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, ...

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 smallest one”.

Refer to the circle on the left as “the lightest one”.

(1) A context c over scales S is a set of objects such that: for each o ∈ c, there is a scale s ∈ S such that either o has the least degree on s (o = arg min∈c P(s(o))) or the highest degree on s (o = arg max∈c P(s(o))).

(2) Extremity Game, in general:
   a. Nature chooses a context c and a target object o ∈ 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. [size] = λx. size(x)
   b. [-est] = λx,y. λx′,x ∈ c and ∀x′ ∈ c, P(x) ≥ P(x′)
   c. [-least] = λx,y. λx′,x ∈ c and ∀x′ ∈ c, P(x) ≤ 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.

Sequence:

\begin{align*}
\text{target} & \quad m_1 \quad m_2 \quad m_{\text{seq}} \\
\text{sample} & \quad o_1 \quad o_2 \quad o_3 \\
\text{Basic Sender} & \quad \text{Attentional Receiver} \\
\text{(a)} & \quad \text{(b)} \\
\end{align*}

Results

<table>
<thead>
<tr>
<th>dims</th>
<th>mean</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.975</td>
<td>0.006</td>
</tr>
<tr>
<td>2</td>
<td>0.985</td>
<td>0.003</td>
</tr>
<tr>
<td>3</td>
<td>0.731</td>
<td>0.062</td>
</tr>
</tbody>
</table>

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


Acknowledgements

Thanks to Jeff Barrett, Emmanuel Chemla, Marco Baroni, Irony Johnson, Jakub Szymanik for helpful comments and discussion. An extended version of this paper was presented at the Workshop on Evolutionary Explorations of Computational 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 Union Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 681328 (Colas)