Why is computationalism relevant to language acquisition?

Cem Bozşahin

Middle East Technical University (ODTÜ)
bozsahin@metu.edu.tr

Psychology Seminars, Koç University
January 11, 2011
http://kybele.psych.cornell.edu/ edelman/

Rationalists do it by the rules.

Empiricists do it to the rules.
**Computationalism in psychology:** treating symbols as relating to the nature of representations, i.e. to their encoding in the mind (see Bickhard 1996 for overview).

**Computationalism in the rest of CogSci:** computational aspects that make a problem easy or difficult.

Computationalist psychology (Monaghan and Christiansen, 2004; Tenenbaum and Xu, 2000)

Non-computationalist computer models (e.g. ACT-R; Anderson 1976)
Cognitivism in CogSci: Qualitatively different problems

Computationalism in CogSci: Quantitatively different tasks (i.e., *same* problem, with some task-specific knowledge)

Empiricist in heart, interactionist at work
Some Piagetian stages

Period of Sensorimotor activity

   Stage of reflexes

   Stage of primary circular reactions

   Stage of coordination of secondary circular reactions

Period of Operational thought

Period of Formal operations
Computer scientist’s view of computation

Formal language theory (theory of descriptions)

Automata theory (theory of computing with descriptions)

Complexity theory (theory of algorithms and their computability)

  Space and time complexity

  effective computability and efficient computability
Running times for 1 microsecond unit operation; from Garey and Johnson (1979: Fig.1.2)

<table>
<thead>
<tr>
<th>Time complexity function</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>.00001</td>
<td>.00002</td>
<td>.00003</td>
<td>.00004</td>
<td>.00005</td>
<td>.00006</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
</tr>
<tr>
<td>$n^2$</td>
<td>.0001</td>
<td>.0004</td>
<td>.0009</td>
<td>.0016</td>
<td>.0025</td>
<td>.0036</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
</tr>
<tr>
<td>$n^3$</td>
<td>.001</td>
<td>.008</td>
<td>.027</td>
<td>.064</td>
<td>.125</td>
<td>.216</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
<td>second</td>
</tr>
<tr>
<td>$n^5$</td>
<td>.1</td>
<td>3.2</td>
<td>24.3</td>
<td>1.7</td>
<td>5.2</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>seconds</td>
<td>seconds</td>
<td>minutes</td>
<td>minutes</td>
<td>minutes</td>
</tr>
<tr>
<td>$2^n$</td>
<td>.001</td>
<td>1.0</td>
<td>17.9</td>
<td>12.7</td>
<td>35.7</td>
<td>366</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>second</td>
<td>minutes</td>
<td>days</td>
<td>years</td>
<td>centuries</td>
</tr>
<tr>
<td>$3^n$</td>
<td>.059</td>
<td>58</td>
<td>6.5</td>
<td>3855</td>
<td>$2 \times 10^8$</td>
<td>$1.3 \times 10^{13}$</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>minutes</td>
<td>years</td>
<td>centuries</td>
<td>centuries</td>
<td>centuries</td>
</tr>
</tbody>
</table>
Objects of computing

strings
- algebra
  - formal
  - language
  - theory
- calculus
  - parsing
  - and
  - automata
  - theory

functions
- algebra
  - domain
  - type theory
- calculus
  - lambda
  - calculus

algorithms
- complexity
  - theory
Some options for lang. acquisition

Nouns-first acquisition (common view; e.g. Gentner 1982)

No-one claims verbs-first (NB. Brown 1998; Tardif 1996)

Why not?

Computationally easier first
What makes a problem computationally easy or difficult?

Ambiguity

Non-determinism (not always the same thing as amb.)

Completeness and decidability

Memory kind, and its management

Frequency

Algorithms vs. heuristics
Radical lexicalization

First step towards getting a grip on computational properties.

