Statistical Learning of Complex Questions

Hartmut Fitz (h.fitz@rug.nl)
Center for Language and Cognition Groningen, Oude Kijk in't Jatstraat 26
9712EK Groningen, the Netherlands

Abstract

The problem of auxiliary fronting in complex polar questions occupies a prominent position within the nature versus nurture controversy in language acquisition. We employ a model of statistical learning which uses sequential and semantic information to produce utterances from a bag of words. This linear learner is capable of generating grammatical questions without exposure to these structures in its training environment. We also demonstrate that the model performs superior to *n*-gram learners on this task. Implications for nativist theories of language acquisition are discussed.

Keywords: Language acquisition; complex syntax; poverty of the stimulus; statistical learning; distributional information.

Introduction

It is a central question in language acquisition which aspects of our knowledge of language are learned from experience and which are part of our biological endowment for language. Nativist arguments often identify some property of a language and argue that it is not learnable from typical child-directed speech. By abductive reasoning, innate language-specific knowledge is offered as the best explanation of why children come to know this property regardless. The problem of auxiliary fronting in so-called complex polar questions (CPQ hereafter) is a key issue in this nature versus nurture debate.

According to Chomsky (1980), English yes/no-questions are formed from declaratives by displacing an auxiliary. The sentence "The man is happy" transforms into a question by subject-auxiliary inversion: "Is the man happy?". Declaratives with a relative clause can contain two identical auxiliaries as in "The man that is hungry is happy". Chomsky asked how children could learn that the main clause auxiliary should be placed in front, rather than the auxiliary which comes first. Only the former rule yields a grammatical CPQ.

(1) a. Is the man that is hungry happy?
    b. *Is the man that hungry is happy?

He claimed that children have no basis in experience to adopt the correct rule since examples such as (1-a) do not occur in child-directed speech. In addition, children should adopt the rule which generates (1-b) because (i) it is supported by experience of simple yes/no-questions and (ii) the correct rule is "far more complex" in that it requires sensitivity to the hierarchical structure of a sentence. But children rarely, if ever, make mistakes as in (1-b) (Crain & Nakayama, 1987; Ambridge, Rowland, & Pine, 2008). They do not seem to generalize in a structure-independent way. To explain this error-free behavior, Chomsky postulated innate structure-dependent constraints on learning.

The above formulation of this poverty-of-the-stimulus argument makes a number of controversial assumptions. There is accreting evidence, for instance, that learning the syntax of questions does not involve learning movement rules (Dabrowska & Lieven, 2005; Estigarribia, 2009). An inadequate description of the learning target might obscure empiricist solutions to the problem. Secondly, auxiliary fronting has been isolated from all the rest of language. Although there is consensus that structures (1-a) are highly infrequent, the input environment of a child might provide other sources of indirect evidence for the correct rule (Pullum & Scholz, 2002). Another critical assumption is that the structure-independent rule (1-b) is simpler and should be preferred in the absence of innate constraints. If there is no reason to believe that children should overgeneralize there is no explanatory necessity for such constraints. The nativist argument would be preempted.[1]

Despite these reservations, it is clear that any theory of language acquisition which places more emphasis on the role of experience needs to explain how the syntax of complex questions can be acquired. Ideally, such an explanation demonstrates that a concrete, implemented learning mechanism built on justifiable assumptions can acquire auxiliary fronting from plausible input distributions.

Linear versus hierarchical models

Recently, several models of language learning have been proposed which explicitly address the issue of auxiliary fronting. These models can roughly be divided into linear and hierarchical approaches. Linear models do not explicitly represent the hierarchical structure of a sentence's organization into phrases and clauses. All models briefly discussed here share the assumption that CPQ do not occur in child-directed speech, they learn solely from indirect evidence.

