Natural scene statistics and the structure of visual signs over human history

Mark Changizi
Sloan-Swartz Center for Theoretical Neurobiology
Caltech

KITP, Sept, 2004

TWO PARTS, I and II

0. General motivation for visual linguistics.

I. Natural scene statistics and shapes of visual signs.

II. Combinatorial structure of letters.
PART 0

General motivation for visual linguistics
<table>
<thead>
<tr>
<th>Cognition</th>
<th>Laboratory</th>
<th>Natural production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>Cognitive psychology</td>
<td>Linguistics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognition</th>
<th>Laboratory</th>
<th>Natural production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>Visual psychophysics</td>
<td>Linguistics</td>
</tr>
</tbody>
</table>
## Natural Scene Statistics and the Structure of Visual Signs Over Human History

<table>
<thead>
<tr>
<th></th>
<th>Laboratory</th>
<th>Natural production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognition</strong></td>
<td>Cognitive psychology</td>
<td>Linguistics</td>
</tr>
<tr>
<td><strong>Vision</strong></td>
<td>Visual psychophysics</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Laboratory</th>
<th>Natural production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognition</strong></td>
<td>Cognitive psychology</td>
<td>Linguistics</td>
</tr>
<tr>
<td><strong>Vision</strong></td>
<td>Visual psychophysics</td>
<td><em>Visual linguistics</em></td>
</tr>
</tbody>
</table>

Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech (KITP Understanding the Brain Program 9/07/04)
Visual linguistics

Study of the relationship between the visual system and human-produced visual signs.

Two visual linguistics research directions thus far…
Two visual linguistics research directions thus far…

(1) Why visual signs are shaped the way they are.
• Are there any regularities in the shapes found in human-produced images?
• Why are letters and other visual symbols shaped the way they are?
• Ecological explanation/natural statistics.

(2) The complexity and redundancy of writing systems.
• How many strokes per letter?
• How combinatorially do strokes combine into letters?
• How redundant?
• Are there laws across the hundreds of writing systems over history?
• Does the visual system prefer these settings? If so, why?
PART I

Natural scene statistics and shapes of visual signs

...or...

Why visual signs are shaped the way they are.
My ecological hypothesis

The configurations of strokes found in writing systems are selected to match the configurations of contours found in natural scenes, because that is what the visual system is good at processing.
What I mean by “shape”: *configurations*

\[
\mathcal{L} = \{\land, \lor, \land \cdots\}
\]

\[
\mathcal{T} = \{\times, \vee, \land \cdots\}
\]

\[
\mathcal{X} = \{+ , \times, \land \cdots\}
\]

All the possible configuration types

\[
\begin{array}{cccccc}
\text{l} & \text{L} & \text{T} & \text{X} \\
\text{Y} & \text{Y} & \text{K} & \text{K} & \text{Y} & \text{K} \\
\Delta & \Delta & \Delta & \Delta & \Delta & \Delta \\
\text{21} & \text{22} & \text{23} & \text{24} & \text{25} & \text{26} \\
\text{27} & \text{28} & \text{29} & \text{30} & \text{31} & \text{32} \\
\text{33} & \text{34} & \text{35} & \text{36} & \text{37} & \text{38} \\
\text{39} & \text{40} & \text{41} & \text{42} & \text{43} & \text{44} \\
\end{array}
\]
Intuition pumping on sources of junctions in natural scenes, I

- Contiguous contours
- Partial occlusion
- Stacks

Intuition pumping on sources of junctions in natural scenes, II

- COMMON
- LESS COMMON
- EVEN LESS COMMON
- RARER
- RARER

Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech (KITP Understanding the Brain Program 9/07/04)
Intuition pumping on sources of junctions in natural scenes, III

Occlusion interpretation:

COMMON    COMMON    RARER    COMMON

ECOLOGICAL PREDICTION

(a) Natural scene measurements

(b) Semi-a-priori predictions
(a) Predictions from natural scene measurements

Sample of rural images
Configuration spectrum for rural images

Sample of urban images
Configuration spectrum for *urban* images

\[
y = 0.99x - 0.031 \\
R^2 = 0.846
\]

Configuration spectra for natural images

\[
y = 0.93x^{0.99} \\
R^2 = 0.846
\]
(b) Predictions from a semi-a-priori analysis
DATA

(1) Characters from 115 non-logographic writing systems.


