# AUXILIARY FRONTING TRANSFORMATIONS AND RULE SIMPLICITY: AN INTERIM REPORT

Tomohiro Fujii and Sota Ninomiya
Yokohama National University

## 1. Introduction

In discussing "properties of language that can reasonably be supposed not to have been learned," Chomsky (1975: 30–31) formulates a structure-independent hypothesis for question formation as follows.

> *Hypothesis 1*: The child processes the declarative sentence from its first word (i.e., from "left to right"), continuing until he reaches the first occurrence of the word "is" (or others like it: "may," "will," etc.); he then preposes this occurrence of "is," producing the corresponding question (with some concomitant modifications of form that need not concern us).

This hypothesis is contrasted with "Hypothesis 2," its structure-dependent counterpart (Chomsky, *op. cit*.).

> *Hypothesis 2*: the child analyzes the declarative sentence into abstract phrases; he then locates the first occurrence of "is" (etc.) that follows the first noun phrase; he then preposes this occurrence of "is," forming the corresponding question.

The structure-dependent hypothesis, needless to say, explains multi-clause interrogative sentences such as (1), on top of simplex ones such as (2), whereas the structure-independent one fails to. The renowned 'Poverty of the Stimulus' argument maintains that the learner acquiring Hypothesis 2 would nevertheless have no evidence against Hypothesis 1 in the input and hence Hypothesis 1 need be excluded from the learner's hypothesis space.[1]

(1) Can the boy who is over there swim?
    cf. The boy who is over there can swim.

(2) Can the boy swim?
    cf. The boy can swim.

Chomsky thus wrote, "[Hypothesis 1] works quite well. It is also extremely simple."

---

[1] Pullum and Scholz (2002) point out that examples with the second auxiliary fronted can be found in child-directed speech. According to Legate and Yang's (2002) often cited study of two CHIDES corpora, 'second auxiliary inverted examples make up 0.068% and 0.045 % of the matrix yes/no and *wh*-questions in the corpora. See Section 5.

The debate as to whether English question formation makes a case for the Poverty of the Stimulus argument has still drawn much attention (see Pearl 2021, which contains a review of the debate to date). The limited scope of this paper is to make the setting of Chomsky's argument more explicit and concrete by proposing a simplicity metric that can measure how *simple* these hypotheses are. The premise that Hypothesis 1 is simpler than its alternative is not trivial for the validity of the argument. If Hypothesis 1 were more complex than Hypothesis 2, the in-advance exclusion of structure-independent hypotheses might be unnecessary. An evaluation metric, which selects a simpler hypothesis over its alternative when both explains the corpus, would do. In what follows, a probabilistic model is proposed that allows us to evaluate the simplicity of transformational rules including those implementing Hypotheses 1 and 2. Inspired by Perfors, Tanenbaum and Regier (2006, 2011), we define a Bayesian prior to characterize the parsimony of an adjunction rule. Perfors et al. propose that the complexity of a rewriting grammar can be calculated by looking at the number of vocabulary items (i.e., terminals and non-terminals) $V$, the number of nonterminals $n$, the number of rewriting or production rules $P$, and the number of rule $i$'s right-hand-side symbols $N_i$, and so on. For example, compare mini-grammars generating nominals, G1 in (3) and G2 in (4). (In the framework of Perfors et al., terminal symbols are what are standardly considered preterminal symbols. Grammars generate strings of symbols, not actual words.)

(3) *Grammar G1*

NP → det n | det adj n

(4) *Grammar G2*

NP → det N′
N′ → adj N′ | n

The prior for each grammar is calculated by the following equation.

(5)

$$p(G) = p(P)\ \ p(n) \prod_{i=1}^{P} p(N_i)\ \ \prod_{j=1}^{N_i} \frac{1}{V}$$

*p(P)* is the probability of the grammar having *P* production rules; *p(n)* is the probability of it having *n* nonterminals; $p(N_i)$ is the probability of rule *i* having $N_i$ right-hand siblings; and *1/V* is the probability of the selected category symbol, say, *adj* being selected as a right-hand sibling *j* from the vocabulary items.[2] As shown in Table 1, Grammar G1 scores higher than Grammar G2 does, meaning that the former is simpler than the latter under the given data.