(1) \[ S \rightarrow NP \ VP \quad Det \rightarrow \text{every} \]
\[ NP \rightarrow \text{Name} \quad N \rightarrow \text{chemist} \]
\[ NP \rightarrow \text{Det N} \quad \text{Name} \rightarrow \text{Kafka} \]
\[ VP \rightarrow V_{iv} \quad V_{iv} \rightarrow \text{arrived} \]
\[ VP \rightarrow V_{tv} \ NP \quad V_{tv} \rightarrow \text{adored} \]

\[ NP=S/VP \quad \text{and} \quad VP=S\backslash NP. \quad \text{Hence} \quad NP=S/(S\backslash NP) \]
S → NP VP  Det → every
NP → Name  N → chemist
NP → Det N  Name → Kafka
VP → $V_{iv}$  $V_{iv}$ → arrived
VP → $V_{tv}$ NP  $V_{tv}$ → adored

(2)  
$V_{tv}=VP/\text{NP}$  $V_{iv}=VP$  $NP=VP\backslash V_{tv}$
$NP=S/VP$  $VP=S/\text{NP}$  $Det=NP/N$
Name=NP  $N=NP\backslash \text{Det}$
Hence  
$V_{tv}=(S\backslash \text{NP})/\text{NP}$  $V_{iv}=S\backslash \text{NP}$
$NP=(S\backslash \text{NP})\backslash((S\backslash \text{NP})/\text{NP})$
$NP=S/(S\backslash \text{NP})$
\[
S \rightarrow \text{NP } \text{VP} \\
\text{NP} \rightarrow \text{Name} \\
\text{NP} \rightarrow \text{Det N} \\
\text{VP} \rightarrow V_{iv} \\
\text{VP} \rightarrow V_{tv} \text{ NP}
\]

\[
\text{Det} \rightarrow \text{every} \\
\text{N} \rightarrow \text{chemist} \\
\text{Name} \rightarrow \text{Kafka} \\
V_{iv} \rightarrow \text{arrived} \\
V_{tv} \rightarrow \text{adored}
\]

\[
(3) \quad \text{every} := \text{Det} = \text{NP}/\text{N} = (\text{S}/(\text{S}\backslash\text{NP}))/\text{N} \\
\text{chemist} := \text{N} = \text{NP}\backslash\text{Det} = \text{NP}\backslash(\text{NP}/\text{N}) \\
\text{Kafka} := \text{Name} = \text{NP} = \text{S}/\text{VP} = \text{S}/(\text{S}\backslash\text{NP}) \text{ and} \\
\quad (\text{S}\backslash\text{NP})\backslash((\text{S}\backslash\text{NP})/\text{NP}) \\
\text{arrived} := \text{VP} = \text{S}\backslash\text{NP} \\
\text{adored} := \text{VP}/\text{NP} = (\text{S}\backslash\text{NP})/\text{NP}
\]
Computational considerations

The child faces (PF,LF) pairs.

Syntax is the hidden variable (Zettlemoyer and Collins, 2005; Siskind, 1996)

Ambiguity

Frequency

Algorithmic complexity (power set construction)

short strings first; contiguity assumption

Needed for completeness
constraining the child’s hypothesis space

string of words $w_1 w_2 \cdots w_n$ (or syllables; Çöltekin and Bozsahin 2007)

Powerset construction for the hypotheses: $O(2^n)$

Contiguous subset construction: $O(n^2)$

more constraints: short and more frequent strings first
<table>
<thead>
<tr>
<th>Time complexity function</th>
<th>size $n$</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>second</td>
<td>.00001</td>
<td>.00002</td>
<td>.00003</td>
<td>.00004</td>
<td>.00005</td>
<td>.00006</td>
</tr>
<tr>
<td>$n^2$</td>
<td>second</td>
<td>.0001</td>
<td>.0004</td>
<td>.0009</td>
<td>.0016</td>
<td>.0025</td>
<td>.0036</td>
</tr>
<tr>
<td>$n^3$</td>
<td>second</td>
<td>.001</td>
<td>.008</td>
<td>.027</td>
<td>.064</td>
<td>.125</td>
<td>.216</td>
</tr>
<tr>
<td>$n^5$</td>
<td>second</td>
<td>.1</td>
<td>3.2</td>
<td>24.3</td>
<td>1.7</td>
<td>5.2</td>
<td>13.0</td>
</tr>
<tr>
<td>$2^n$</td>
<td>second</td>
<td>.001</td>
<td>1.0</td>
<td>17.9</td>
<td>12.7</td>
<td>35.7</td>
<td>366</td>
</tr>
<tr>
<td>$3^n$</td>
<td>second</td>
<td>.059</td>
<td>58</td>
<td>6.5</td>
<td>3855</td>
<td>$2 \times 10^8$</td>
<td>$1.3 \times 10^{13}$</td>
</tr>
</tbody>
</table>
A thought experiment

Inspired by Steedman and Hockenmaier (2007); Siskind (1996); Zettlemoyer and Collins (2005); Kwiatkowksi et al. (2010)

(4) Eat veggies

b. eat:=NP:eat'      veggies:=S\NP:λx.veg'x
c. eat:=NP:veg'      veggies:=S\NP:λx.eat'x
e. *eat:=S\NP:λx.eat'x      veggies:= NP: veg'
(6) No veggies.

\[ \frac{4}{20} \] percent of the possibilities, out of a total of 20, can relate veggies to \( veg' \) as a noun or verb.

