3,8)	0.916	0.921	0.928	-	-
D(3,16)	0.822	0.820	0.817	-	-
D(4,4)	0.921	0.932	0.945	0.965	-
D(4,8)	0.881	0.890	0.903	0.923	-
D(5,4)	0.861	0.883	0.907	0.927	0.949

Table 1: Normalized mutual information over context conditions ranging from coarse (0 shared attributes) to fine (2-4 shared attributes depending on the dataset) when speakers are trained context-aware.

References

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019). Anti-efficient encoding in emergent communication. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/31ca0ca71184bbdb3de7b20a51e88e90-Abstract.html

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*, *127*(4), 591–621. https://doi.org/10.1037/rev0000186

Fang, F., Sinha, K., Goodman, N., Potts, C., & Kreiss, E. (2022). Color Overmodification Emerges from Data-Driven Learning and Pragmatic Reasoning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*. https://escholarship.org/uc/item/9kn7n6qb

Franke, M., & Degen, J. (2016). Reasoning in Reference Games: Individual- vs. Population-Level Probabilistic Modeling. *PLOS ONE*, *11*(5), e0154854. https://doi.org/10.1371/journal.pone.0154854

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Hawkins, R. X. D., Franke, M., Smith, K., & Goodman, N. D. (2018). Emerging abstractions: Lexical conventions are shaped by communicative context. *Proceedings of the 40th annual conference of the cognitive science society (CogSci)*, 463–468. http://cocolab.stanford.edu/papers/HawkinsEtAl2018-Cogsci.pdf

Hu, J., Levy, R., & Zaslavsky, N. (2022). Scalable pragmatic communication via self-supervision. *ICML Workshop on Self-Supervised Learning for Reasoning and Perception*. https://doi.org/10.48550/arXiv.2108.05799

Lazaridou, A., Peysakhovich, A., & Baroni, M. (2017). Multi-agent cooperation and the emergence of (natural) language. *International Conference on Learning Representations*. https://openreview.net/forum?id=Hk8N3Sclg

Mu, J., & Goodman, N. (2021). Emergent Communication of Generalizations. *Advances in Neural Information Processing Systems*, *34*, 17994–18007. https://proceedings.neurips.cc/paper_files/paper/2021/file/9597353e41e6957b5e7aa79214fcb256-Paper.pdf

Ohmer, X., Duda, M., & Bruni, E. (2022). Emergence of Hierarchical Reference Systems in Multi-agent Communication. *Proceedings of the 29th International Conference on Computational Linguistics*, 5689–5706. https://aclanthology.org/2022.coling-1.501

Sedivy, J. C. (2003). Pragmatic Versus Form-Based Accounts of Referential Contrast: Evidence for Effects of Informativity Expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23. https://doi.org/10.1023/A:1021928914454

White, J., Mu, J., & Goodman, N. D. (2020). Learning to refer informatively by amortizing pragmatic reasoning. https://doi.org/10.48550/arxiv.2006.00418