In addition to simplicity, likelihood (i.e., how likely a grammar generates the set of trees that

[2] For the ease of exposition, we ignore discussion of how to assign wrights to productions in priors. The weights for the first rule and second rule of Grammar G1, represented as $\langle\theta_1, \theta_2\rangle$, for instance, can be ⟨0.3, 0.7⟩ or ⟨0.56, 0.44⟩ or something else. See Perfors et al. (2011) for how to treat weighted productions for priors.

it assigns to the strings of the corpus) is also a criterion for grammar choice in the Perfors et al. framework. Suppose that the corpus consists of *det n* and *det adj n*. If the rules rewriting the same nonterminal are equally weighted, each NP rule in Grammar G1 is assigned 0.5. In the other grammar, the NP rule is assigned 1 while each N' rule 0.5. Then, the two grammars' likelihoods turn out to be as shown in Table 1. Again, the flat grammar scores higher than the hierarchical grammar under the given corpus. These likelihood and the prior probabilities taken together return posterior probabilities. As far as the given corpus goes, the model selects Grammar G1.

Table 1. Natural logarithms of the prior, likelihood and posterior probabilities for Grammars G1 and G2.

| | Prior | Likelihood | Posterior |
|---|---|---|---|
| Grammar G1 | -12.477 | -1.386 | -13.863 |
| Grammar G2 | -14.979 | -2.079 | -17.058 |

If, however, more new constructions are added to the corpus, a grammar from the flat hypothesis gets more complex quickly whereas its CFG counterpart does not. Suppose for example that nominal *det adj adj n* are added to the corpus. G1 would need a new rule, "NP → det adj adj n," but G2 does not. This way, hierarchical grammars are eventually selected over their flat counterparts without direct evidence. This is one of the major results of Perfors et al. (2006, 2011).

The present work discusses transformational rules rather than rewriting grammars. We focus on the issue of how the priors of transformational rules, especially adjunction rules, can be calculated in order to determine a structure-independent rule is simpler than its structure-dependent counterpart. The task of measuring how well a transformation combined with a CFG fits a training corpus (i.e., the likelihood probability) is not undertaken here, being left for future research.

## 2. Transformational Rules

This section aims at formulating structure-dependent and structure-independent Aux-to-C movement transformations. The first to discuss is the standard structure-dependent transformation that empirically works. Consider (6), which is formulated as a 4-term, obligatory and root transformation.[3] (Again, terminals are presented in lower case.)

[3] The rule is assumed to be optional without involving the question complementizer in, e.g., Chomsky (1957). We follow Baker (1978: 119) assuming that the rule's SA requires the involvement of a question complementizer and that it is obligatory. Note that we assume there are two complementizers in the vocabulary: $c_{+Q}$ and $c_{-Q}$. This means that we treat either of them a simplex symbol, which is not the case in a recent understanding. This non-standard treatment of features is due to the fact that CFGs only handle terminals and nonterminals.

(6) *Front the matrix Aux:* Structure-dependent; Root; Obligatory

| | | | | | |
|---|---|---|---|---|---|
| SA: | $c_{+Q}$, | NP, | aux, | | |
| SC: | 1 | 2 | 3 | 4 | => |
| | 3*1 | 2 | | 4 | |

A transformation comprises a Structural Analysis (SA) (or Structural Description) and a Structural Change (SC). A rule's SD determines what trees the rule may applies to while its SC tells what happens to input trees as a result of application of the rule. The rule given in (6) is understood as follows: This rule applies to a tree that can be analyzed as a terminal that is $c_{+Q}$ (Term 1) followed by a piece that is dominated by NP (Term 2), followed by a terminal that is aux (Term 3), followed by anything (Term 4). The SC then dictates that no change takes place in Term 2 and Term 4 and that the third element *aux* Chomsky-adjoins to the right of the first element $c_{+Q}$. Chomsky-adjunction, indicated by * in the SC, is simply what is known as adjunction in the Principles and Parameters framework.[4] Finally, the third-term terminal in the SA gets deleted.

