Psych 215L: Language Acquisition

Lecture 15
Poverty of the Stimulus:
Structure Dependence

Reminder: Poverty of the Stimulus

The Logic of Poverty of the Stimulus (The Logical Problem of Language Acquisition)

1) Suppose there are some data.
2) Suppose there is an incorrect hypothesis compatible with the data.
3) Suppose children behave as if they never entertain the incorrect hypothesis.

   Addendum (interpretation): Or children converge on the correct hypothesis much earlier than expected (Legate & Yang 2002).

Conclusion: Children possess innate knowledge ruling out the incorrect hypothesis from the hypothesis space considered.

Addendum (Interpretation): The initial hypothesis space does not include all hypotheses. Specifically, the incorrect ones of a particular kind are not in the child's hypothesis space.

Legate & Yang (2002): Poverty of the Stimulus Lives

Child Input

Very frequent
Is Hoggle t_{is} running away from Jareth?

Very infrequent, if ever
Can someone who can solve the Labyrinth t_{can} show someone who can't how?

Perfors, Tenenbaum, & Regier (2011): Or does it?

Some Issues

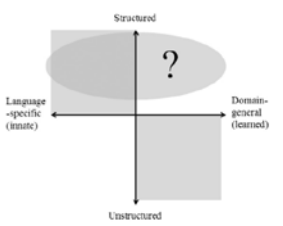
Unclear how much evidence is "enough". Forms do occur, even if they do so rarely. Moreover, may be better to consider forms not in isolation, but in a larger context.

"Our findings suggest that it is vital to consider the learnability of entire candidate grammars holistically. While crucial data that would independently support any one generalization (such as the auxiliary-fronting rule) may be very sparse or even nonexistent, there may be extensive data supporting other, related generalizations; this can bias a rational learner towards making the correct inferences about the cases for which the data is very sparse….The need to acquire a whole system of linguistic rules together imposes constraints among the rules, so that an a priori unbiased learner may acquire constraints that are based on the other linguistic rules it must learn at the same time."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Some Issues

It's possible to have both domain-general learning abilities and structured representations.



Perfors, Tenenbaum, & Regier (2011): Or does it?

Some Issues

Previous statistical accounts haven't connected with the argument that preferring hierarchical structures must be innate.

"PoS arguments begin with the assumption – taken by most linguists as self-evident – that language does have explicit hierarchical phrase structure, and that linguistic knowledge must at some level be based on representations of syntactic categories and phrases that are hierarchically organized within sentences. The PoS arguments are about whether and to what extent children's knowledge about this structure is learned via domain-general mechanisms, or is innate in some language-specific system. Critiques based on the premise that this explicit structure is not represented as such in the minds of language users do not really address this argument..."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Some Issues

Previous statistical accounts also are somewhat difficult to interpret.

"For instance, the networks used by Reali and Christiansen (2005) and Lewis and Elman (2001) measure success by whether they predict the next word in a sequence or by comparing the prediction error for grammatical and ungrammatical sentences. These networks lack not only a grammar-like representation; they lack any kind of explicitly articulated representation of the knowledge they have learned. It is thus difficult to say what exactly they have learned about linguistic structure – despite their interesting linguistic behavior once trained."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Some Issues

Working within an ideal learner framework, to show the inference is possible from the data. It remains to be seen whether it's possible for children, given their memory and processing limitations.

"We are not proposing a comprehensive or mechanistic account of how children actually acquire language...setting this challenge aside allows us to focus with more clarity on those aspects of learnability that classic PoS arguments address: claims about what data might be sufficient for learning, or what language-specific prior knowledge must be assumed in order to make learning possible…If we can show that such learning is in principle possible, then it becomes meaningful to ask the algorithmic-level question of how a system might successfully and in reasonable time search the space of possible grammars to discover the best-scoring grammar."

Perfors, Tenenbaum, & Regier (2011): Or does it?

A depiction of the Poverty of the Stimulus

"…many versions of the PoS argument assume that the T is language-specific: in particular, that T is the knowledge that linguistic rules are defined over hierarchical phrase structures."