In the framework of data-oriented parsing, Bod (2009) showed that derivations of parse trees for grammatical CPQ are shorter (or more probable) than those for ungrammatical CPQ given the model's language input. Since there was a primitive subtree substitution operation built into the learning mechanism, it seems to beg the question of whether structure-dependent processing can be learned or should be considered innate. Perfors, Tenenbaum, and Regier (2006) demonstrated that an ideal Bayesian learner favors a hierarchical over a linear grammar to fit a training corpus. This grammar could parse grammatical CPQ while the linear grammar could not. The model, however, did not strictly learn grammars from data, but rather selected one from a given set. How grammar selection bears on the process of child-language acquisition needs to be elucidated. They argued that linear mod-

[1] More detailed discussion of the assumptions behind this nativist argument can be found in Fitz (2009).

els have little to contribute to the auxiliary fronting debate because structure-dependent processing requires hierarchical representations. This assumption has been challenged by a number of linear approaches. If a linear model, learning auxiliary fronting from raw data, behaves in a manner consistent with structure-dependent processing, this would suggest that explicit representations of hierarchical structure might be superfluous. Clark and Eyraud (2006) proposed a linear alignment learner which substituted relativized NPs for simple NPs if they occurred in identical contexts in the corpus. As a result, the learner could generate grammatical CPQ. The model can be criticized on similar grounds as the model by Bod. A simple recurrent network was used by Lewis and Elman (2001) to learn CPQ from an artificial language with some success but their results are anecdotal at best. The most widely received linear approach used *n*-gram learners on untagged corpora of child-directed speech (Reali & Christiansen, 2005). The authors showed that a Bigram model could reliably classify pairs of grammatical and ungrammatical CPQ by assigning higher sentence probability to the former on 96% of the tested items. They suggested that indirect statistical information extracted from strings of words might be sufficient for children to infer the correct rule of auxiliary fronting. These results were scrutinized by Kam, Stoyneshka, Tornyova, Fodor, and Sakas (2008). They argued that the success of the Bigram model was largely due to a single distinguishing bigram which was supported by accidental phonological facts about English. When they added structural and lexical diversity to the test items, the model failed. Moreover, they argued that the bigram approach might not be valid cross-linguistically.

The Adjacency-Prominence learner

In our own work we aimed at showing that these difficulties could be overcome by a linear statistical model which in addition to *n*-gram based sequence learning (adjacency) uses meaning to constrain sentence production (prominence). The statistical information on which this learner draws has two components. The adjacency statistics was collected over bigrams in the training corpus. It measured how often two words which co-occurred in sentences, occurred adjacent to each other. The key addition over *n*-gram models was the prominence statistics. The learner tracked which words frequently preceded other words in the input environment. Words which on average were found earlier in a sentence than other words are considered more prominent. Using this statistics, a hierarchy was created which ordered words in an utterance in terms of their prominence. More prominent words then tended to be sequenced earlier in production. While the adjacency statistics selected words based on the previous word in an utterance, the prominence statistics selected words based on their prominence relation with remaining words in an utterance. This process is illustrated schematically in Figure 1. Both statistics were combined into the Adjacency-Prominence learner (AP-learner for short). This model of syntax learning was introduced in Chang, Lieven, and Tomasello (2008) where it was tested on a variety of typologically-distinct languages. Formal definitions of the

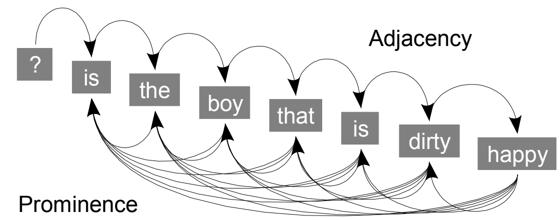

Figure 1: Adacency-prominence statistics for the CPQ *Is the boy that is dirty happy?* (adapted from Chang et al. (2008)).

two kinds of statistics are given in Table 1. Note that the adjacency statistics differs from forward transitional probabilities because bigram counts are normalized by the frequency of word pairs instead of the first unigram. Note also that the prominence statistics of a word is a sum over its relation with other words. To give a comparable weight to the adjacency statistics, it was multiplied by the number of remaining words in an utterance.