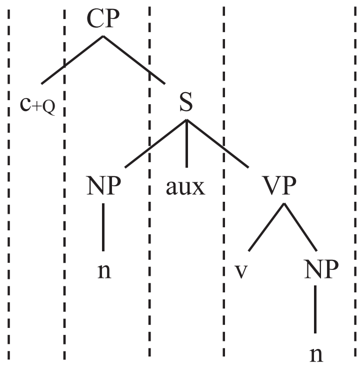
Note that no symbols are present in Term 4 in the SA. By this, we mean any string, including a null string, can satisfy the specification of this term. Terms like this are often conditioned using *variables* like "X" or "Y" in the literature. The reason for choosing the non-standard notation will be clear later in the paper. Note also that further information can be added. Such information includes information on rule type such as "Structure-(in)dependent," "Obligatory," and "Root," as seen in (6). The first is already introduced above. The obligatory-optional distinction is a familiar notion since Chomsky (1957) and will be taken up shortly. As for the third distinction, we can think of it as requiring that the landing site of the moving Aux, C, be a matrix element (Emonds 1976). Though these features do not necessarily play a crucial role in comparing structure-dependent and independent hypotheses, they surely help to increase the grammars' descriptive adequacy.

To see more details of how a P-marker is mapped onto a new P-marker, observe that the tree for the sentence in (7) can be "factorized" in the way indicated by dotted lines in (8). This factorization allows the tree undergo the structure-dependent transformation in (6), yielding (9).

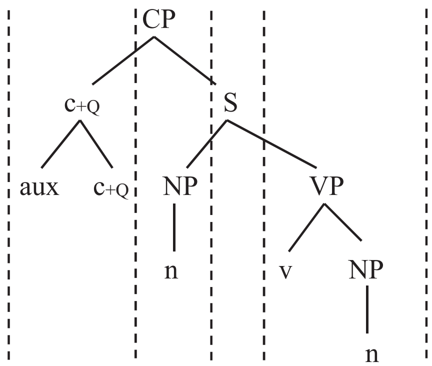
(7) $c_{+Q}$ n aux v n.

[4] Note incidentally that adjunction is an elementary transformation. Other elementary transformations include substitution, deletion, and possibly some others. The prior probability of a transformation that will be proposed below is meant to cover adjunction only.

(8)




(9)




Turn to Hypothesis 1. One way of instantiating this structure-independent hypothesis is shown in (10).

(10) *Front the first Aux*: Structure-independent; Root; Obligatory

| | | | | | |
|---|---|---|---|---|---|
| SA: | c+Q, | | , | aux, | |
| SC: | 1 | 2 | 3 | 4 | => |
| | 3*1 | 2 | | 4 | |

*where no aux exists in 2.*

A few technical notes are in order here. First, adoption of structure-independent transformations does not entail that the base component of the grammar be a flat grammar. It is possible to formulate this rule to apply to trees generated by a CFG. Recall that Perfors et al. showed the superiority of CFGs over flat grammars. Let's assume for the sake of argument that CFGs can be paired with not only a purely "hierarchical" transformational component but also with one containing "linear" rules. Nothing technically forces us to say that structure-independent rules have to operate over flat representations.

Then, how can the structure-dependent-independent distinction be characterized by using the type of rule format employed here? The answer is that structure-independent rules can be formulated as those operating on *terminal symbols* exclusively. Being structure-independent, the 'Front the first Aux' rule mentions no non-terminals in its SA.

Next, the rule in (10) incorporates an analyzability condition on the SA, which must be met by trees in the rule's domain of application. To express the property of being "the first Aux," (10) says *c+Q* can be followed by any terminal string except that the string must contain no auxiliary element. This condition assures that the auxiliary occurring in Term 3 be the first auxiliary in the sentence. Freidin (1991) observes that the possibility of structure-independent rules like "Front the first Aux" can be eliminated by "a constraint on the form of transformations that prohibits the counting property." The formulation of the rule in (10) suggests that transformations are so powerful that the system may acquire a counting property without introducing a counting tool independently. In other words, given the general properties of transformations, prohibition of counting seems impossible to accomplish unless stipulated.[5] Somewhat ironically, however, it is this very powerfulness of the device that enables us to measure the simplicity of various versions of auxiliary fronting.