(3) Thousands of non-linguistic symbols (Dreyfuss, *Symbol Sourcebook*, 1972)
The non-logographic writing systems…

<table>
<thead>
<tr>
<th>Name and example characters</th>
<th>Kind of system</th>
<th>Date</th>
<th>Phyllogeny</th>
<th>H</th>
<th>H - d</th>
<th>H + d</th>
<th>H - d</th>
<th>H + d</th>
<th>H - d</th>
<th>H + d</th>
<th>H - d</th>
<th>H + d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Japanese</td>
<td>logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Arabic</td>
<td>logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Mayan</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Maya</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Mayan</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Maya</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Maya</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Maya</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>Maya</td>
<td>non-logographic</td>
<td>1500</td>
<td>3000</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
</tbody>
</table>

Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech
(KITP Understanding the Brain Program 9/07/04)
Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech
(KITP Understanding the Brain Program 9/07/04)
Not a tautology

Scribbles by children

Shorthand (six systems)

Also, “random strokes” and tilings.

Visual versus motor, I

Trademarks versus shorthand

Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech
(KITP Understanding the Brain Program 9/07/04)
Visual versus motor, II

Minimum # hand-sweeps

# angles in stimulus (stimulus complexity)

“Visual” wins
Testing the ecological hypothesis, I

Testing the ecological hypothesis, II

Dr. Mark Changizi, Sloan-Swartz Center for Theoretical Neurobiology, Caltech
(KITP Understanding the Brain Program 9/07/04)
Conclusion to Part I:

- There are regularities in the kinds of structures/shapes found in visual signs over human history.
- They appear to not be driven primarily by motor optimization.
  (trademarks, hand-sweeps, cursive and shorthand)
- Estimates of the ecological frequency of configuration types accords well with the frequencies across visual signs, suggesting that visual sign structures have been selected to match the contour-combinations found in natural scenes, because that is what the human visual system is good at processing.

PART II

Combinatorial structure of letters
How do writing systems get larger?

Universal type approach

- Stroke types
- Characters

Invariant-length approach

- Stroke types
- Characters

Length L increases \( \sim \log C \)

- \# stroke types B invariant.

- Length L invariant.

- \( C \sim B^4 \), where \( d \leq L \), so

- B increases \( \sim C^{4d} \)
Writing systems appear, then, to follow the invariant-length approach.

Recall that this implies that the number of stroke types must increase…
# stroke types does indeed increase, as expected from the invariant-length.

But from the stroke-type plot, we can also infer how combinatorial characters are.
Length $L = 3$ strokes per character, so the maximum number of
degrees of freedom relating stroke types and characters is 3.
I.e., $C \sim B^3$.

But the actual relationship between them from the plot is…
$C = 0.268 \times B^{1.587}$. The combinatorial degree, $d = 1.587 \approx 1.5$.

Redundancy $R = 1 - d/L \approx 50\%$.

There is another way to estimate combinatorial degree and
redundancy…

If combinatorial, then stroke types in larger writing systems
must interact with a greater number of other stroke types.
Average degree per node does, indeed, increase. Namely approximately as $\delta \sim C^v$, where $v \approx 0.24$.

$$C \sim B^{\delta L-1} - BC^{v(L-1)}.$$  

$$C \sim B^{1/[1-v(L-1)]}.$$  

So, we can compute combinatorial degree $d = 1/[1-v(L-1)]$. For $v = 0.24$, $d = 1.5$.

Redundancy $R \equiv 1 - d/L = 50\%$, as before.
Summary of Part II:

- Writing systems increase in size via the invariant-length approach, not the universal-stroke-type approach.  
  \[ \text{Suggests some upper limit to visual processing. (?) \} ]

- The average number of strokes per character is \( \approx 3 \).  
  \[ \text{Why? Subitizing limit?} \]

- The number of stroke types increases, with combinatorial degree exponent of \( \approx 1.5 \). (Via two distinct kinds of estimate.) Therefore, only half of the possible degrees of freedom are utilized, and thus redundancy \( \approx 50\% \). \[ \text{Presumably useful for discriminability.} \]

Summary of ALL:

- Visual signs have been selected to be shaped like the ecology.

- Letters in writing systems have been selected to have approximately 50% redundancy.

- Visual linguistics—the study of human visual signs—may be a promising approach to discovering fundamental principles governing vision. (E.g., there is more likely to be mechanisms in the visual system for processing \( \top \) than \( \bot \).)

- Future directions include:
  - Human judgments of stimulus complexity.
  - Camouflage and animal visual signalling.