Table 1: AP-learner statistics.

$C(w_{n-1}, w_n)$	Frequency of bigram $w_{n-1}w_n$
$\text{Pair}(w_a, w_b)$	Frequency of words w_a, w_b occurring together in the same sentence in any order
$P(w_a, w_b)$	Frequency of word w_a occurring before w_b in a sentence at any distance
Length	Number of words in bag-of-words
η	Learning rate[2]
Adjacency	$\text{Adj}(w_n) = C(w_{n-1}, w_n)/\text{Pair}(w_{n-1}, w_n)$
Prominence	$\text{Pro}(w_n) = \sum_{w_b} P(w_n, w_b)/\text{Pair}(w_n, w_b)$ where w_b are all words in the bag (except w_n)
Adjacency-Prominence	$\text{AP}(w_n) = \text{Length} \times \text{Adj}(w_n) + \text{Pro}(w_n) \times \eta$

Evaluation

The performance of the AP-learner was evaluated in a sentence generation task. We assumed that speakers aim to produce utterances which express the meaning they intend to convey. To approximate constraints that meaning places on sentence production, a target utterance was split into an unordered bag-of-words. The learner then had to use its syntactic knowledge, extracted from the training corpus, to order this bag-of-words. Sentences were produced incrementally one word at a time. At each word position, all words in the bag were competing for the next slot in the utterance. The

[2]The learning rate was used to balance the contribution of both statistics to word choice. It was held fixed across experiments.

learner could use forward probabilities from the preceding word (adjacency) but also the prominence ordering over the words in the bag to predict the next word. The prominence value for a given word could dynamically change as the set of word options diminished during production.

Training and test items were identified as questions or declaratives by prepending a marker `quest`/`decl` to each sentence. Utterance generation was initialized by creating a bag-of-words including the marker for the target sentence. For each word in the bag, the adjacency-prominence statistics was collected and the word with the highest combined value was selected (see Table 1). The word was appended to the marker and removed from the bag-of-words. This procedure continued recursively until the bag was empty. The string of words produced by the learner was compared with the target utterance and its grammatical alternatives. For instance, the bag-of-words obtained from "Is the dog that is run -ing happy?" also generated "Is the dog that is happy run -ing?". If the learner produced either form, the sentence prediction accuracy count was incremented. Likewise, if either of the ungrammatical alternatives with a displaced embedded clause auxiliary was produced, the output was counted as a structure-independent generalization error.

Reali and Christiansen (2005) tested their *n*-gram learners in a grammaticality judgement task in which CPQs with lower cross-entropy were classified as grammatical. Our learner, in contrast, had to actually produce sentences from a bag-of-words and not merely classify them. Statistical information sufficient for classification might not be suitable for production. Chang et al. (2008) argued that bag-of-word generation is an adequate task to assess and compare statistical learners across languages.

The remainder of this paper is organized as follows. First, we demonstrate that the AP-learner can learn the syntax of complex questions in the absence of positive evidence and that overgeneralization does not occur. Then we compare the AP-learner with *n*-gram models and show that it performs superior. Finally, we identify conditions under which the AP-learner does make structure-independent errors. Such conditions arguably do not obtain in child-language acquisition. We conclude with a discussion of our results.

Method

Language input

The AP-learner was trained on an artificial English-like language with transitives and intransitives as basic construction types. From these constructions, simple declaratives, simple polar questions, complex declaratives, and polar questions with relative clauses could be generated (see Table 2). The language had number and noun-verb agreement, tense (past/present) and aspect (progressive/simple). Nouns could be animate and inanimate, or substituted by pronouns. Over a lexicon of 104 words and inflectional morphemes the language generated approximately 2.8×10^9 distinct sentences. Ambridge et al. (2008) suggested that structure-independent

Table 2: Structures generated by the artificial language.

