Input & Universal Grammar

Charles Yang
Department of Linguistics, Computer Science & Psychology
Institute for Research in Cognitive Science
University of Pennsylvania

SLRF Formal and Functional Symposium
Pittsburgh 2012

Input & Universal Grammar

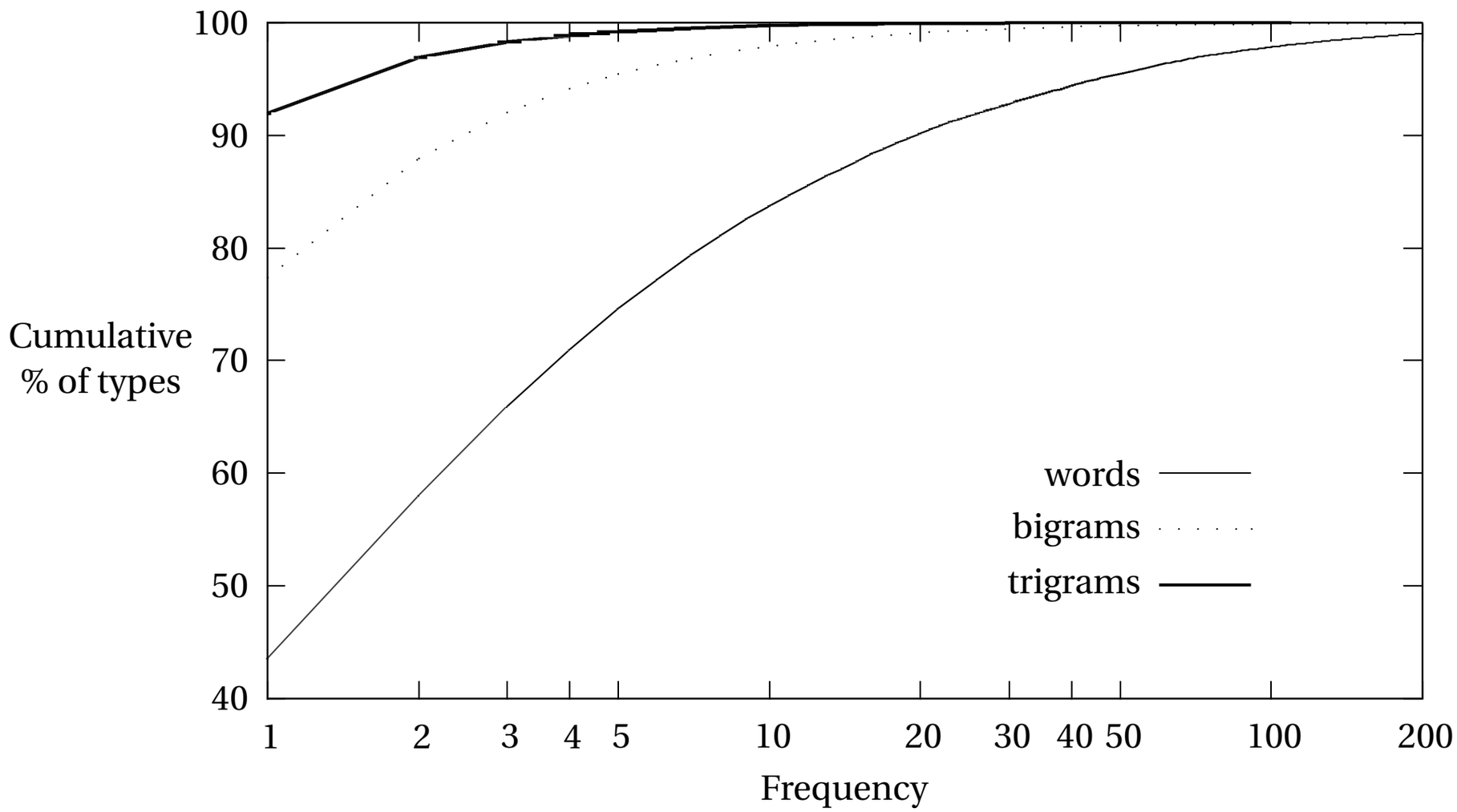
- Input: A statistical look at distributional information
- Output: Quantitative and cross-linguistic patterns of development
- Mechanisms of learning: Why UG can make use of input-driven, probabilistic, and domain general learning processes
- Some speculations on L2 acquisition

Input & Usage Effects?