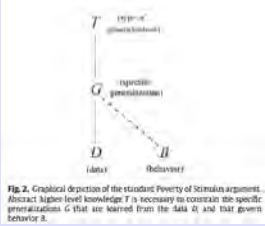
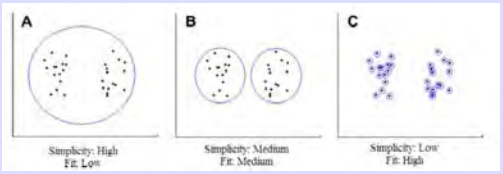


Fig. 2. Graphical depiction of the standard Poverty of Stimulus argument. Abstract higher-level knowledge T is necessary to constrain the specific generalizations G that are learned from the data D, and that govern behavior B.

Perfors, Tenenbaum, & Regier (2011): Or does it?

Bayesian learning: Tradeoffs

A Bayesian learner finds a balance between fit to the data (likelihood) and simplicity of the explanation (hypothesis prior). (Would prefer the middle one below)

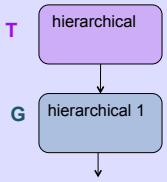


Perfors, Tenenbaum, & Regier (2011): Or does it?

Bayesian Model Selection

First, pick a type of grammar T (ex: linear, regular, hierarchical).

Then, pick an instance of T, G, from which the data D are generated.

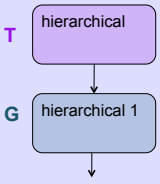


"Is the dwarf who is being teased grumpy?"

Perfors, Tenenbaum, & Regier (2011): Or does it?

Posterior probability of G and T, given D

$$p(G,T|D) \propto_p (D|G)_p(G|T_p(T)$$

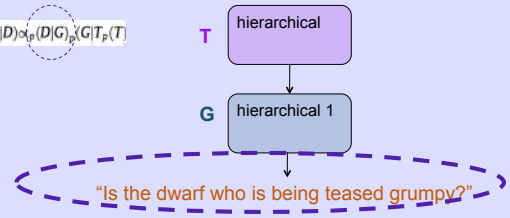


"Is the dwarf who is being teased grumpy?"

Perfors, Tenenbaum, & Regier (2011): Or does it?

Posterior probability of G and T, given D

$p(G,T|D) \propto_p (D|G)_p(G|T_p(T)$

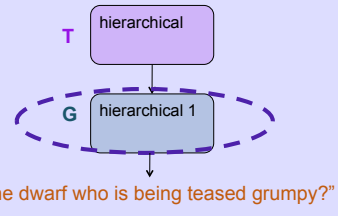


is proportional to the probability of generating the data from G [p(D | G)]

Perfors, Tenenbaum, & Regier (2011): Or does it?

Posterior probability of G and T, given D

$p(G,T|D) \propto_p (D|G)_p(G|T_p(T)$



is proportional to the probability of generating the data from G [p(D | G)], multiplied by the probability of G, given the type of grammar T chosen [p(G|T)].

Perfors, Tenenbaum, & Regier (2011): Or does it?

The Corpus, slightly simplified

Adam corpus (American English), each word (mostly) replaced with its syntactic category:

determiners (det) [ex: *the, a, an*]
adjectives (adj) [ex: *adorable, stinky*]
prepositions (prep) [ex: *to, from, of*]
proper nouns (prop) [ex: *Jareth, Sarah, Hoggle*]
infinitives (to) [ex: *to* in *I want to go*]
participles (part) [ex: She would have *gone*, I'm *going*]
infinitive verbs (vinf) [ex: I want to *go*]
auxiliary verbs (aux) [ex: he *can* go]
complementizers (comp) [ex: I thought *that* I should go.]
wh-question words (wh) [ex: *what* are you doing]
nouns (n) [ex: *cat, penguin, dream*]
comments (c) [ex: *mmhm*]
pronouns (pro) [ex: *he, she, it, one*]

Adverbs (ex: *too, very*) and negations (ex: *not*) were removed from all sentences.

Perfors, Tenenbaum, & Regier (2011): Or does it?