Sentence type	Example
Simple declarative	The guys buy it.
Simple polar question	Was the dog sleeping?
Complex declarative	A girl that is hitting him plays.
Complex polar question	Is a cat that is grumpy thirsty?

generalizations such as

(2) Are the boys **that running** are eating?

may not occur in development because they violate word co-occurrence patterns in English (boldface bigram). In similar vein, Kam et al. (2008) argued that the good performance of the Reali and Christiansen (2005) model was due to these relative clause initial bigrams. To ensure that our learner could, in principle, generalize erroneously, we separated plural markers and inflectional morphemes for tense and aspect from the word stem. Thus sentence (2) was represented in our artificial language as

(3) Are the boy -s **that run** -ing are eat -ing?

The boldface bigram occurred frequently in the training corpus, for example in sentences like "The boy -s that run are kick -ing the toy". This made it more difficult for our learner to retain the embedded clause auxiliary in CPQs.

Results

Experiment 1

The first experiment tested whether the AP-learner was able to produce correct CPQs when trained only on simple declaratives, simple polar questions and declaratives with relative clauses. The learner was trained on 20.000 sentences randomly generated from the artificial language. 50% of these were simple sentences, the others were complex. 50% of the simple sentences were questions, the others were declaratives. Crucially, the training corpus did not contain any instance of a CPQ or any other question with a relative clause. Thus, it was tested whether the statistical information contained in the trained structures was sufficient for the AP-learner to generalize to the syntax of the novel CPQs. If so, this would support the idea that indirect evidence from frequent structures which are attested in child-directed speech might be sufficient to learn the correct subject-auxiliary inversion rule for complex polar questions.

The test set contained 40 CPQs randomly generated by the artificial grammar. All CPQs had an intransitive main clause. 20 had a center-embedded intransitive relative-clause (II), and 20 had a transitive relative-clause. Half of the transitive embeddings were subject-relativized (ITS), the other half were object-relativized (ITO). All tested CPQs were ambiguous in that the main clause auxiliary was identical with the embedded clause auxiliary. Auxiliaries could be singular or plural, past or present tense. Three actual test questions are listed in Table 3. In contrast to the study of Reali and Christiansen (2005), the set of tested CPQs was structurally diverse (intransitive and transitive embeddings, subject- and object-relativized) and not limited to the auxiliary "is".

Table 3: Sample test questions.

Type	Example
II	Were the boy -s that were dirty play -ing ?
ITS	Was a brother that was push -ing them hungry ?
ITO	Is a cat that a boy is chase -ing jump -ing ?

When evaluating the learner's output for ITS and ITO questions, only those grammatical alternatives were considered which preserved clause type and the grammatical role of the relativized constituent. For instance, when tested on ITOs, the learner's utterance had to have an intransitive main clause, and the transitive embedding had to have an object gap in order to count as an accurate production. The results of this experiment are shown in Figure 2.[3] The mean sentence prediction accuracy was 91.25% versus 8.75% incorrect productions. On CPQs of type II, the AP-learner reached 100% accuracy. Slightly lower was the accuracy on ITS (94%) and ITO structures (71%). This difference between subject- and object-relativized transitives is consistent with developmental data on relative-clause acquisition in English-speaking children (Diessel & Tomasello, 2005). The AP-learner made mistakes on this task, it did not produce all test questions correctly. Importantly, however, none of the learner's incorrect productions matched an ungrammatical CPQ which reflected structure-independent generalization. Although the AP-learner did not experience any instance of a CPQ in training, it correctly generalized the syntax of subject-auxiliary inversion from simple polar questions and declaratives with relative clauses to the formation of complex questions. When we added either ambiguous CPQs or CPQs with mixed number, tense and aspect (or both) to the training set, the learner's performance did not improve on any of the tested question types. These results suggest that the distributional information contained in simple polar questions and complex declaratives support the learning of structure-dependent generalizations even if the learner does not explicitly represent the hierarchical organization of CPQs into clauses and phrasal units. Since both these structures—simple questions and relative clause constructions—typically occur in child-directed speech, children might be exposed to sufficient indirect evidence to induce the syntax of auxiliary fronting in the absence of positive examples.