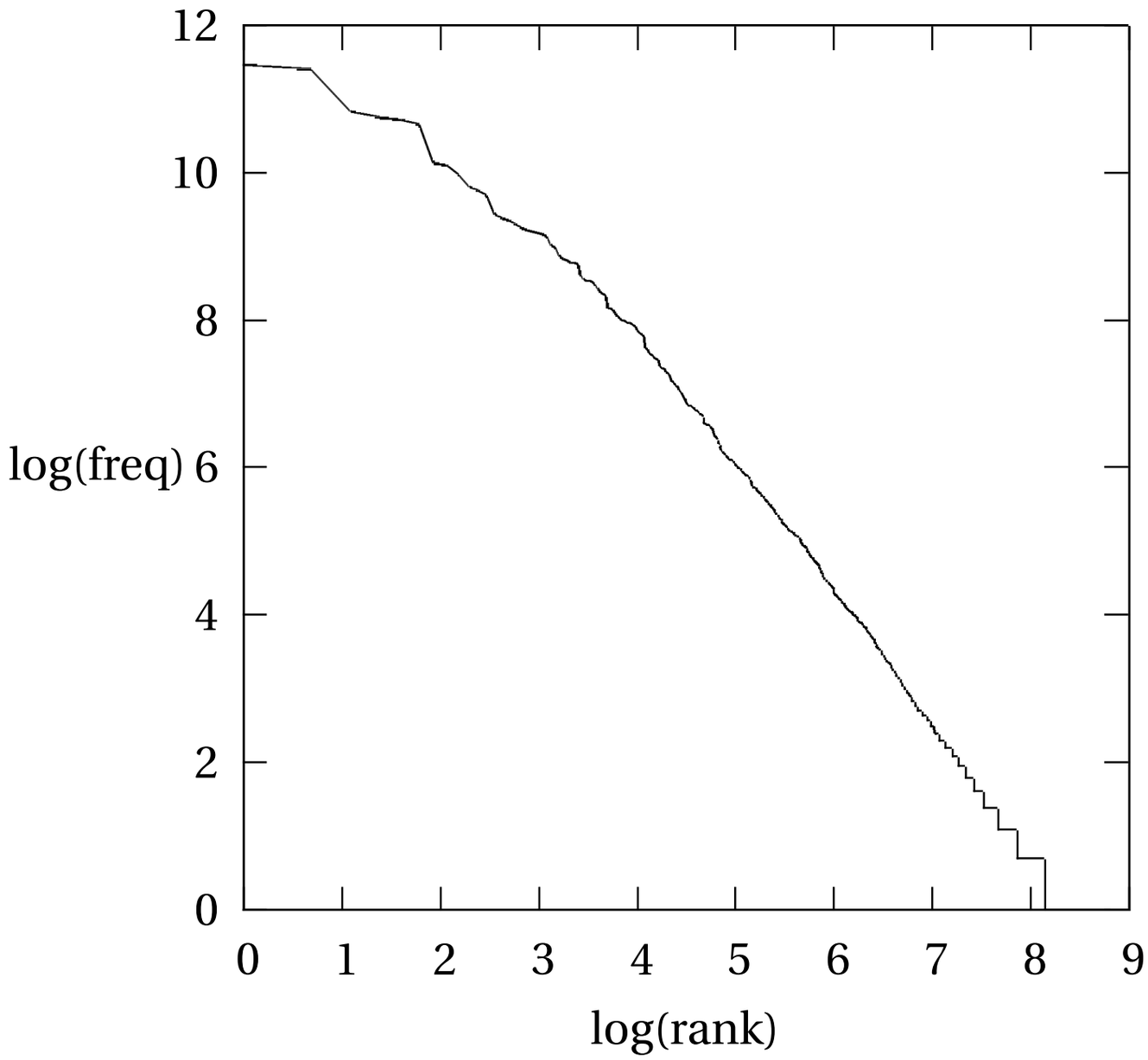
- Frequency effects, limited extent of diversity ("verb islands"), etc.
- "**give me X**", a highly frequent expression, is often cited as evidence of the child using formulaic expressions
- From the Harvard children (Adam, Eve, Sarah)
  - **give me**: 93, **give him**: 15, **give her**: 12, or **7.75 : 1.23 : 1**
  - **me**: 2870, **him**: 466, **her**: 364, or **7.88 : 1.28 : 1**