The Corpus, slightly simplified

Ungrammatical and the most complex grammatical sentences were also removed: (available at http://www.psychology.adelaide.edu.au/personalpages/staff/amyperfors/research/cognitionpos/index.html).

Note: This biases the model against the more complex hierachical grammars.

topicalized sentences
ex: "Here he is."

(some) sentences with subordinate clauses
ex: "if you want to."

(some) sentential complements
ex: "He thought that she ought to watch the movie."

conjunctions (ex: *and, or, but*)

serial verb constructions
ex: "You should go play outside."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Test corpora

Separate by frequency (idea: less complex sentences occur more frequently)

Level 1 (500+ times) = 8 unique types
Level 2 (300+ times) = 13 types
Level 3 (100+ times) = 37 types
Level 4 (50+ times) = 67 types
Level 5 (10+ times) = 268 types
Level 6 (complete corpus) = 2338 unique types, including interrogatives, wh-questions, relative clauses, prepositional and adverbial phrases, command forms, and auxiliary as well as non-auxiliary verbs.

Relative clauses might be particularly useful for Y/N questions with multiple auxiliaries.

Perfors, Tenenbaum, & Regier (2011): Or does it?

The grammars

Structure-dependent, hierarchical grammar (smaller & larger): represented with context-free phrase structure rules

Context-free grammar CFG-S

NP → NP PP | NP CP | NP C | N | det N | adj N | pro | prop
N → n | adj N

Context-free grammar CFG-L

NP → NP PP | NP CP | NP C | N PP | N CP | N C | pro PP | pro CP | pro C | prop PP | prop CP | prop C | N | det N | adj N | pro | prop
N → n | adj N

Perfors, Tenenbaum, & Regier (2011): Or does it?

The grammars

Flat grammars: one production per observed sentence (good fit, maximum complexity)

Flat grammar

S → pro aux part
S → adj n aux n prep det n
S → det n v n
S → pro aux adj n comp pro v

Perfors, Tenenbaum, & Regier (2011): Or does it?

The grammars

Regular grammars: hierarchical branching in only one direction (rules: A → a or A → aB) , with varying levels of complexity & fit

Regular grammar REG-N

NP → pro | prop | n | det N | adj N | pro PP | prop PP | n PP | det N_{PP} | adj N_{PP} | pro CP | prop CP | n CP | det N_{CP} | adj N_{CP} | pro C | prop C | n C | det N_C | adj N_C
N → n | adj N
N_{PP} → n PP | adj N_{PP}
N_{CP} → n CP | adj N_{CP}
N_C → n C | adj N_C

Regular grammar REG-M

NP → pro | prop | n | det N | adj N | pro PP | prop PP | n PP | pro CP | prop CP | n CP | pro C | prop C | n C
N → n | adj N | n PP | n CP | n C

Regular grammar REG-B

HP → pro | prop | n | det N | adj N | pro HP | prop HP | n HP | pro CP | prop CP | n CP | pro C | prop C | n C | prep HP | prep | adj | adj HP | to V_{inf}
N → n | adj N | n HP | n CP | n C

Perfors, Tenenbaum, & Regier (2011): Or does it?

The grammars

The relationship between grammar complexity.

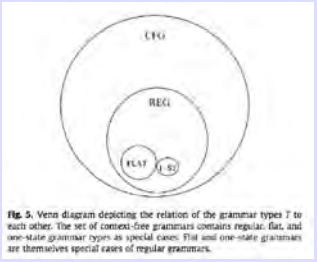


Fig. 5. Venn diagram depicting the relation of the grammar types *T* to each other. The set of context-free grammars contains regular, flat, and one-state grammar types as special cases. Flat and one-state grammars are themselves special cases of regular grammars.

Perfors, Tenenbaum, & Regier (2011): Or does it?

Likelihoods for the grammars

Two component adaptor grammar model of Goldwater et al. (2006) and Johnson et al. (2007)

(1) [grammar] Assign probability distribution over infinite syntactic forms accepted in the language

(2) [adaptor] Generate finite observed corpus from that probability distribution (use power-law generation, so a few syntactic types are very frequent while most are infrequent)

Perfors, Tenenbaum, & Regier (2011): Or does it?