The likelihood of \( no \) meaning e.g. veggies is \( \frac{2}{20} \).

(7) \hspace{1cm} \text{Experience 1} \hspace{1cm} \text{Experience 2 (with chocolate)}

<table>
<thead>
<tr>
<th>Eat veggies</th>
<th>No veggies</th>
</tr>
</thead>
<tbody>
<tr>
<td>eat := S/NP:eat'</td>
<td>no := S/NP:no'</td>
</tr>
<tr>
<td>veggies := S\NP:veg'</td>
<td>veggies := S\NP:no'</td>
</tr>
<tr>
<td>:veg'</td>
<td>:veg'</td>
</tr>
<tr>
<td>:eat'</td>
<td>:eat'</td>
</tr>
<tr>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>:veg'</td>
<td>:veg'</td>
</tr>
<tr>
<td>:choc'</td>
<td>:choc'</td>
</tr>
<tr>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>:no'</td>
<td>:no'</td>
</tr>
<tr>
<td>:veg'</td>
<td>:veg'</td>
</tr>
<tr>
<td>:choc'</td>
<td>:choc'</td>
</tr>
</tbody>
</table>
(8) Veggies gone.

Before this experience, half of the four veggies:=$veg'$ hypotheses consider this relation to be mediated by $S\backslash NP$, and the other half by $NP$.

(9) veggies := $S/\text{NP}$ :$veg'$   gone := $S\backslash\text{NP}$ :$veg'$

$\text{NP}$ :$\text{gone}'$ $\text{NP}$ :$\text{gone}'$

$\text{NP}$ :$\text{veg}'$ $\text{NP}$ :$\text{veg}'$

$\text{NP}$ :$\text{gone}'$ $\text{NP}$ :$\text{gone}'$

The child is $\frac{3}{6}$ likely to believe the connection is mediated by $\text{NP}$, $\frac{2}{6}$ by $S\backslash\text{NP}$, and $\frac{1}{6}$ by $S/\text{NP}$, in just three scenes.
Knowledge of words as subpieces of grammar and its use.

(10) veggies := {NP@3\(\frac{3}{6}\), S\(\frac{2}{6}\), S/NP@1\(\frac{1}{6}\)}: veg'

(11) veggies := {S\NP:veg'@2\(\frac{2}{14}\), S\NP:eat'@1\(\frac{1}{14}\),
               S\NP:no'@1\(\frac{1}{14}\), S\NP:choc'@1\(\frac{1}{14}\),
               S/NP:veg'@1\(\frac{1}{14}\), S/NP:gone'@1\(\frac{1}{14}\),
               NP:veg'@3\(\frac{3}{14}\), NP:eat'@1\(\frac{1}{14}\),
               NP:no'@1\(\frac{1}{14}\), NP:choc'@1\(\frac{1}{14}\), NP:gone'@1\(\frac{1}{14}\) }