Input: **Very** boring

- Zipf's law: Much of language is repetitions of a few, while most distinct items occur rarely
- Linguistic combinations produce an even large space of possibilities (e.g., bigrams, trigrams, morphology, rules/ constructions)



Wall Street Journal



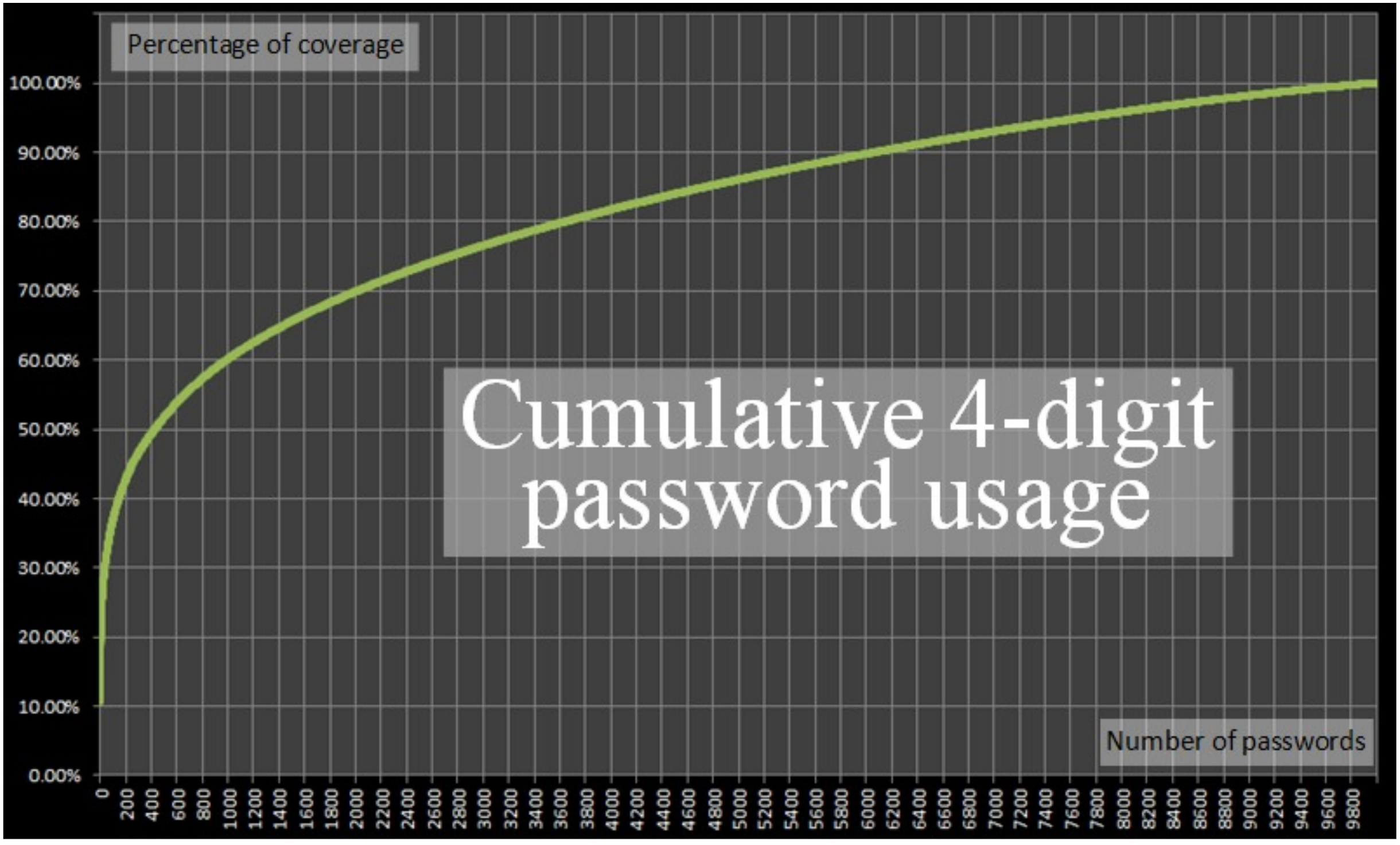
- Rules and their frequencies from the Penn Treebank (log-log scale)