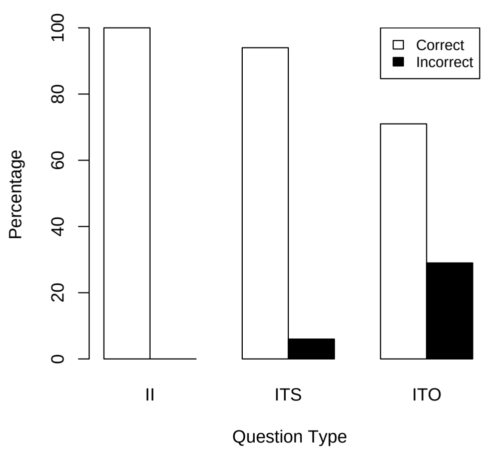

Figure 2: AP-learner tested on three kinds of CPQ.

[3]All modelling data reported here are averaged over ten randomly generated training sets to ensure that results were robust with regard to the artificial language used to create input environments.

Experiment 2

In the previous experiment, the AP-learner showed differences in production accuracy between II, ITS and ITO questions. To trace the origin of differential performance, it was helpful to compare the AP-learner with Bi- and Trigram models of statistical learning. Both these models were trained, tested and evaluated in exactly the same way as the AP-learner. Figure 3 shows the prediction accuracy of the different models by CPQ type. All models displayed the same qualitative behavior in that II questions were easier to produce than ITS, which were easier than ITO. Both *n*-gram models performed similar to the AP-learner on II questions. These CPQs were shorter than the other question types and thus had fewer choice points for prediction error. Moreover, ungrammatical II questions frequently contained word sequences which were not supported by the training corpus (e.g., "that happy"). The models followed a simple principle of non-monotonic learning to produce grammatical II questions: in the absence of evidence to the contrary, embedded clause auxiliaries should not be omitted. The Trigram model came close to the AP-learner on ITS questions (82%), whereas the Bigram model dropped below 40% accuracy. Errors made by the Bigram model

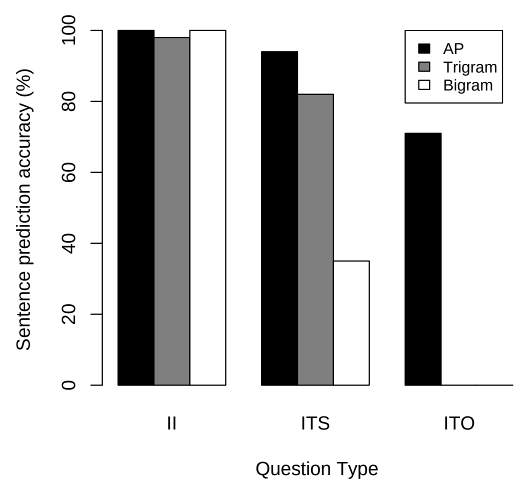

Figure 3: AP-learner in comparison with *n*-gram models.

mostly occurred sentence-initially (e.g., "quest Is chase"), whereas Trigram model errors mostly occurred in the relative clause (e.g., wrong verb type). The AP-learner was less vulnerable to these kinds of errors because it did not rely exclusively on co-occurrence frequencies. In addition to adjacency, the model could also use the prominence statistics which informed it that a subject should precede a verb form in the main clause and that a transitive verb should be produced in the relative clause (instead of an intransitive) when there was a direct object (e.g., a pronoun) left to sequence in the bag-of-words. Neither *n*-gram model produced any correct ITO question, whereas the AP-learner produced 71% correct ITOs. The Bigram model made the same errors as in ITS questions and sequenced a verb form after the initial auxiliary. The Trigram model often converted ITOs into grammatical ITS questions. The AP-learner also made such conversion errors, but less frequently. Again, the prominence statistics helped the model to place subject noun phrases before the verb form in transitive embeddings and this information was not available to the other models.