The analyzability condition found in (10) helps to formulate other structure-independent rules that have been discussed in the literature. For instance, the 'no aux' condition enables us to express the rule "Front the second Aux," as in (11).

(11) *Front the second Aux*: Structure-independent; Root; Obligatory

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SA: | c+Q, | , | aux, | , | aux, | | |
| SC: | 1 | 2 | 3 | 4 | 5 | 6 | => |
| | 5*1 | 2 | 3 | 4 | | 6 | |

*where* (i) *no aux exists in 2 and (ii) no aux in 4*.

This time, Terms 2 and 4 are conditioned. It is clear that the rule becomes more complex as the ordinal number assigned to the moving auxiliary becomes larger. Likewise, (12) and (13) are the rule "Front any Aux" and "Front the last Aux," respectively. Freidin (1991) discusses the former while Lasnik and Uriagereka (2002) discuss the latter.

(12) *Front any Aux*: Structure-(in)dependent; Root; Obligatory

| | | | | | |
|---|---|---|---|---|---|
| SA: | c+Q, | , | aux, | | |
| SC: | 1 | 2 | 3 | 4 | => |
| | 3*1 | 2 | | 4 | |

(13) *Front the last Aux*: Structure-independent; Root; Obligatory

| | | | | | |
|---|---|---|---|---|---|
| SA: | c+Q, | , | aux, | | |
| SC: | 1 | 2 | 3 | 4 | => |
| | 3*1 | 2 | | 4 | |

*where no aux exists in 4*.

---

[5] So far as the analyzability condition "No aux in ..." is available. See Chomsky (1965: 56) and Bach (1974: 97) for discussion of restricting possible analyzability conditions to Boolean conditions. Note that the 'no aux' condition requires more than Boolean conditions: It involves existential quantification. See also Bach (1974: 217) for other examples of analyzability conditions violating this Boolean requirement.

In passing, the 'Front any Aux' rule can be structure-dependent or structure-independent, as Freidin (1991) correctly observes. In relation to the way it is formulated in (12), the distinction lies in whether to include nonterminals in the vocabulary for the transformation. It will be shown below that the structure-independent 'Front any Aux' rule is simpler than its structure-dependent counterpart.

The next section is devoted to proposing a way of scoring these transformational rules by simplicity.

## 3. The Prior Probability

Suppose that a CFG *G* has *V* vocabulary items and *n* nonterminals. A transformational rule *R* operates on P-markers or trees generated by *G*. *R* can be (i) obligatory or optional, (ii) root or non-root, and (ii) structure-dependent or independent. It is assumed that each of these choices concerning rule type is made with parameter 0.5. For example, Extraposition, an optional, non-root and structure-dependent transformation, and the 'Front the first Aux' rule are equally valued at this stage of rule construction. Both have probability 0.125 at this point.

Next suppose that transformation *R* have *t* terms, where $t \geq 1$. The probability for this choice, *p(t)*, is assumed to be a geometric distribution; see Perfors et al. 2006, 2011, where geometric distributions are used to determine *p(n)* and *p(N)*. This geometric distribution represents the probability that the *t*th Bernoulli trial is the first success. The probability of having two terms is regarded as the probability of getting a head (success) in the second trial of coin flipping after getting a tail (failure) in the first trial. We simply assume the success probability *P* is 0.5.

(14) $$p(t) = (1 - P)^{t-1} \; P$$
where $t \geq 1$

Hence, as the number of terms of a rule becomes larger, *p(t)* becomes smaller and therefore the rule becomes more complex, which seems to be in accord with our intuition.

How is the prior of each term in the SA calculated? Term *i* ($1 \leq i \leq t$) in *R*'s SA has $s_i$ symbols ($0 \leq s_i$). The string of symbols could be "NP", "aux v", etc., but it could also be zero. Terms having no symbols serve as variables, as mentioned above. Thus, how many symbols to use for a given term is determined first, and then what symbols to use is determined. Each symbol is selected from the set of *V′* vocabulary items. *V′* equals *V* for structure-dependent transformations (i.e., the sum of the number of terminals and the number of nonterminals in *G*) , whereas *V′* equals *V′* minus *n* for structure-independent transformations (i.e., the number of terminals). In theoretical literature, the symbols to use for specifying a term are not limited to terminals and nonterminals defined in the CFG. Labeled brackets such as "[s", for instance, can also be used and have actually been used. The set of symbols assumed to be available here would need to be augmented if one wished to include those symbols.