Verb Islands in adult language (>1Mil)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
put	**401**	**164**	**124**	**15**	**12**	**12**	**11**	**10**	**8**	**5**
tell	**245**	65	49	49	45	36	22	16	14	13
see	**152**	100	38	32	28	21	14	14	12	11
want	**158**	83	36	24	19	15	13	9	5	4
let	**238**	38	32	23	22	17	8	6	3	3
give	**115**	92	59	32	31	7	5	5	5	5
take	**130**	57	30	21	18	15	14	9	8	7
show	**100**	34	27	21	19	17	12	8	7	7
got	**58**	37	14	12	11	9	7	7	7	4
ask	**45**	41	27	24	12	10	8	8	4	2
make	**67**	20	12	10	9	7	7	4	3	2
eat	**67**	42	14	8	6	5	5	3	3	3
like	**39**	13	9	6	4	4	4	4	3	3
bring	**43**	30	17	15	10	10	3	3	3	3
hear	**46**	22	13	9	6	4	4	3	3	3
total	**1904**	**838**	**501**	**301**	**252**	**189**	**137**	**109**	**88**	**75**

put: it, your, them, him, my, her, em, you, his, water

PIN Number Analysis



Matches and ***Mismatches***

- Roger Brown (1973, *A First Language*): word order errors are "triflingly few"
  - Children must be able to learn the basic rules of grammar with 2-3 million sentences
- Yet a great deal of surprises remain, especially if we relate them to the distributions of linguistic patterns in the input

Abundant Input, Late Learning

- Missing subjects in child English (Bloom 1973, Hyam 1986)
  - __ want look a man.
- Missing objects as well (Wang et al. 1992)
  - Look at __. __ go a little higher
- Null subject stage last about 3 years but an overwhelming amount of child directed English input do contain the subject, as English is an obligatory subject language (unlike Chinese, Japanese, Spanish, Italian, etc.)

Lateness is not Universal

		English	Italian	Chinese
Adults	**Subject**	**~ 0%**	**70%**	**50%**
	Object	**0%**	**0%**	**20%**
Children	**Subject**	**30%**	**~70%**	**~50%**
	Object	**8%**	**0%**	**~20%**