Results based on data types

Log probability = smaller negative number means more probable

Hierarchical grammars preferred once more complex structures (levels 4, 5, and 6) are included in the data to be accounted for.

Corpus	Probability	FLAT	REG-N	REG-M	REG-B	1-ST	CFG-S	CFG-L
Level 1	Prior	−99	−148	−124	−117	−94	−155	−192
	Likelihood	−17	−20	−19	−21	−36	−27	−27
	Posterior	**−116**	−168	−143	−138	−130	−182	−219
Level 2	Prior	−630	−455	−442	−411	−201	−357	−440
	Likelihood	−134	−147	−157	−162	−275	−194	−177
	Posterior	−764	−601	−599	−573	**−476**	−551	−617
Level 3	Prior	−1198	−663	−614	−529	−211	−454	−593
	Likelihood	−282	−323	−333	−346	−553	−402	−377
	Posterior	−1480	−985	−947	−875	**−764**	−855	−970
Level 4	Prior	−5839	−1550	−1134	−850	−234	−652	−1011
	Likelihood	−1498	−1761	−1918	−2042	−3104	−2078	−1955
	Posterior	−7337	−3311	−3052	−2892	−3338	**−2730**	−2967
Level 5	Prior	−10,610	−1962	−1321	−955	−244	−732	−1228
	Likelihood	−2858	−3378	−3584	−3816	−5790	−3917	−3703
	Posterior	−13,466	−5338	−4905	−4772	−6034	**−4649**	−4931
Level 6	Prior	−67,612	−5231	−2083	−1390	−257	−827	−1567
	Likelihood	−18,118	−24,454	−25,695	−27,123	−40,108	−27,312	−26,111
	Posterior	−85,730	−29,685	−27,779	−28,513	−40,365	−28,139	**−27,678**

Perfors, Tenenbaum, & Regier (2011): Or does it?

Why the transition to the hierarchical grammars occurs

"What kind of input is responsible for the transition from linear grammars to grammars with hierarchical phrase structure? The smallest three corpora contain very few elements generated from recursive productions (e.g., nested prepositional phrases or relative clauses) or sentences using the same kind of phrase in different positions (e.g., a prepositional phrase modifying an NP subject, an NP object, a verb, or an adjective phrase). While a regular grammar must often add an entire new subset of productions to account for these elements, a context-free grammar need add fewer (especially CFG-S). As a consequence, the flat and regular grammars have poorer generalization ability and must add proportionally more productions in order to parse a novel sentence."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Why the larger hierarchical grammar is preferred at the very last level

"The larger context-free grammar CFG-L outperforms CFG-S on the full corpus, probably because it includes non-recursive counterparts to some of its recursive productions. This results in a significantly higher likelihood since less of the probability mass is invested in recursive productions that are used much less frequently than the non-recursive ones. Thus, although both grammars have similar expressive power, the CFG-L is favored on larger corpora because the likelihood advantage overwhelms the disadvantage in the prior."

Perfors, Tenenbaum, & Regier (2011): Or does it?

Results using automatically generated grammars, rather than hand-crafted ones

Log probability = smaller negative number means more probable

Same results: hierarchical grammars preferred when more complex data is in the input.