After being specified the way described above, Term $i$ can be further specified by analyzability conditions, as noted in the previous section. A rule can require that a certain term contain an NP, for example. There are $c_i$ conditions for Term $i$ ($0 \leq c_i \leq V'$). An analyzability condition $k$ for Term $i$ is expressed as " +" or "–" followed by a symbol selected from the vocabulary. The plus sign means Term $i$ must contain the item that follows while the minus sign means the term must not. Also, it should be stressed that a category symbol appears only once per term for the purpose of analyzability conditions: "the term must contain an NP" does not occur more than once and excludes "the term must not contain an NP". Thus, the probability of the category symbol for the $k$th analyzability condition of a given term being selected is $1/(V'+1-k)$. Once a symbol is selected, either "+" or "–" is added to the symbol with probability 0.5. Summarizing, $1/2(V'+1-k)$ is the equation that determines what the analyzability conditions for a term are like.

The expressive power of the analyzability conditions on terms defined that way is obviously not sufficient. Identity in reference, for example, cannot be expressed in the current vocabulary because the proposed system lacks the two-place predicate "identical." The Reflexivization transformation and some others therefore cannot be implemented. Also, lexical features such as [+Q] cannot serve as analyzability conditions in the system. For the current purposes, though, the set of analyzability conditions made available here suffices for us to define various Aux-to-C movements.

The prior probability of an SA can be computed by using (15).

(15)

$$\prod_{i=1}^{t} p(s_i)\left(\frac{1}{V'}\right)^{s_i} \begin{cases} p(c_i) & \text{where } c_i = 0 \\ p(c_i)\displaystyle\prod_{k=1}^{c_i} \frac{1}{2(V'+1-\mathrm{k})} & \text{where } c_i \geq 1 \end{cases}$$

Note that although $p(s_i)$ and $p(c_i)$, like $p(t)$, are geometric distributions, they are different from $p(t)$ in that Term $i$ could represent a null string and involve no analyzability conditions. They are assumed to be geometric distributions of another kind: $(1\text{-}P)^n P$. This represents a probability of getting a head (success) after $n$ failures in coin flipping.

(16) a. $p(s_i) = (1-P)^{s_i}P$
where $n \geq 0$.

b. $p(c_i) = (1-P)^{c_i}P$
where $n \geq 0$.

As for the prior of an SC, the term that undergoes fronting is determined by selecting one term from the $t$ terms and the landing site is selected from the other terms than the already selected "moving" term. As noted in footnote 4, other transformations than adjunction are out of the scope here. (17) cannot be employed for deletion, for example.

(17)

$$\left(\frac{1}{t}\right)\left(\frac{1}{t-1}\right)$$

To put all these together, the prior probability of a transformational rule given rule type T is given by:

(18)

$$P(R|T) = \left(\frac{1}{2}\right)^2 p(t)\prod_{i=1}^{t} p(s_i)\left(\frac{1}{V'}\right)^{s_i}\left\{\begin{array}{l} p(c_i) \\ \text{where } c_i = 0\text{; otherwise,} \\ p(c_i)\prod_{k=1}^{c_i}\frac{1}{2(V'+1-k)} \end{array}\right\}\left(\frac{1}{t}\right)\left(\frac{1}{t-1}\right)$$

## 4. Scoring the Rules

For the sake of concreteness, the discussion that follows assumes the CFG given in (19), which comprises four non-terminals, four terminals, and seven productions.

(19) *CFG*

a. Four non-terminals: CP, S, NP, VP

b. Four terminals: $c_{+Q}$, $c_{-Q}$, aux, n, v

c. Sven productions

CP → $c_{+Q}$ S | $c_{-Q}$ S

S → NP aux VP

NP → n

VP → v | v NP | v CP

Which CFG is chosen as the base component affects the prior of an auxiliary fronting rule in a certain way. Most crucial is how many vocabulary items are provided by the CFG. Adding a new rule "VP → v NP NP" to the set of productions would not affect the scores of the prior of any transformation because no new vocabulary items are added by introducing the new rule. Addition of a new vocabulary item, say, "det," however, would lead the prior scores to drop because the number of vocabulary items that must be referred to in constructing rules increases by one.