Children age: <3;0

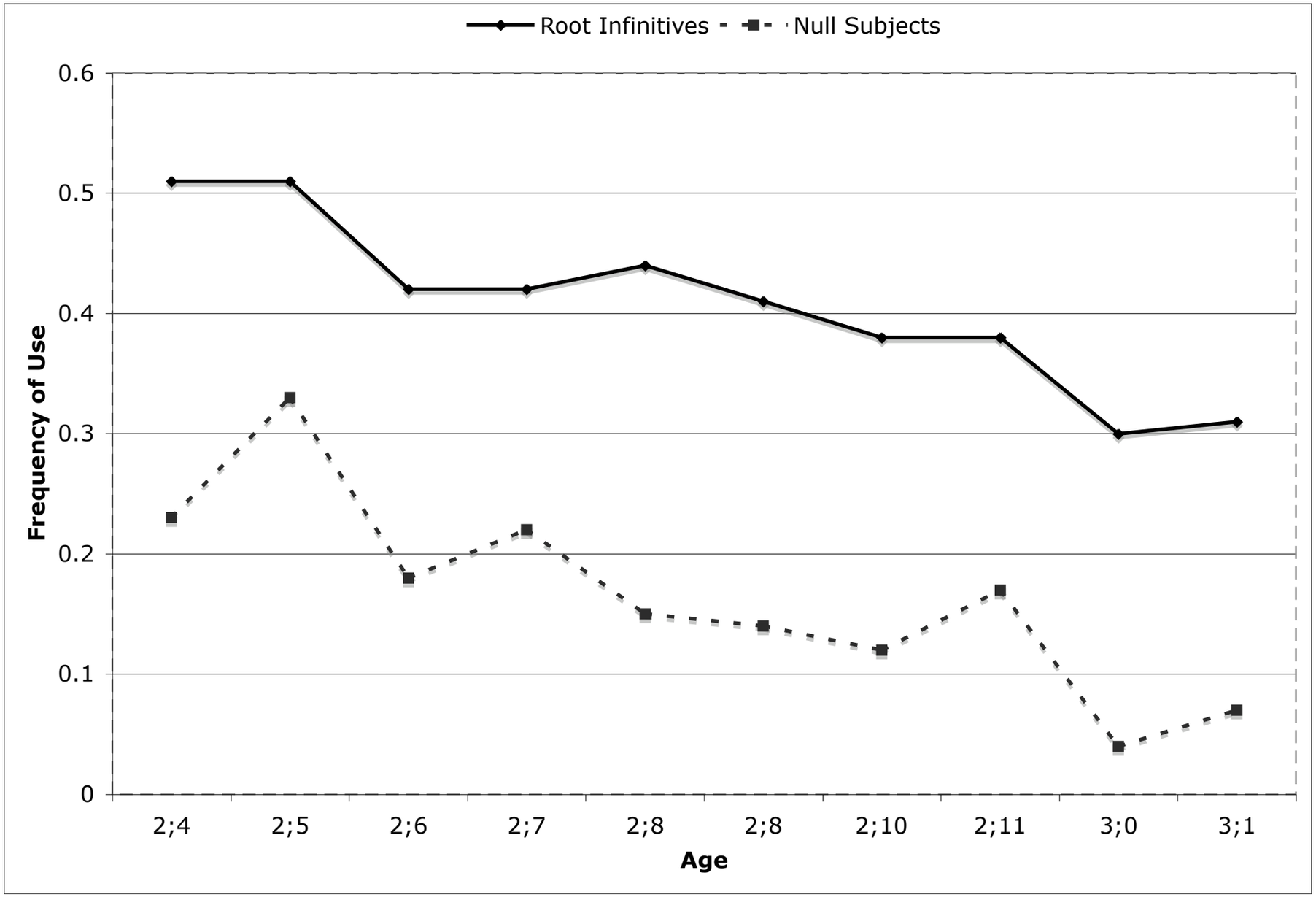
Data from Wang et al. (1992), Valian (1991), Bates (1978)

“Luke, look at the input ...”