### Keren’s first words (Dromi, 1987)

<table>
<thead>
<tr>
<th>Age</th>
<th>Child’s conven.</th>
<th>conven. form</th>
<th>Hebrew word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>10(12)</td>
<td>haw</td>
<td>(?)</td>
<td>a dog’s bark</td>
<td>12(16)</td>
</tr>
<tr>
<td>11(16)</td>
<td>?aba</td>
<td>(aba)</td>
<td>Father</td>
<td>12(18)</td>
</tr>
<tr>
<td>11(17)</td>
<td>?imaima</td>
<td>(?)</td>
<td></td>
<td>12(19)</td>
</tr>
<tr>
<td>11(18)</td>
<td>ham</td>
<td>(?)</td>
<td>said while eating</td>
<td>12(20)</td>
</tr>
<tr>
<td>12(3)</td>
<td>mu</td>
<td>(?)</td>
<td>a cow’s moo</td>
<td>12(23)</td>
</tr>
<tr>
<td>12(3)</td>
<td>?ia</td>
<td>(?)</td>
<td>a donkey’s bray</td>
<td>12(25)</td>
</tr>
<tr>
<td>12(8)</td>
<td>pil</td>
<td>(pil)</td>
<td>an elephant</td>
<td>12(25)</td>
</tr>
<tr>
<td>12(11)</td>
<td>buba</td>
<td>(buba)</td>
<td>a doll</td>
<td>12(25)</td>
</tr>
<tr>
<td>12(13)</td>
<td>pipi</td>
<td>(pipi)</td>
<td>urine</td>
<td>12(25)</td>
</tr>
</tbody>
</table>
### Tad’s first words (Gentner, 1982) AmE

<table>
<thead>
<tr>
<th>Age (m.)</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td></td>
<td>duck</td>
<td>daddy</td>
<td>dipe (diaper)</td>
<td>keys</td>
</tr>
<tr>
<td>eye</td>
<td></td>
<td>cow</td>
<td>bath</td>
<td>truck</td>
<td>cheese</td>
</tr>
<tr>
<td>down</td>
<td></td>
<td>boo</td>
<td>bottle</td>
<td>spoon</td>
<td>apple</td>
</tr>
<tr>
<td>eye</td>
<td></td>
<td>cow</td>
<td>bath</td>
<td>truck</td>
<td>cheese</td>
</tr>
<tr>
<td>down</td>
<td></td>
<td>boo</td>
<td>bottle</td>
<td>spoon</td>
<td>apple</td>
</tr>
</tbody>
</table>

Note: The table lists the words Tad learned at different ages, with examples including dog, eye, and down.
## Mik and Xan’s first verbs

*(Brown, 1998)*  
**VSO Tzeltal**

<table>
<thead>
<tr>
<th>Mik (1;5–2;0)</th>
<th>Xan (1;3–2;2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba</td>
<td>we’</td>
</tr>
<tr>
<td>go/allgone</td>
<td>eat tortillas</td>
</tr>
<tr>
<td>la’</td>
<td>chu’</td>
</tr>
<tr>
<td>come!</td>
<td>suckle breast</td>
</tr>
<tr>
<td>we’</td>
<td>ay</td>
</tr>
<tr>
<td>eat tortillas</td>
<td>exist, be located</td>
</tr>
<tr>
<td>ak’</td>
<td>boj</td>
</tr>
<tr>
<td>give</td>
<td>cut with machete</td>
</tr>
<tr>
<td>tzak</td>
<td>k’ux</td>
</tr>
<tr>
<td>take, grasp in hand</td>
<td>eat beans, crunchy things</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>lo</td>
<td>lo</td>
</tr>
<tr>
<td>eat fruit, soft things</td>
<td></td>
</tr>
</tbody>
</table>
Experiments: where do we fit in?

Before experiment design (what data can falsify comp. assumptions?)

After experiment (fine structure of data; alternative explanations)

Some adult nouns can be child verbs.

Turkish nouns do not cluster well (Ketrez, 2003).

Reporting (part of speech counts aren't informative)

   real data (in temporal sequence)

   commonly: noun group, verb group, POS,
Falsifying computationalist assumptions

Some short verbs are not learned early even when they are frequent and unambiguous.

Some frequently-used ambiguous long nouns can be learned early.

Infrequent but short nouns can be learned early.

Some ambiguous but short nouns can be learned early.
Conclusion

Word use is rule use. Rule use is statistical but not arbitrary.

We all agree that something narrows down the child’s hypothesis space.

Computationalism is empirically easy to test with child data.

Computationalism cannot be pigeon-holed to representationalism.

It is about nature of tasks and complexity of task-specific knowledge.

Abstracting away from obvious explanations to nouns-first acq.

and no evidence for verbs-first acquisition

We would like to take part in experiments.
Thank you.

Thanks: Cogs 541 class at ODTÜ (Lang Acq.), co-taught with Deniz Zeyrek Seçkin Can and Orhan Demir (for data pointers)
Participants of ODTÜ Cogsci workshop on Psycholinguistics (Sept. 2010)
Belma Haznedar
Duygu Özge
Theo Marinis
References