Table 2 illustrates how the prior probability of a rule is calculated, taking the "Front the first Aux" rule as an example. While the CFG has eight vocabulary items, the relevant items here are four terminals. This reflects the fact that structure-independent rules only make reference to terminal symbols. Assuming the CFG in (19), the priors for the five hypotheses are shown in Table 3, ranked by logged probability.

The two 'Front any Aux' rules are the most preferred two rules. The structure-independent rule is preferred most because there are (i) only two non-variable terms (Terms 1 and 3), (ii) no conditions on terms, and (ii) four terminals as vocabulary items. The structure-dependent version is the same except that four nonterminals are also included in the vocabulary. Thus, the former is slightly simpler than the latter. The 'Front the first Aux' and 'Front the matrix Aux' rules are placed 3rd and $4^{th}$, respectively. Notice that all other things being equal, the first and last Aux hypotheses are preferred to any *n*th Aux hypotheses ($n \geq 2$). It can also be inferred that the larger *n* becomes, the more complex the rule.

Table 2. Calculation of the logged prior probability for the 'Front the first Aux' hypothesis (10). The number of relevant vocabulary items (= *V'*) is four, and the parameter for geometric distributions is 0.5.

| Property of the rule | | Component | Probability |
|---|---|---|---|
| -Dependent, +Root, +Obl | | Rule type | 0.5*0.5*0.5 |
| # terms (= $t$) | 4 | $p(t)$ | 0.0625 |
| # symbols in Term 1 (= $s_1$) | 1 | $p(s^1)$ | 0.25 |
| Term 1: $c_{+Q}$ | | $(1/V)^{s1}$ | 0.25 |
| # conditions on Term 1 (= $c_1$) | 0 | $p(c1)$ | 0.5 |
| # symbols in Term 2 (= $s_2$) | 0 | $p(s^2)$ | 0.5 |
| Term 2: | | $(1/V)^{s2}$ | 1 |
| # conditions for Term 2 (= $c_2$) | 1 | $p(c^2)$ | 0.25 |
| Condition on Term 2: -Aux | | $(1/2V)$ | 0.125 |
| # symbols in Term 3 (= $s_3$) | 1 | $p(s^3)$ | 0.25 |
| Term 3: aux | | $(1/V)^{s3}$ | 0.25 |
| # conditions for Term 3 (= $c_3$) | 0 | $p(c^3)$ | 0.5 |
| # symbols in Term 4 (= $s_4$) | 0 | $p(s^4)$ | 0.5 |
| Term 4: | | $(1V)^{s4}$ | 1 |
| # conditions for Term 4 (=$c_4$) | 0 | $p(c^4)$ | 0.5 |
| SC | | $(1/t)*(1/t-1)$ | 0.083 |
| | | Logged Prior | -19.814 |

Table 3. Rankings of the six hypotheses with the CFG in (19), which has four nonterminals and four terminals.

| Ranking | Transformational rule | Prior |
|---|---|---|
| 1 | Front any Aux (Structure-independent) | -17.041 |
| 2 | Front any Aux (Structure-dependent) | -18.427 |
| 3 | Front the first Aux | -19.814 |
| 3 | Front the last Aux | -19.814 |
| 5 | Front the matrix Aux | -21.200 |
| 6 | Front the second Aux | -29.741 |

Finally, it can be demonstrated that the addition of a new terminal to extend the CFG (unsurprisingly) makes the scores for all the rules a little worse, but the rankings are kept intact. Compare Tables 3 and 4.

Table 4. Rankings of the six hypotheses when the CFG is given one more terminal.

| Ranking | Hypothesis | Prior |
|---|---|---|
| 1 | Front any Aux (Structure-independent) | -17.487 |
| 2 | Front any Aux (Structure-dependent) | -18.663 |
| 3 | Front the first Aux | -20.483 |
| 3 | Front the last Aux | -20.483 |
| 5 | Front the matrix Aux | -21.553 |
| 6 | Front the second Aux | -30.857 |

## 5. Discussion and Conclusion

The discussion in the previous sections suggests the following general conclusions.

