Language, cognition and computation $\emptyset$: What can a child do to the rules?

Cem Bozşahin

Middle East Technical University (ODTÜ)

bozsahin@metu.edu.tr

Cognitive Science Colloq., Boğaziçi University

January 10, 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)

Time complexity function	size n 10	20	30	40	50	60
n	.00001 second	.00002 second	.00003 second	.00004 second	.00005 second	.00006 second
n^2	.0001 second	.0004 second	.0009 second	.0016 second	.0025 second	.0036 second
n^3	.001 second	.008 second	.027 second	.064 second	.125 second	.216 second
n^5	.1 second	3.2 seconds	24.3 seconds	1.7 minutes	5.2 minutes	13.0 minutes
2^n	.001 second	1.0 second	17.9 minutes	12.7 days	35.7 years	366 centuries
3^n	.059 second	58 minutes	6.5 years	3855 centuries	2×10^8 centuries	1.3×10^{13} centuries

Objects of computing

- strings
  - algebra
    - formal language theory
  - calculus
    - parsing and automata theory
- functions
  - algebra
    - domain and 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	Det	$\rightarrow$	every
	NP	$\rightarrow$	Name	N	$\rightarrow$	chemist
	NP	$\rightarrow$	Det N	Name	$\rightarrow$	Kafka
	VP	$\rightarrow$	V_{iv}	V_{iv}	$\rightarrow$	arrived
	VP	$\rightarrow$	V_{tv} NP	V_{tv}	$\rightarrow$	adored

NP=S/VP and **VP=S\NP**. Hence **NP=S/(S\NP)**

S	$\rightarrow$	NP VP	Det	$\rightarrow$	every
NP	$\rightarrow$	Name	N	$\rightarrow$	chemist
NP	$\rightarrow$	Det N	Name	$\rightarrow$	Kafka
VP	$\rightarrow$	V_{iv}	V_{iv}	$\rightarrow$	arrived
VP	$\rightarrow$	V_{tv} NP	V_{tv}	$\rightarrow$	adored

(2) **V_{tv}=VP/NP** **V_{iv}=VP** **NP=VP\\V_{tv}**

NP=S/VP **VP=S\\NP** Det=**NP/N**

Name=**NP** **N=NP**\Det

Hence **V_{tv}=(S\\NP)/NP** **V_{iv}=S\\NP**

NP=(S\\NP)\\((S\\NP)/NP)

NP=S/(S\\NP)

S	$\rightarrow$	NP VP	Det	$\rightarrow$	every
NP	$\rightarrow$	Name	N	$\rightarrow$	chemist
NP	$\rightarrow$	Det N	Name	$\rightarrow$	Kafka
VP	$\rightarrow$	V_{iv}	V_{iv}	$\rightarrow$	arrived
VP	$\rightarrow$	V_{tv} NP	V_{tv}	$\rightarrow$	adored

(3) every := Det = $\mathbf{NP}/\mathbf{N}$ = $\mathbf{(S/(S\backslash NP))/N}$

chemist := N = $\mathbf{NP}\backslash\mathrm{Det}$ = $\mathbf{NP\backslash(NP/N)}$

Kafka := Name = $\mathbf{NP}$ = $\mathbf{S/VP{=}S/(S\backslash NP)}$ and $\mathbf{(S\backslash NP)\backslash((S\backslash NP)/NP)}$

arrived := $\mathbf{VP = S\backslash NP}$

adored := $\mathbf{VP/NP = (S\backslash NP)/NP}$

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: $O(n^2)$ $\sum_{i=1}^{n} i$

more constraints: short and more frequent strings first

Time complexity function	size n 10	20	30	40	50	60
n	.00001 second	.00002 second	.00003 second	.00004 second	.00005 second	.00006 second
n^2	.0001 second	.0004 second	.0009 second	.0016 second	.0025 second	.0036 second
n^3	.001 second	.008 second	.027 second	.064 second	.125 second	.216 second
n^5	.1 second	3.2 seconds	24.3 seconds	1.7 minutes	5.2 minutes	13.0 minutes
2^n	.001 second	1.0 second	17.9 minutes	12.7 days	35.7 years	366 centuries
3^n	.059 second	58 minutes	6.5 years	3855 centuries	2×10^8 centuries	1.3×10^{13} centuries