Kam et al. (2008) argued that Bigram models are not sufficient to learn the syntax of complex questions from noisy, realistic corpora. Our results support their findings for idealized input environments. The AP-learner was superior to both *n*-gram models when tested CPQs could not reliably be generated from a bag-of-words based on forward probabilities alone.

Experiment 3

As mentioned in the introduction, Chomsky's argument for the innateness of structure-dependent constraints on language learning has two prongs. Children have no basis in experience to infer the correct rule for auxiliary fronting, and they should overgeneralize by displacing the linearly-first auxiliary as witnessed in simple polar questions in their language input. In Experiment 1, we found no evidence for either claim. The AP-learner could produce more than 90% grammatical CPQs without having experienced such structures in training. Although the model made some mistakes, it never produced ungrammatical CPQs in which the embedded clause auxiliary was omitted. In a third experiment we tried to elicit overgeneralizations by creating input conditions which mislead the AP-learner into producing structure-independent errors. To do this, we distinguished multiple word tokens with markers in forward order of their occurrence within one sentence. Question (3), for instance, was now represented as

(4) are1 the1 boy1 -s1 that1 run1 -ing1 are2 eat1 -ing2 ?

After the model had produced a CPQ from a marked bag-of-words, the markers were removed and the output was compared with the equally unmarked target questions (grammatical and ungrammatical versions).

Distinguishing constituents in this way created clause-specific similarities between auxiliaries in different structures. The auxiliary are1 in test item (4) resembled the auxiliary in simple polar questions and the embedded clause auxiliary in complex declaratives from the training set. The auxiliary are2 resembled the main clause auxiliary in complex declaratives. These similarities were picked up by the adjacency-prominence statistics, as shown in Figure 4. Now the AP-

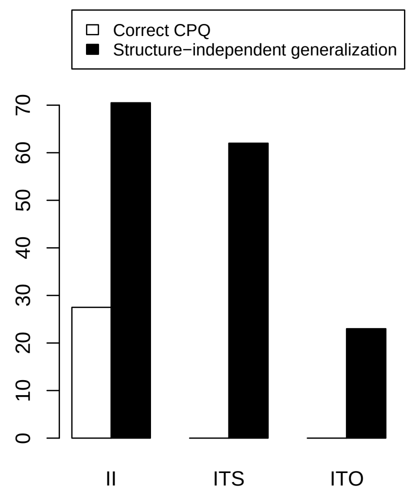

Figure 4: Structure-independent errors occurred when multiple auxiliaries were distinguished in the corpus.

learner produced only 13.75% correct CPQs. Out of the total incorrect CPQs, 65.5% were structure-independent errors in which the question-initial auxiliary was omitted from the relative clause rather than the main clause. Hence, the AP-learner could be forced to generalize erroneously when constituents were forward marked. Children, however, learn the syntax of questions from input which is not marked in this way. It is therefore not self-evident, as Chomsky suggests, that children should adopt the wrong auxiliary fronting hypothesis in the absence of innate constraints. In order to substantiate this claim, one would have to argue that children perceptually distinguish and track multiple auxiliary tokens in a way similar to our AP-learner in the above experiment. Unless this can be done convincingly, there is no reason to believe that children should overgeneralize. As a consequence, it is no longer puzzling that they in fact rarely do (Crain & Nakayama, 1987; Ambridge et al., 2008). Moreover, we do not need to posit innate constraints on learning as the best explanation of why they do not. One crucial premiss of the poverty-of-the-stimulus argument breaks away. Experiments 1 & 3, we believe, jointly shift the burden of proof back to those who claim that a biological endowment for structure-dependent processing is necessary to block overgeneralization.