Abundant Input, Late Learning

- Extensive use of Root Infinitives that should be tensed
  - English: Papa have it.
  - Dutch: thee drinken (tea drink-$_{INF}$)
  - French: Dormir petit bébé (sleep-$_{INF}$ little baby)
  - German: mein Kako hinstelln (my chocolate milk put-$_{INF}$)
  - Hebrew: Malon lauf (balloon fly-$_{INF}$)

Optional Infinitives and Null Subjects Together



Large dataset from a Dutch learner (data from Haegeman 1996)

Little Input, Early Learning

- The placement of verbs in French
  - Jean voit souvent/pas Claude. (“John sees often/not Claude”)
- Only 7% of the sentences in child-directed French show this pattern (Yang 2002), yet children learn this property of French by the time of two word combinations (1;8, Pierce 1992)
  - marches pas (“works not”)
  - pas la poupée dormir (“not the doll sleep”)
- Similar findings in similar languages, and languages like English pattern very differently

Same Grammar, Differential Learning

- Germanic languages have Verb Second (V2)
  - Dutch: Dit boek las ik gisteren. (“this book read I yesterday”)
  - Norwegian: Det vet æ ikkje. (“that know I not”)
- But Dutch and German children take over 3 years to use V2 reliably as they produce a lot of verb initial utterances (Clahsen 1986, Haegeman 1996), while Norwegian children learn V2 as early as 2;0 (Westergaard 2009)

Central Questions

- Can learning primarily consist of memorization and lexically specific rules?
  - For detailed assessment of usage-based learning, seeYang (2011, *Proc. Assoc. Comp. Ling.*)
- What combination of grammar model and learning model will give the best account of child language?

Industrial Lessons

- Statistical parsing: Learn from pre-parsed tree structures (e.g., Wall Street Journal, Brown Corpus)
- Start: a large set of probabilistic CFG rules

$$\text{S} \xrightarrow{p} \text{NP VP}, \ \text{S} \xrightarrow{1-p} \text{NP VP}$$

- Training: adjust the probabilities of rule expansions so that they maximize the likelihood of the training data
- Testing: run the resulting grammar on new data
- State of art parsers: low 90% (impressive but still a long way to go)

Why Google hasn't solved everything

(a) $VP \rightarrow V\ NP$

(b) $VP \rightarrow V_{drink}\ NP$

(c) $VP \rightarrow V_{drink}\ NP_{water}$

Rule Type	F-score
a+b+c	~89.0%
a+b	~88.4%
a	~84%

- Multiple forms of rules are present in training, ranging from general to lexical
- One can vary certain types to test their effectives in generalization (Gildea 2001 *Proc. ACL*, Bikel 2004, *Comp. Ling.*)
- Storing construction or lexically specific rules offers virtually no payoff in data coverage (Yang 2011, *Proc. ACL*)
- The range of grammar (output) is enormous, but the learning data (input) is limited and grows far too slowly

Are the best rules good enough?

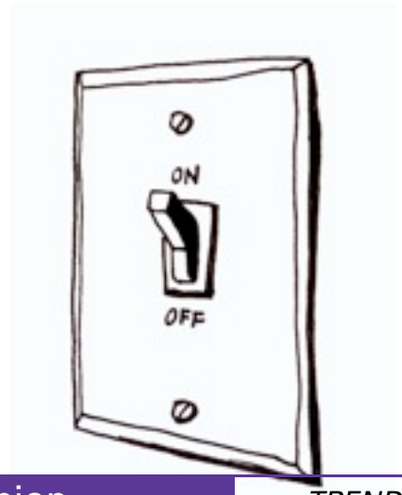
- Formal learnability is one thing; the developmental test from child language is ultimately more important
- S→ NP VP will be learned quickly: >95% of the English data
- VP→ V_{FIN} *pas* will be learned slowly: 7% of the French data
- But French children learn verb placement ***early*** and English children learn the use of subjects ***late***!
- What kind of (grammar, learning) combination would take the input and produce the output like children?

UG + Learning from Input

- Parameters ≈ Principal Component Analysis
- "Child competence is identical to adult competence"
- "Parameters are set very early"
- Magic and More Magic ...

