The typology of lexical classes in emergent languages

Edoardo Ponti\textsuperscript{1}, Dieuwke Hupkes\textsuperscript{2}, Diana Rodriguez\textsuperscript{2} and Elia Bruni\textsuperscript{3}

\textsuperscript{1}University of Cambridge, \textsuperscript{2}University of Amsterdam, \textsuperscript{3}Universitat Pompeu Fabra

Interaction and the Evolution of Linguistic Complexity
University of Edinburgh, June 18, 2019
Introduction

L. Steels, 2015
Referential games

Data: Shapes, Andreas et al (2016)
What do they say?

bo bo di la la
What do they say?

bo bo di la la  di la bo ke la
What do they say?

bo bo di la la  di la bo ke la

What can we do with these languages?
Research questions

- Do lexical classes emerge in the agents’ languages?
Research questions

▸ Do *lexical classes* emerge in the agents’ languages?
▸ Does this depend on the hyperparameters $L$ and $|V|$?
Setup

- Do *lexical classes* emerge in the agents’ languages?
- Does this depend on the hyperparameters $L$ and $|V|$?

- 9 different setups:
  - Initial vocabulary sizes $|V|$: 7, 14 or 28
  - Maximum lengths $L$: 3, 5 or 10
### Some statistics

<table>
<thead>
<tr>
<th>Settings</th>
<th>Language Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average L</td>
</tr>
<tr>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>28</td>
</tr>
</tbody>
</table>
What do they say?

$|V| = 7$, $L=5$   bo bo di la la   di la bo ke la
What do they say?

| $|V| = 7, L=3$ | la di di | la mu di |
| $|V| = 7, L=5$ | bo bo di la la | di la bo ke la |
| $|V| = 28, L=5$ | ti fa do ke ti | ti ti ke do la |
What is their language like?

- Topographic similarity (Lazaridou et al, 2018)
- Causal influence (Lowe et al., 2019)
- Representational similarity (Bouchacourt et al, 2018)
- Message distinctness (Choi et al., 2018)
- Perplexity per word (Havrylov and Titov, 2017)
What is their language like?

Two types of information:

- What images do the messages refer to (Semantic)
- What do the messages look like (‘Syntactic’)

Semantic analysis

Local Mutual Information

\[ \text{LMI}(symb; feat) = p(symb, feat) \cdot \log \frac{p(symb|feat)}{p(symb)} \]

(Evert, 2005)
## Semantic Analysis

Purity of words and features

|     | $|V|$ |     | $|V|$ |
|-----|-----|-----|-----|
|     | 7   | 14  | 28  | 7   | 14  | 28  |
| 3   | 0.37| 0.35| 0.24| 0.28| 0.22| 0.15|
| 5   | 0.29| 0.31| 0.31| 0.25| 0.21| 0.17|
| 10  | 0.29| 0.36| 0.28| 0.22| 0.21| 0.12|

Word purity

Feature purity
## Semantic Analysis

Highest scoring features

| $|V|$ | L  | Feature  | Purity |
|---|----|----------|--------|
| 7  | 3  | triangle | 0.74   |
| 7  | 5  | right    | 0.82   |
| 7  | 10 | lower    | 0.52   |
| 14 | 3  | middle   | 0.44   |
| 14 | 5  | right    | 0.46   |
| 14 | 10 | right    | 0.60   |
| 28 | 3  | triangle | 0.39   |
| 28 | 5  | left     | 0.38   |
| 28 | 10 | lower    | 0.29   |

**Table:** Highest scoring feature per setup.
Syntax

Hidden Markov Model with Hierarchical Dirichlet Process

(Johnson and Willsky, 2013; Teh et al., 2005)
Syntactic Analysis

Hyperpriors

\[ \alpha \text{ hyper-prior.} \quad \gamma \text{ hyper-prior.} \]
Overlap between semantic and syntactic clusters

| $|V|$ | $L$ | B-cubed | NMI |
|---|---|---|---|
| 7 | 3 | 0.426 | 0.464 |
| 7 | 5 | 0.244 | 0.466 |
| 7 | 10 | 0.346 | 0.378 |
| 14 | 3 | 0.371 | 0.284 |
| 14 | 5 | 0.395 | 0.234 |
| 14 | 10 | 0.266 | 0.267 |
| 28 | 3 | 0.320 | 0.189 |
| 28 | 5 | 0.224 | 0.076 |
| 28 | 10 | 0.167 | 0.096 |
Some intermediate conclusions