Discussion and conclusions

Using a statistical model of syntactic development adapted from Chang et al. (2008), we demonstrated that the syntax of complex polar questions was learnable to a high degree of accuracy even when these structures were not present in the language input to the model. The tested questions were more diverse, both lexically (auxiliaries) and structurally (relative clause types), than the items used in Reali and Christiansen (2005) which may answer to some of the criticism posed by Kam et al. (2008). Our learner, however, was collecting more than *n*-gram statistics to accomplish this task. In addition to adjacency, it used a prominence ordering over words that were left to sequence. Words which were more prominent in sentences of the learner's experience were more accessible for production. Thus, the AP-learner was not relying on the presence or absence of particular bigrams to produce grammatical questions and it outperformed several *n*-gram models. Importantly, it was also shown that errors the learner made did not reflect structure-independent generalizations. To elicit these errors, the learning environment had to be manipulated such that it no longer resembled natural language input to children. This casts some doubt on the claim that children should overgeneralize in the absence of innate constraints.

On the other hand, our learner was trained on an artificial English-like language which did not exhibit the noisiness, diversity and distributional properties of child-directed speech. Our results should therefore be interpreted as a proof-of-concept that under idealized conditions a statistical learner which draws on sequential and semantic information can learn the syntax of complex polar questions from simpler and similar structures in the input. It remains to be tested whether this approach scales to real corpora and in particular whether it works for different languages which permit complex polar questions other than auxiliary-initial ones (Kam et al., 2008).

We do not suggest here that the AP-statistics is all that is needed to learn the syntax of complex polar questions. For one thing, the learner made mistakes where adults do not. The inclusion of meaning constraints (bag-of-words and prominence hierarchy) into a statistical learning model was not sufficient to guarantee error-free learning or rule out the production of grammatical alternatives. Tighter semantic constraints and additional sources of information might be necessary.

Compared with other models which have previously been proposed to show the data-driven learnability of auxiliary fronting, the AP-learner did not make assumptions about the nature of syntactic representations in children, or the operations performed on such representations. Our model learned from untagged raw text by means of simple, domain-general mechanisms and did not incorporate language-specific knowledge or biases. The model's task to produce rather than classify sentences is closer to experimental paradigms in developmental psychology than grammaticality judgement. Incremental word prediction is consistent with current theories of language processing (Pickering & Garrod, 2007). Furthermore, the evaluation procedure did not depend on language-specific assumptions about syntactic categories or on sentence probabilities which are difficult to interpret. Even though the AP-learner did not explicitly represent the hierarchical structure of complex questions or syntactic rules operating on such representations, it performed as if it respected the structure-dependence of auxiliary fronting. Thus, surface distributional information might be sufficient for a statistical learner to resolve the Chomskyan challenge.

Acknowledgments

Thanks to Franklin Chang for helpful discussions.

References

Ambridge, B., Rowland, C. E., & Pine, J. M. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science*, *32*, 222–255.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*, *33*, 752–793.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research*, *9*, 198–213.

Chomsky, N. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky* (M. Piattelli-Palmarini, Ed.). Cambridge, MA: Harvard University Press.

Clark, A., & Eyraud, R. (2006). Learning auxiliary fronting with grammatical inference. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* Vancouver, BC.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, *63*(3), 522–543.

Dabrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*, 437–474.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, *81*(4), 882–906.

Estigarribia, B. (2009). Facilitation by variation: Right-to-left learning of English yes/no questions. *Cognitive Science*, *34*, 68–93.

Fitz, H. (2009). *Neural syntax*. ILLC dissertation series, University of Amsterdam.

Kam, X., Stoyneshka, I., Tornyova, L., Fodor, J., & Sakas, W. (2008). Bigrams and the richness of the stimulus. *Cognitive Science*, *32*, 771–787.

Lewis, J., & Elman, J. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the 26th Annual Boston University Conference on Language Development.* Somerville, MA.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the stimulus? A rational approach. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* Vancouver, BC.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*, 105–110.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*, 9–50.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*, 1007–1028.