- Use parameters
  - a model of language variation and child learning errors
- Do not use Magic
  - use a model of learning that is gradual and takes input into account

From Trigger to Dimmer



ELSEVIER
Opinion
TRENDS in Cognitive Sciences Vol.8 No.10 October 2004

Universal Grammar, statistics or both?

- The Variational model (Yang 2002 Oxford UP)
- Parameter values are associated with probabilities (p: VO, 1-p: OV)
  - try out a value, reward/punish) based on success/failure
  - learning rate: magnitude of change, subject to individual variation
  - More tokens of parameter **signatures**, faster learning

Signatures & Learning

- Verb raising in French: 7% input, very early acquisition
- Learning the use of subject in English
  - Hearing "I eat pizza" doesn't no good because it does not disambiguate the types of grammars the learner considers
- Expletive subject sentences
  - "There is a cookie on the floor" (1%)
- Signature for Chinese-type topic drop: null objects (12%)
- The most comprehensive study of a realistic parameter domain (Fodor & Sakas 2012 *Language Acquisition*) shows that most if not all parameters have signatures, which make learning feasible

From Input to Output

TABLE 1 | Statistical Correlates of Parameters in the Input and Output of Language Acquisition

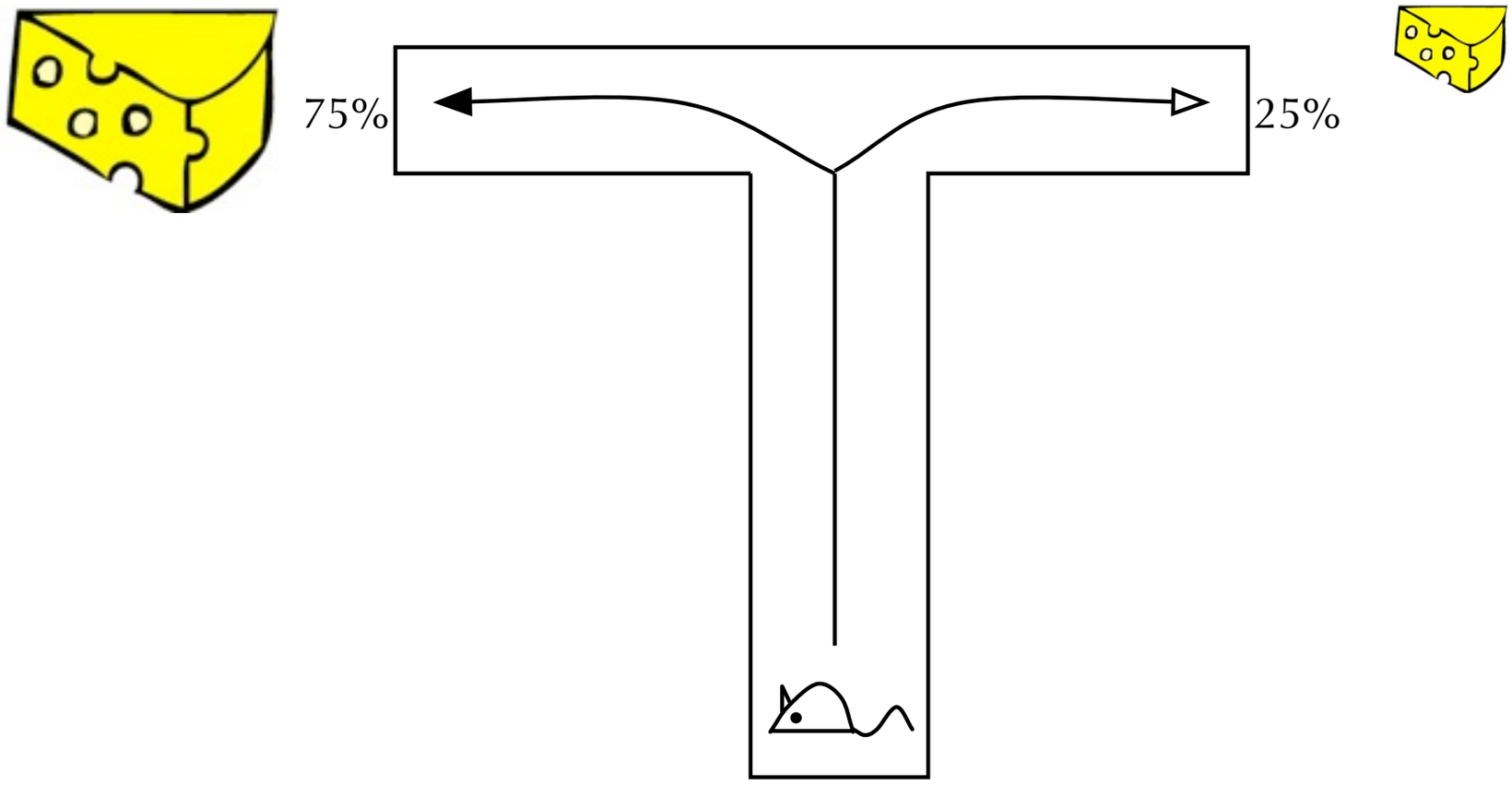
Parameter	Target	Signature	Input Frequency (%)	Acquisition
Wh fronting	English	Wh questions	25	Very early
Topic-drop	Chinese	Null objects	12	Very early
Prodrop	Italian	Null subjects in questions	10	Very early
Verb raising	French	Verb adverb/*pas*	7	1.8
Obligatory subject	English	Expletive subjects	1.2	3.0
Verb second	German/Dutch	OVS sentences	1.2	3.0–3.2
Scope marking	English	Long-distance questions	0.2	>4.0

