Paying Attention to Function Words

Shane Steinert-Threlkeld

Institute for Logic, Language and Computation, Universiteit van Amsterdam

S.N.M.Steinert-Threlkeld@uva.nl --- https://www.shane.st

Introduction

Two major types of linguistic expressions [4]:

- Content words: nouns, verbs, adjectives
- Function words: determiners, tense, conjunctions, prepositions, complementizers, ...

> Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe;
> All mimsy were the borogoves,
> And the mome raths outgrabe.
> — From 'Jabberwocky', Carroll [1]

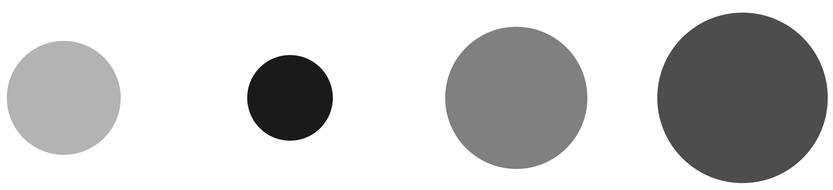
Crucial questions for explaining the emergence of *compositional* communication:

- Why have human languages evolved to exhibit this division of labor between content and function words?
- How could such a distinction have emerged in the first place?

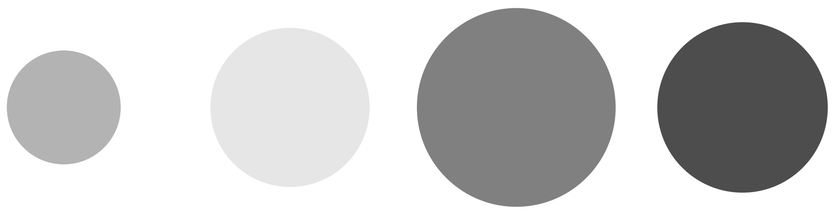
Contributions

1. Why existing approaches don't explain this distinction [longer version]
2. A new signaling game [3, 5], with variable contexts and gradable properties
3. The emergence of function words by reinforcement learning and *attention*

A Signaling Game with Varying Contexts



Refer to the circle on the left as "the lightest one".



Refer to the circle on the left as "the smallest one".

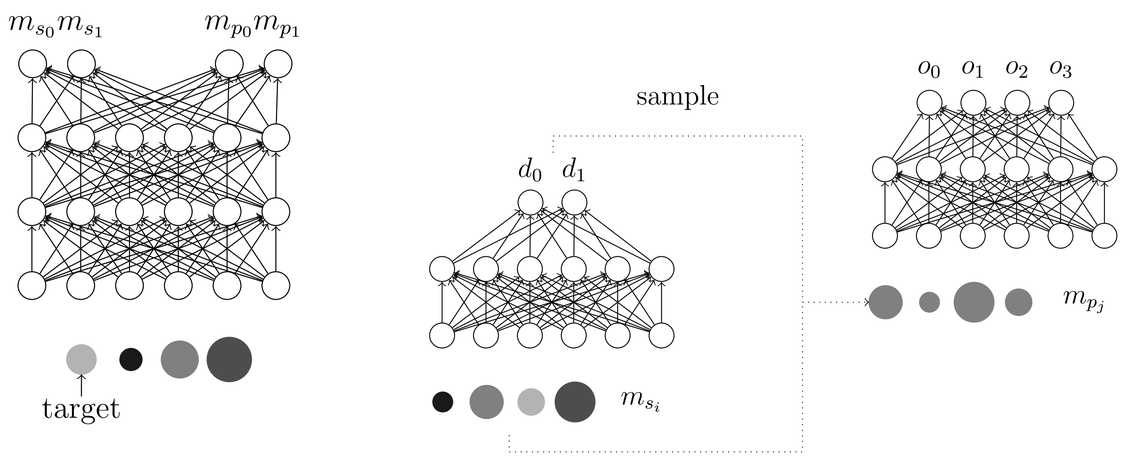
(1) A *context* c over scales S is a set of objects such that: for each $o \in c$, there is a scale $s \in S$ such that either o has the least degree on s ($o = \arg\min_{o' \in c} s(o')$) or the highest degree on s ($o = \arg\max_{o' \in c} s(o')$).

(2) Extremity Game, in general:
   a. Nature chooses a context c and a target object $o \in c$.
   b. The sender sees c and o and sends a message m from some set of messages M.
   c. The receiver sees c and m and chooses an object o' from c.
   d. The play is successful (and the two agents equally rewarded) if and only if $o' = o$.

(3) Toy semantics for a gradable adjective and superlative morphemes.
   a. $[\![\text{size}]\!] = \lambda x.s_{\text{size}}(x)$
   b. $[\![\text{-est}]\!]^c = \lambda P_{\langle e,d\rangle}.\lambda x_e.x \in c \text{ and } \forall x' \in c, P(x) \succeq P(x')$
   c. $[\![\text{least}]\!]^c = \lambda P_{\langle e,d\rangle}.\lambda x_e.x \in c \text{ and } \forall x' \in c, P(x) \preceq P(x')$

Experiment

Similar to [2], we train agents to play this game using REINFORCE [6], varying (a) number of properties and (b) receiver architecture type.
Code + data: `https://github.com/shanest/function-words-context`