A thought experiment

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

(4) Eat veggies

(5) a. *eat*:=**S**$/$**NP:**eat' *veggies*:=**NP:**veg'
b. *eat*:=**NP:**eat' *veggies*:=**S**$\backslash$**NP**: $\lambda x.veg'x$
c. *eat*:=**NP:**veg' *veggies*:=**S**$\backslash$**NP**: $\lambda x.eat'x$
d. **eat*:=**NP:**eat' *veggies*:= **S**$/$**NP**: $\lambda x.veg'x$
e. **eat*:=**S**$\backslash$**NP:**$\lambda x.eat'x$ *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)

Experience 1				Experience 2 (with chocolate)			
Eat veggies				No veggies			
eat := $\mathbf{S/NP}$	$:eat'$	veggies := $\mathbf{S\backslash NP}$	$:veg'$	no := $\mathbf{S/NP}$	$:no'$	veggies := $\mathbf{S\backslash NP}$	$:no'$
	$:veg'$		$:eat'$		$:veg'$		$:veg'$
$\mathbf{NP}$	$:eat'$	$\mathbf{NP}$	$:veg'$		$:choc'$		$:choc'$
	$:veg'$		$:eat'$	$\mathbf{NP}$	$:no'$	$\mathbf{NP}$	$:no'$
					$:veg'$		$:veg'$
					$:choc'$		$:choc'$

(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**/**NP**	:veg'	gone :=	**S**$\backslash$**NP**	:veg'
			:$gone'$			:$gone'$
		NP	:veg'		**NP**	:veg'
			:$gone'$			:$gone'$

The child is $\frac{3}{6}$ likely to believe the connection is mediated by **NP**, $\frac{2}{6}$ by **S**$\backslash$**NP**, and $\frac{1}{6}$ by **S**/**NP**, in just three scenes.

Knowledge of words as subpieces of grammar and its use.

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

(11) veggies := $\{\mathbf{S}\backslash\mathbf{NP}{:}\mathit{veg}'@\frac{1}{14}$...

$$
\begin{aligned}
\text{veggies} := \{ & \mathbf{S}\backslash\mathbf{NP}{:}\mathit{veg}'@\tfrac{2}{14}, && \mathbf{S}\backslash\mathbf{NP}{:}\mathit{eat}'@\tfrac{1}{14}, \\
& \mathbf{S}\backslash\mathbf{NP}{:}\mathit{no}'@\tfrac{1}{14}, && \mathbf{S}\backslash\mathbf{NP}{:}\mathit{choc}'@\tfrac{1}{14}, \\
& \mathbf{S}/\mathbf{NP}{:}\mathit{veg}'@\tfrac{1}{14}, && \mathbf{S}/\mathbf{NP}{:}\mathit{gone}'@\tfrac{1}{14}, \\
& \mathbf{NP}{:}\mathit{veg}'@\tfrac{3}{14}, && \mathbf{NP}{:}\mathit{eat}'@\tfrac{1}{14}, \\
& \mathbf{NP}{:}\mathit{no}'@\tfrac{1}{14}, && \mathbf{NP}{:}\mathit{choc}'@\tfrac{1}{14}, \quad \mathbf{NP}{:}\mathit{gone}'@\tfrac{1}{14} \\
\}
\end{aligned}
$$

Keren's first words (Dromi, 1987) Hebrew (Israel)

Age m(d)	Child's word	conven. form					
10(12)	haw	(?)	a dog's bark	12(16)	hita	(?)	going out for a walk
11(16)	?aba	(aba)	Father	12(18)	tiktak	(?)	sound of clock
11(17)	?imaima	(?)		12(19)	cifcif	(?)	bird's tweet
11(18)	ham	(?)	said while eating	12(20)	hupa	(?)	accom. making sudden contact w/ground
12(3)	mu	(?)	a cow's moo	12(23)	dio	(dio)	giddi up
12(3)	?ia	(?)	a donkey's bray	12(25)	hine	(hine)	here
12(8)	pil	(pil)	an elephant	12(25)	?ein	(?ein)	all gone
12(11)	buba	(buba)	a doll	12(25)	na?al	(na?al)	a shoe
12(13)	pipi	(pipi)	urine	12(25)	myau	(?)	a cat's meow

