The Reading Brain as a Statistical Learning Machine

Davide Crepaldi

International School for Advanced Studies (SISSA), Trieste

davide.crepaldi@sissa.it
https://lrlac.sissa.it
Tweets at @CrepaldiDavide

Heriot–Watt University, Edinburgh, 26 September 2018
Reading is a human wonder
We’re extremely good readers . . .

- We can read 8-letter words in ~35ms (e.g., Forster and Davis, 1984)
- We gather information about ~20 letters every ~200ms (e.g., Rayner, 1998)
- We read ~300 words per minute (e.g., Pelli et al., 2007)
... with no genetic endowment

- Written language isn’t observed universally
- Literacy isn’t acquired spontaneously
Word morphology
Arbitrariness

- elephant
- table
- heat
- drum
Arbitrariness. Really?

- elephant
- table
- heat
- drum
- preheat
- juicer
- bioweapon
- guesstimate
- untweet (?)
A break into arbitrariness

meaning

morphology

form
The core idea

Morphology creates probabilistic regularities in language form and in form–to–meaning mapping. The brain codes for these regularities and uses them during processing.
Positional constraints in morpheme perception
Morpheme positional constraints

- KINDNESS and NESSKIND
- PREHEAT and HEATPRE
- CATWALK and WILDCAT
- OVERHANG and HANGOVER
Blind to suffixes?

- (GASFUL vs. GASFIL) vs. (FULGAS vs. FILGAS)

(Crepaldi et al., 2010)
Blind to prefixes?

- (PREHOSE vs. PLEHOSE) vs. (HOSEPRE vs. HOSEPLE)
Stems everywhere?

- ####
- dealer
- 500 ms
- 30-55 ms
- DEAL
- 2000 ms
- press button
Stems everywhere?

▷ (fishgold–GOLDFISH vs. kacnvrqw–GOLDFISH) vs. (tonebari–BARITONE vs. suyzchmw–BARITONE)

(Crepaldi et al., 2013)
Orthography–to–Semantic Consistency (OSC)
Form as a cue to meaning

CORN
- Get all words that start with CORN
- Take their semantic representations
- Compute their similarity
- Take the mean

Orthography–Semantic Transparency (OSC)
- The internal consistency of the ‘form’ family in terms of meaning
- How similar in meaning are words similar in form
- How good of a cue to meaning is form
**Table 6.** Results of the regression analysis on the lexical decision latencies extracted from the BLP for a large set of random words

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.5922</td>
<td>.0109</td>
<td>602.89</td>
<td>.0001</td>
</tr>
<tr>
<td>Word frequency</td>
<td>-0.0308</td>
<td>.0009</td>
<td>33.41</td>
<td>.0001</td>
</tr>
<tr>
<td>Word FS</td>
<td>-0.0041</td>
<td>.0021</td>
<td>1.97</td>
<td>.0495</td>
</tr>
<tr>
<td>Word length</td>
<td>0.0035</td>
<td>.0013</td>
<td>2.74</td>
<td>.0061</td>
</tr>
<tr>
<td>OSC</td>
<td>-0.0254</td>
<td>.0066</td>
<td>3.84</td>
<td>.0002</td>
</tr>
</tbody>
</table>

(Marelli et al., 2014)
OSC also explains ERPs

(Amenta et al., in prep.)
The core idea

Morphology creates probabilistic regularities in language form, and in form–to–meaning mapping. The brain codes for these regularities and uses them during processing.
The core idea

*Language* shows probabilistic regularities in its form, and in form–to–meaning mapping. The brain codes for these regularities and uses them during processing.
Orthography in Baboons
Baboons learn words (Grainger et al., 2012)
Baboons extract knowledge about letter stats

A

All first words

Nonword responses (%)

0 20 40 60 80 100

First words Nonwords Difference

DAN ART CAU DOR VIO ARI

B

Last 50 first words

Nonword responses (%)

0 20 40 60 80 100

First words Nonwords Difference
Baboons extract knowledge about letter stats
The lesson from Baboons

- We don’t need language to do visual word identification
- Visual word learning proceeds through letter stats (perhaps)
Novel word learning in humans
A new lexicon

- 200 novel words (e.g., mefoal), 6 to 9 letters long
- Each novel word presented 3 times, interspersed with 600 non-words (e.g., paltoon)
- Lexical decision with feedback
Letter stats distinguish words and non-words in the novel language
Minimal bigram frequency rules
Not a useful cue
A stronger cue
Still, minimal bigram frequency is what matters
The lesson from humans

- We do code for letter stats
- We don’t seem to figure out the informative cue in a novel lexicon though, we just go for minimal bigram frequency
- Unfamiliar script?
Phantom words

(Endress and Mehler, 2009)
Pseudofonts

(Vidal et al., 2017)
Phantom words in reading

... S S S S S D S S S S D S S S S D S S S ...
Phantom words in reading
Longer words, same story
The lesson from phantom words

- We code for bigrams as we learn novel words.
- No strong commitment to bigrams, we didn’t try much else—the core point is that we capture sub-lexical stats in a word learning task.
The core idea

**Language** shows probabilistic regularities in its form, and in form-to-meaning mapping. The brain codes for these regularities and uses them during processing.
A new approach to reading

- Scripts can be seen as fully-fledged visual systems
- They can be studied as such, without language
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning
A new approach to reading

- Scripts can be seen as fully-fledged visual systems
- They can be studied as such, without language
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning
The Ministry of Education, University and Research in Italy; the ESRC, the British Academy and the Leverhulm Trust in the UK.

The ERC Starting Grant 679010 (StatLearn)

Jarek Lelonkiewicz and Yamil Vidal Dos Santos at SISSA; Simona Amenta in Gent; Marco Marelli at Milano Bicocca; Kathy Rastle at Royal Holloway; Colin Davis at Bristol.
The Reading Brain as a Statistical Learning Machine

Davide Crepaldi

International School for Advanced Studies (SISSA), Trieste

davide.crepaldi@sissa.it
https://lrlac.sissa.it
Tweets at @CrepaldiDavide

Heriot–Watt University, Edinburgh, 26 September 2018