Corpus	Probability	FLAT	REC-N	REG-M	REG-B	1-ST	CFG-S	CFG-L
Level 1	Prior	−99	−99	−99	−99	−94	−133	−148
	Likelihood	−17	−19	−20	−19	−36	−26	−25
	Posterior	**−116**	−118	−119	−118	−130	−159	−173
Level 2	Prior	−030	−385	−423	−384	−201	−355	−404
	Likelihood	−134	−151	−158	−155	−275	−189	−188
	Posterior	−764	−536	−581	−539	**−476**	−544	−592
Level 3	Prior	−1158	−653	−569	−529	−211	−433	−521
	Likelihood	−282	−320	−339	−346	−553	−402	−380
	Posterior	−1480	−973	−909	−875	**−764**	−835	−901
Level 4	Prior	−5839	−1514	−1099	−837	−234	−566	−798
	Likelihood	−1808	−1770	−1868	−2008	−3104	−2088	−1093
	Posterior	−7337	−3284	−2967	−2845	−3338	**−2654**	−2789
Level 5	Prior	−10,610	−1771	−1279	−956	−244	−615	−817
	Likelihood	−2856	−3514	−3618	−3816	−5790	−3931	−3781
	Posterior	−13,466	−5285	−4897	−4772	−6034	**−4546**	−4598
Level 6	Prior	−67,612	−5169	−2283	−1943	−257	−876	−1111
	Likelihood	−18,118	−24,299	−25,303	−25,388	−40,108	−27,032	−25,889
	Posterior	−85,730	−29,468	−27,586	−27,311	−40,365	−27,908	**−27,000**

Perfors, Tenenbaum, & Regier (2011): Or does it?

Results using data tokens, rather than data types

Different results: Linear grammars always preferred, no matter how complex the data.

Why?

"The corpus of sentence tokens contains almost ten times as much data, but no concomitant increase in the variety of sentences (as would occur if there were simply more types, corresponding to a larger dataset of tokens). Thus the likelihood is weighted relatively more strongly relative to the prior (which does not change); this works against the context-free grammars, which overgeneralize more."

Implications:
Children need a bias to evaluate grammars based on data types, rather than data tokens. (Innate bias, but domain-specific or domain-general?)

Perfors, Tenenbaum, & Regier (2011): Or does it?

Results using data types, but data based by age input (rather than level of complexity)

Hierarchical grammars preferred at all ages (even earliest ages have sufficient complexity in the input).

Corpus	Probability	FLAT	REC-N	REC-M	REC-B	1-ST	CFC-S	CFC-L
Epoch 0 (2:1)	Prior	−3998	−1915	−1349	−1166	−244	−698	−864
	Likelihood	−881	−1255	−1321	−1322	−2199	−1489	−1448
	Posterior	−4849	−3180	−2670	−2488	−2433	**−2187**	−2312
Epoch 1 (2:1–2:6)	Prior	−22,832	−3791	−1974	−1728	−257	−838	−1055
	Likelihood	−5945	−7811	−8223	−8104	−13,123	−8834	−8467
	Posterior	−28,777	−11,602	−10,197	−9832	−13,380	−9672	**−9522**
Epoch 2 (2:1–3:1)	Prior	−14,908	−4193	−2182	−1836	−257	−895	−1099
	Likelihood	−9250	−12,184	−12,815	−12,724	−20,334	−13,675	−13,099
	Posterior	−44,158	−16,357	−14,977	−14,560	−20,591	−14,540	**−14,195**
Epoch 3 (2:1–3:5)	Prior	−48,459	−4621	−2202	−1832	−257	−876	−1111
	Likelihood	−12,909	−17,153	−17,975	−17,918	−28,487	−19,232	−18,417
	Posterior	−61,368	−21,774	−20,177	−19,780	−28,744	−20,108	**−19,528**
Epoch 4 (2:1–4:2)	Prior	−59,525	−4881	−2242	−1903	−257	−876	−1111
	Likelihood	−15,945	−21,317	−22,273	−22,293	−35,284	−23,830	−22,793
	Posterior	−75,570	−26,198	−24,515	−24,196	−35,541	−24,706	**−23,904**
Epoch 5 (2:1–5:2)	Prior	−67,612	−5169	−2283	−1943	−257	−876	−1111
	Likelihood	−18,118	−24,299	−25,303	−25,388	−40,108	−27,032	−25,889
	Posterior	−85,730	−29,468	−27,586	−27,311	−40,365	−27,908	**−27,000**

Perfors, Tenenbaum, & Regier (2011): Or does it?

Accounting for data

Note that the hierarchical grammars can account for much of the most complex data, even if they're only trying to account for less complex data.

Also, the CFG-L grammar trying to account for Adam data can account for between 87 and 94% of sentences from a completely different data set (Sarah).