(a) Basic Sender (b) Attentional Receiver

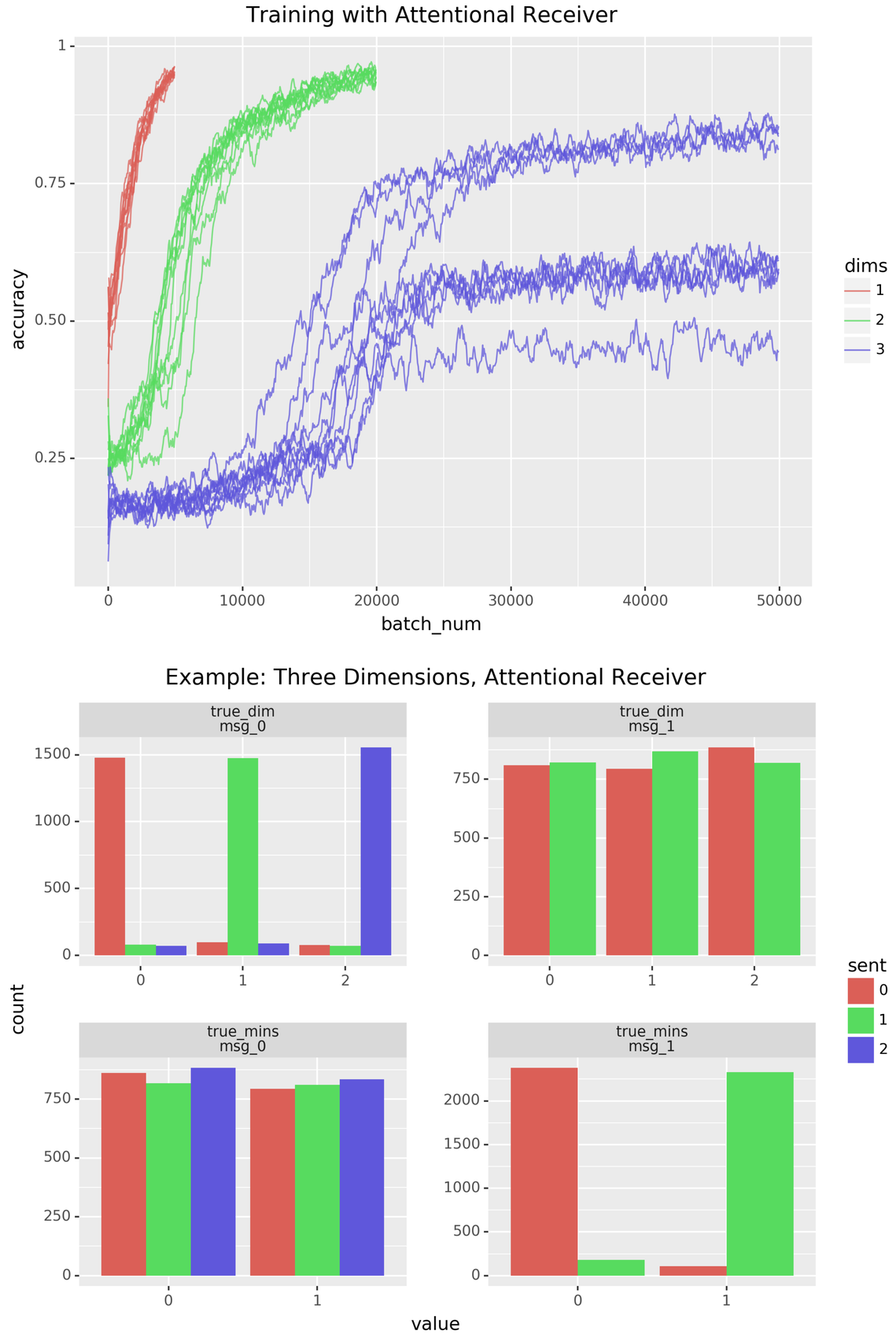
Results

dims	mean	std
1	0.975	0.006
2	0.985	0.003
3	0.731	0.062

(a) Basic Receivers

dims	mean	std
1	0.959	0.005
2	0.964	0.005
3	0.697	0.144

(b) Attentional Receivers



Future Research

- Fewer assumptions about what aspects of the input to pay attention to
- RNNs as sender/receiver, with costs for:
  - Vocabulary size
  - Length of messages

References

[1] Lewis Carroll. *Through the Looking-Glass, and What Alice Found There*. Macmillan, 1871.

[2] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. "Multi-Agent Cooperation and the Emergence of (Natural) Language". In: *International Conference of Learning Representations (ICLR2017)*. 2017. arXiv: 1612.07182. URL: http://arxiv.org/abs/1612.07182.

[3] David Lewis. *Convention*. Blackwell, 1969.

[4] Luigi Rizzi and Guglielmo Cinque. "Functional Categories and Syntactic Theory". In: *Annual Review of Linguistics* 2.1 (2016), pp. 139–163. DOI: 10.1146/annurev-linguistics-011415-040827.

[5] Brian Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, 2010.

[6] Ronald J Williams. "Simple statistical gradient-following algorithms for connectionist reinforcement learning". In: *Machine Learning* 8.3-4 (1992), pp. 229–256.

Acknowledgements

Thanks to Jeff Barrett, Emmanuel Chemla, Meica Magnani, Iris van de Pol, and Jakub Szymanik for helpful comments and discussion. An extended version of this paper was presented at the Workshop on Evolutionary Explanations of Compositional Communication at the Biennial Conference of the Philosophy of Science Association: see http://philsci-archive.pitt.edu/15274/. This work was supported by funding from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. STG 716230 CoSaQ.