- Transformational rules can be given the counting property without introducing an independent counting device.
- All other things being equal, structure-independent transformations are simpler than their structure-dependent counterparts.
- All other things being equal, transformational rules affecting any auxiliary are simpler than their counterparts affecting a selected auxiliary.
- All other things being equal, the first (or last) Aux hypothesis is simpler than any *n*th Aux hypothesis ($n \geq 2$).

These results support Chomsky’s conjecture, seen at the outset of the paper, that the ‘Front the first Aux’ rule is simpler, compared to the ‘Front the matrix Aux’ one.

Many important questions have to be left untouched here. How are the likelihood probabilities of the rules calculated? What (kind of) training corpus should be used to measure the likelihoods? and so on and so forth. Note that, as Pullum and Scholz (2002) and Legate and Yang (2002) observe, actual child-directed corpora contain "direct evidence" against the 'Front the first Aux' rule. Chomsky's original argument is based on the assumption that the learner only uses simple questions to choose among the alternative hypotheses, as noted earlier. Critical in the present context is that the property of the corpora in question would prevent us from calculating the likelihood probabilities of the structure-independent hypotheses in a meaningful way except "Front any Aux." The likelihood probability of the 'Front the first Aux' rule, for instance, would be zero under any of the corpora containing utterances like (1). For question sentences with the "second" auxiliary inverted cannot be parsed by the grammar, and the grammar with the correct structure-dependent rule would be more highly valued in a trivial way. There are two ways of finding out the likelihood probability of the first-Aux alternative. One way is by using a hypothetical corpus containing no multiple-auxiliary questions while continuing to make the Chomsky assumption. The other way is by using a more realistic corpus containing examples like (1) and introducing a descriptively more adequate linear rule. One such linear hypothesis can be found in the Lasnik-Uriagereka list of auxiliary fronting rules (Lasnik and Uriagereka 2002): the "Front the first or last Aux" rule. Let us conclude the paper with the following question. Is this rule formulable in the transformational framework adopted here? We wish to take up this and aforementioned open questions in a future study.

## References


Chomsky, Noam (1957) *Syntactic Structures*, Mouton, The Hague.

Chomsky, Noam (1965) *Aspects of Theory of Syntax*, MIT Press, Cambridge, Mass.

Chomsky, Noam (1975) *Reflections on Language*, Pantheon, New York.

Bach, Emmon (1974) *Syntactic Theory*, Holt, Rinehart and Winston, New York.

Baker, Carl L. (1978) *Introduction to Generative-Transformational Syntax*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Emonds, Joseph (1976) *A Transformational Approach to English Syntax: Root, Structure-preserving, and Local Transformations*, Academic Press, New York.

Freidin, Robert (1991) "Linguistic Theory and Language Acquisition: A Note on Structure-Dependence," *Behavioral and Brain Sciences* 14: 618–619.

Lasnik, Howard and Juan Uriagereka (2002) "On the Poverty of the Challenge," *The Linguistic Review* 19: 147–150.

Legate, Julie and Charles Yang (2002) "Empirical Re-assessment of Stimulus of Poverty Arguments," *The Linguistic Review* 19: 151–162.

Pearl, Lisa (2021) "Poverty of the Stimulus Without Tears," *Language Learning and Development.* https://www.tandfonline.com/doi/full/10.1080/15475441.2021.1981908

Perfors, Andrew, Joshua B. Tenenbaum, and Terry Regier (2006) "Poverty of the Stimulus? A Rational Approach." In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Ron Sun and Naomi Miyake, eds., New York & London, Psychology Press, 663–668.

Perfors, Andrew, Joshua B. Tenenbaum, and Terry Regier (2011) "The Learnability of Abstract Syntactic Principles," *Cognition*, 118: 306–338.

Pullum, Geffery K. and Barbara C. Scholz (2002) “Empirical Assessment of Stimulus Poverty Arguments,” *The Linguistic Review* 19: 9–50.