Proportion of sentences in the full corpus that are parsed by smaller grammars. The Level 1 grammar is the smallest grammar of that type that can parse the Level 1 corpus. All Level 6 grammars can parse the full (Level 6) corpus.

Grammar	FLAT (%)	REG-N (%)	REG-M (%)	REG-B (%)	1-ST (%)	CFG-S (%)	CFG-L (%)
X types							
Level 1	0.3	0.7	0.7	0.7	100	2.4	2.4
Level 2	1.4	3.7	5.1	5.5	100	31.5	16.4
Level 3	2.6	9.1	9.1	32.7	100	53.1	46.8
Level 4	10.9	59.7	61.2	75.2	100	87.6	82.7
Level 5	18.7	68.8	80.3	88.0	100	91.8	88.7
X tokens							
Level 1	5.9	32.6	32.6	32.6	100	40.2	40.2
Level 2	21.4	58.8	61.7	60.7	100	76.4	66.7
Level 3	25.4	72.5	70.9	79.6	100	87.8	85.8
Level 4	34.2	92.5	94.3	95.4	100	98.3	97.5
Level 5	36.9	95.9	97.6	98.5	100	99.0	98.6

Perfors, Tenenbaum, & Regier (2011): Or does it?

Making the right generalizations

The hierarchical grammars are the only grammars that generalize correctly - they have rules to parse the grammatical utterances (even ones not in the input) and no rules able to parse the ungrammatical utterances.

Ability of each grammar to parse specific sentences. The complex declarative sentence "Eagles that are alive can fly" occurs in the Adam corpus. Only the context-free grammars can parse the corresponding complex interrogative sentence.

Type	In input?	Example	Can parse?						
			FLAT	REG-N	REG-M	REG-B	1-ST	CFG-S	CFG-L
Decl Simple	Y	Eagles can fly. (n aux vi)	Y	Y	Y	Y	Y	Y	Y
Int Simple	Y	Can eagles fly? (aux n vi)	Y	Y	Y	Y	Y	Y	Y
Decl Complex	Y	Eagles that are alive can fly. (n comp aux adj aux vi)	Y	Y	Y	Y	Y	Y	Y
Int Complex	N	Can eagles that are alive fly? (aux n comp aux adj vi)	N	N	N	N	Y	Y	Y
Int Complex	N	*Are eagles that alive can fly? (aux n comp adj aux vi)	N	N	N	N	Y	N	N

Perfors, Tenenbaum, & Regier (2011): Or does it?

Generating the right representations

The hierarchical grammars (especially CFG-L) generate the most accurate structural representations for a novel data set (though the regular grammars aren't far behind).

Hand-parsed

Grammar	Precision	Recall	F-score
CFG-L	89.6	90.4	90.0
CFG-S	88.3	89.1	88.7
REG-B	85.1	85.6	85.3
REG-M	85.1	85.6	85.3
REG-N	84.1	84.5	84.3
RB	86.9	87.3	87.1
LB	33.1	33.6	33.4

Perfors, Tenenbaum, & Regier (2011): Or does it?

Implications about what children need

"In general, one must assume either a powerful domain-general learning mechanism with only a few general innate biases that guide the search, or a weaker learning mechanism with stronger innate biases, or some compromise position. Our results do not suggest that any of these possibilities is more likely than the others. Our core argument concerns only the specific need for a bias to *a priori* prefer analyses of syntax that incorporate hierarchical phrase structure. We are arguing that a rational learner may not require such a bias, not that other biases are also unnecessary."

Berwick, Pietroski, Yankama, & Chomsky (2011) Response

Berwick et al. say this doesn't address the PoS problem

Basic argument: Having a hierarchical analysis is the first step to being able to posit structure-dependent rules for transforming one utterance to another (i.e., declarative to interrogative) - but it doesn't mean that you *do* posit those rules rather than structure-independent rules. You have to still know to use that structure when hypothesizing your rules.