Tad's first words (Gentner, 1982) AmE

Age (m.)					
11	dog	16	eye	19	down
12	duck	18	cow		boo
13	daddy		bath		bottle
	yuk		hot		up
	mama		cup		hi
	teh (teddy bear)		truck		spoon
	car	19	kitty		bye
14	dipe (diaper)		pee pee		bowl
	toot toot (horn)		happy		uh oh
	owl		oops		towel
15	keys		juice		apple
	cheese		TV		teeth

Mik and Xan's first verbs (Brown, 1998) VSO Tzeltal

Mik (1;5–2;0)		Xan (1;3–2;2)	
ba	go/allgone	we'	eat tortillas
la'	come!	chu'	suckle breast
we'	eat tortillas	ay	exist, be located
ak'	give	boj	cut with machete
tzak	take, grasp in hand	k'ux	eat beans, crunchy things
⋮	⋮	⋮	⋮
lo	eat fruit, soft things	lo	

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.

Computationalism is empirically easy to test with child data.

Computationalism cannot be pigeon-holed to representationalism.

We work on **what** representations stand for, not **how** they are represented on wetware (but NB. Elman 1990)

Weak computationalism is essentially functional, because we are not in the business of constructing minds,

only understanding how it works, given limited perceptive abilities (across species).

It is not behaviorist. We aim to understand **interaction** of internal processes and the external world, and task-specificity of knowledge

with as few auxiliary assumptions as possible.

Thank you.

Hope to see you throughout the spring term.

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

Anderson, John R. 1976. *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Bickhard, Mark H. 1996. "Troubles with computationalism." In W. O'Donohue and R. Kitchener, eds., *Philosophy of psychology*, 173–183. London: Sage.

Brown, Penelope. 1998. "Children's first verbs in Tzeltal: evidence for an early verb category." *Linguistics*, 36, 4, 713–753.

Çöltekin, Çağrı, and Cem Bozsahin. 2007. "Syllable-based and Morpheme-based Models of Bayesian Word Grammar Learning from CHILDES Database." In *Proc. of the 29th Annual Meeting of Cognitive Science Society*. Nashville, TN.

Dromi, Esther. 1987. *Early lexical development*. Cambridge University Press.

Elman, Jeffrey. 1990. "Finding Structure in Time." *Cognitive Science*, 14, 179–211.

Garey, M.R., and D.S. Johnson. 1979. *Computers and intractability: A guide to NP-completeness*. San Francisco: W.H. Freeman.

Gentner, Dedre. 1982. "Why nouns are learned before verbs: Linguistic Relativity versus Natural Partitioning." In Stan A. Kuczaj II, ed., *Language Development, vol.2: Language, Thought and Culture*, 301–334. Hillsdale, New Jersey: Lawrence Erlbaum.

Ketrez, Nihan F. 2003. "Distributional properties of nouns and verbs in Turkish child-directed speech." *USC Working papers in Linguistics*, 121–144.

Kwiatkowksi, Tom, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. "Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification." In *Proc. of the Conf. on Empirical Methods in Natural Language Processing*. Cambridge, MA.

Monaghan, Padraic, and Morten H. Christiansen. 2004. "What distributional information is useful and usable for language acquisition." In *Proc. of the 26th Annual Meeting of Cognitive Science Society*, 963–968. Mahwah: NJ.

Siskind, Jeffrey. 1996. "A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings." *Cognition*, 61, 39–91.

Steedman, Mark, and Julia Hockenmaier. 2007. "The Computational Problem of Natural Language Acquisition." Ms., University of Edinburgh.

Tardif, Twila. 1996. "Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies." *Developmental Psychology*, 32, 3, 497–504.

Tenenbaum, Joshua B., and Fei Xu. 2000. "Word learning as Bayesian inference." In *Proc. of the 22nd Annual Conf. of the Cognitive Science Society*. Philadelphia.

Zettlemoyer, Luke S., and Michael Collins. 2005. “Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars.” In *Proc. of the 21st Conf. on Uncertainty in Artificial Intelligence*. Edinburgh.