- Large variation for both syntactic and semantic analysis, depending on the initial vocabulary size and maximum message length
- Agents talk primarily about position, and not about shapes and colors
Some intermediate conclusions

- Large variation for both syntactic and semantic analysis, depending on the initial vocabulary size and maximum message length
- Agents talk primarily about position, and not about shapes and colors

There is a framework that addresses the *functional* aspect of language, but we should also take care of the ecosystem that the agents live in.
Preview of a parallel project

Internal and External Pressures

Internal

▶ **Least effort**: Speaking has a cost

External

▶ **Subjective Constancy**: Objects can be recognised under different circumstances
  ▶ *Illumination*
  ▶ *Position*

▶ **Object Frequency**: objects and features are non-uniformly occurring in the real world
Internal pressure for least-effort

<table>
<thead>
<tr>
<th></th>
<th>Acc</th>
<th>Avg Length</th>
<th>Std Length</th>
<th>N tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>0.99</td>
<td>11.0</td>
<td>0.0</td>
<td>20.67</td>
</tr>
<tr>
<td>penalty</td>
<td>0.98</td>
<td>6.10</td>
<td>0.87</td>
<td>13.0</td>
</tr>
</tbody>
</table>
Diagnostic Classifier Accuracy

(rnn) → red

bo bo di la la

(Diagnostic Classifiers, Hupkes et al., 2018)
## Diagnostic Classifier Accuracy

<table>
<thead>
<tr>
<th>Game</th>
<th>Shape</th>
<th>Colour</th>
<th>Size</th>
<th>Hor</th>
<th>Vert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>0.33</td>
<td>0.33</td>
<td>0.50</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.53</td>
<td>0.45</td>
<td>0.60</td>
<td>0.93</td>
<td>0.96</td>
</tr>
</tbody>
</table>
## Diagnostic Classifier Accuracy

<table>
<thead>
<tr>
<th>Game</th>
<th>Shape</th>
<th>Colour</th>
<th>Size</th>
<th>Hor</th>
<th>Vert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>0.33</td>
<td>0.33</td>
<td>0.50</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.53</td>
<td>0.45</td>
<td>0.60</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Location invariance</td>
<td>0.65</td>
<td>0.99</td>
<td>0.91</td>
<td>0.33</td>
<td>0.34</td>
</tr>
</tbody>
</table>
## Diagnostic Classifier Accuracy

<table>
<thead>
<tr>
<th>Game</th>
<th>Shape</th>
<th>Colour</th>
<th>Size</th>
<th>Hor</th>
<th>Vert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>0.33</td>
<td>0.33</td>
<td>0.50</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.53</td>
<td>0.45</td>
<td>0.60</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Location invariance</td>
<td>0.65</td>
<td>0.99</td>
<td>0.91</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>Colour constancy</td>
<td>0.36</td>
<td>0.67</td>
<td>0.60</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>
## Diagnostic Classifier Accuracy

<table>
<thead>
<tr>
<th>Game</th>
<th>Shape</th>
<th>Colour</th>
<th>Size</th>
<th>Hor</th>
<th>Vert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>0.33</td>
<td>0.33</td>
<td>0.50</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.53</td>
<td>0.45</td>
<td>0.60</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Location invariance</td>
<td>0.65</td>
<td>0.99</td>
<td>0.91</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>Colour constancy</td>
<td>0.36</td>
<td>0.67</td>
<td>0.60</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>World distribution</td>
<td>0.68</td>
<td>0.73</td>
<td>0.88</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Recap

► Do lexical classes emerge in the agents’ languages?
  ► Semantic analysis (LMI and Purity)
  ► Syntactic analysis (HMM)
  ► Cluster overlapping

► Answer: a little bit

► Internal and External Pressures
  ► Least effort
  ► Subjective Constancy
  ► Object Frequency

► Diagnostic classification

► Conclusion: the ecosystem matters