"But even if a Bayesian learner can acquire grammars that generate structured expressions…crucially, however, it does not follow that such learners will acquire grammars in which rules are structure dependent. On the contrary…the acquired grammars may still operate structure-independently… PTR seem to assume that if a grammar generates expressions that exhibit hierarchy, then the rules defined over these expressions / structures must be structure dependent…Structured expressions can be (trans)formed by a structure-independent rule…for example, fronting the first auxiliary."

Berwick, Pietroski, Yankama, & Chomsky (2011) Response

One way to think about that critique: A different kind of rule

In particular, learning a grammar that accounts for the strings of the language doesn't cover the behavior that needs accounting for.

Ability of each grammar to parse specific sentences. The complex declarative sentence "Eagles that are alive can fly" occurs in the Adam corpus. Only the context-free grammars can parse the corresponding complex interrogative sentence.

Type	In input?	Example	Can parse?						
			FLAT	REG-N	REG-M	REG-B	1-ST	CFG-S	CFG-L
Decl Simple	Y	Eagles can fly. (n aux vi)	Y	Y	Y	Y	Y	Y	Y
Int Simple	Y	Can eagles fly? (aux n vi)	Y	Y	Y	Y	Y	Y	Y
Decl Complex	Y	Eagles that are alive can fly. (n comp aux adj aux vi)	Y	Y	Y	Y	Y	Y	Y
Int Complex	N	Can eagles that are alive fly? (aux n comp aux adj vi)	N	N	N	N	Y	Y	Y
Int Complex	N	*Are eagles that alive can fly? (aux n comp adj aux vi)	N	N	N	N	Y	N	N

"not good enough"

This is because the "correct" rules involve transforming the declarative into the yes-no question. The CFG grammars just generate the yes-no question directly, without realizing its relation to the declarative (which is that these utterances share a core component of meaning). To this end, the CFG grammars here are not embedded in a model of generating language from underlying thought/meaning.

Perfors, Tenenbaum, & Regier (2011): Or does it?

About the necessity of the type-based analysis

"Our work suggests that if human learners, like our model, are capable of evaluating whether type-based or token-based analyses are *themselves* more appropriate for a given problem, they might rationally decide to favor a more type-based analysis when deciding among grammars (not necessarily for other aspects of language acquisition)…Would a disposition to evaluate grammars within a two-component adaptor-grammar-like framework, or based on type data only, constitute a language-specific or domain-general disposition? It is difficult to say, but the conceptual underpinnings of the adaptor grammar framework are consistent with a domain-general interpretation, emerging due to memory constraints or other cognitive factors."

Perfors, Tenenbaum, & Regier (2011): Or does it?

The importance of learning a system rather than learning a construction

"Our analysis makes a general point that has sometimes been overlooked in considering stimulus poverty arguments, namely that children learn grammatical rules as a part of a system of knowledge…. We have suggested here that even when the data does not appear to explain an isolated inference, there may be enough evidence to learn a larger system of inguistic knowledge – a whole grammar – of which the isolated inference is a part. A similar intuition underlies other arguments about the important role that indirect evidence might play in language acquisition…This point is also broadly consistent with the generative tradition in linguistics…one of whose original goals was to unify apparently disparate aspects of syntax…"

Perfors, Tenenbaum, & Regier (2011): Or does it?

Learning higher-order generalizations first

"One implication of our work is that it may be possible to learn a higher-order abstraction T even before identifying all of the correct lower-level generalizations G that T supports. Therefore, it may be possible for T to operate to constrain G even if T itself is learned… If an abstract generalization can be acquired very early and can function as a constraint on later development of specific rules of grammar, it may function effectively as if it were an innate domain-specific constraint, even if it is in fact not innate and instead is acquired by domain-general induction from data."

Perfors, Tenenbaum, & Regier (2011): Or does it?

How this happens

"While there are infinitely many possible specific grammars G, there are only a small number of possible grammar types T. It may thus require less evidence to identify the correct T than to identify the correct G. More deeply, because the higher level of T affects the grammar of the language as a whole while any component of G affects only a small subset of the language produced, there is in a sense much more data available about T than there is about any particular component of G…Higher-order generalizations may thus be learned faster simply because there is much more evidence relevant to them."