- Parameters have developmental correlates (Yang 2012, *WIREs Cognitive Science*)
- Same parameter, different languages:
  - V2 in Norwegian: **10%** of OVS in input→**early**
  - V2 in Dutch: **1.2%** of OVS input→**late**

Input & Individual Variation

- Never been denied (Chomsky 1965, Wexler & Culicover 1980, Berwick 1985, Gibson & Wexler 1994, Yang 2002)
- No need to appeal to unmotivated and unnecessary theoretical machinery to account for the gaps between children and adults
- Optional Infinitives: verbal morphology that mark tense-you are **not** learning Chinese (Legate & Yang 2007 *Lg. Acq.*)
- Individual level correlation between length of OI stage and the amount of tensed morphology in CDS (Hadley *et al.* 2011 *JSLHR*)
  - Suggests that the source of delay in SLI children may be due to (more general) learning: poor morphological learner (Leonard et al 1992, Rice et al. 2000) make less effective use of the tense information to unlearn the RI usage

Learning & Learning Language



- Variational Model uses Reinforcement Learning (Bush & Mosteller 1951), a very general learning mechanism with broad behavioral and neural support
- Strongly demonstrated in human subjects (children and adults)
  - See especially the work in the acquisition of sociolinguistic variables (Labov and co.)

L2 Acquisition: Re-turning the dimmer?

- The combination of grammar model and learning model
- The Variational Model provides a precise and testable hypothesis for L2 acquisition research
  - Even if the grammar model is **not** parameter based
    - If initial state is L1-independent:
      - L2 learners mirror the time course trajectories of L1 learners
    - If initial state is L1:
      - L2 learners will eschew the time course of L1 learners
    - It's **not** sufficient to study a single parameter: cross-parameter comparison is necessary

Conclusion

- Input is rich and interesting, but it alone won't do the job
  - Also need to be cautious about drawing conclusions from input effects
- Input and Universal Grammar are perfectly consistent
  - Previous conception of the **learning mechanism** needs to be reconsidered
- Input effects in L2 likewise may be assessed accurately by making the grammar-learning interaction very explicit