Do any of our Models of Parameter-Setting Work? Just-so Stories aren't Sufficient

Ken Wexler, MIT

York Workshop, Feb. 13, 2015

Acquisition Prolegomena: *Very Early Parameter Setting*. Basic parameters of clause structure are correctly set at the earliest observable time (beginning of 2 word stage; 18 months, or even earlier). These include, head-complement order (SVO/SOV), Verb to Tense (French/English) V2 or not (German/English), move wh or not, null-subject or not (Italian/English; note English kids' null-subjects are *not* the result of a mis-setting of the parameter, universally agreed by acquisition investigators, including Hyams. Key data: English-speaking kids don't omit subjects of finite verb what an object or adjunct is questioned: #what e eats, #where e is going, unlike Italian kids, who do this all the time, (Bromberg and Wexler, Roeper and Rohrbacher for English, Guasti for Italian).

One data example: V2 in Dutch, utterances counted from 47 kids (1;7-3;2), where "Nonfinite" means non-finite in root contexts (the Optional Infinitive stage). Kids almost universally raise finite verbs and don't raise non-finite verbs. (Wexler, Schaeffer and Bol 2000).

	V1/V2	V-final
Finite 1	953 (99%)	11 (2%)
Nonfinite	20 (1%)	606 (98%)

So we need a fast model of parameter-setting that operates quickly at a fairly young age and with rather limited processing results (kids under 1;6, probably much younger, when this is all operating). Feasible system is needed. But we don't even have a system that actually works in the limit, even if we give up feasibility! Let's discuss.

(1) The plot to date: The standard *Principles and Parameters* framework assumes that parameter values (let's say always binary) are easily set, and (hopefully?) without (much?) error. Major problem: no theory has been shown to have this property. Moreover, there are analyses that show that systems that seem to capture the intended way it works, do *not* appear to have this property.

(2) In particular, Gibson and Wexler (henceforth GW) showed that – in a model in which learners tried to set a parameter by matching a piece of evidence, "local maxima" existed, and the learner could not ultimately get to the correct state (set of parameter values). The learning was "learning on errors", as in Wexler and Hamburger (1973), Wexler and Culicover (1980) and much of psychological learning theory. GW argued that this was the intention of the P and P non-formal statement of the property, and they were simply trying to formalize that informal statement. Their work didn't argue against the P and P intuition; rather it showed that there were problems for the intuition, problems that had to be addressed.

(3) GW spent some time on solutions, including default values of parameters, temporal assumptions, maturation, etc., and perhaps (with maturation) (somewhat) solved the 3-parameter system that they worked on.

(4) Bertolo et al (a couple of papers) expanded the system to 13 parameters (*not* selected for being particularly "learnable", rather, to capture attested syntactic variation), adding quite a bit more syntactic detail, in hopes that the richer system would make the problem more tractable. It actually became worse; there was no way that the standard P and P learning model (Triggers) could learn the parameters in this system, it seemed.

Cue Models

(5) In response to GW and the following trigger papers, Dresher argued that a completely different kind of model was needed; one in which the "cues" for each parameter value were innately specified. He took as a model the system for learning the values of stress parameters in phonology that Dresher and Kaye had worked up.

(6) Actually the Dresher and Kaye system didn't work, wasn't learnable according to the cue analysis, even in Dresher and Kaye (see fn. 34 of Sakas and Fodor for a concise summary of the general issue). Even in such a limited, "finite" system as the stress system in phonology, the method didn't work. It is one reason to not easily gravitate toward a built-in cue system. Other reasons include:

(a) the lack of unambiguous triggers (triggers that necessitate that a parameter be set to a particular value)

(b) the apparent need for a huge number and complexity of built-in triggers, redundantly specifying what should "simply" set the parameter value

(c) the large complexity of computing whether a trigger was unambiguous, if the trigger wasn't specified innately. Can we imagine that a child can do this? (Even an adult).

(d) the lack of any model that seemed to work

(7) Reason (6c) was particularly important given (Janet) Fodor's arguments for a type of cue theory in which the learner calculated on-line whether a piece of evidence constituted an unambiguous trigger (the "Structural Triggers Learner (STL"). Sakas and Fodor (2012) mostly don't argue for STL, their paper isn't about that. Though there is a small section toward the end holding out hope that it will be useful. None of the actual computational arguments, however, argue for STL, rather simply for cues.

(8) Sakas and Fodor (henceforth SF) want a "deterministic" parameter-setting system. This means one that makes no errors, that can set a parameter correctly (to either value) from a piece of evidence, and won't set the parameter if it can't be so set; rather the learner will wait until the unambiguous evidence arrives.

(9) They claim to almost achieve this goal in their 13 parameter system (or to achieve it, given a certain (but acceptable) loosening of definitions.

(10) SF follow the strategy of GW and Bertolo et al, of actually specifying a set of (hopefully correct) syntactic parameters and then investigating precisely the consequences for learning. This isn't a universal strategy. For the most part, papers on an alternative model (Yang) don't operate with this type of real parametric system, and (as I argued in "Grammatical Computation in the OI Stage" (de Villiers and Roeper, eds.)), these papers often or typically don't run any actual learning computations about syntactic parametric systems, especially involving several parameters. So different types of conclusions can be reached.

(11) The desire to be logical and clear and to do real derivations is laudatory (think how we try to do syntax and semantics). This desire also keeps SF on the modest, appropriate side of what kinds of conclusions they can draw. It

helps the science to advance by making this kind of work business as usual – can we advance the state of the art?

THE SET OF PARAMETERS IN SF

(12) Some general properties of the set of 13 parameters:

a. Bears some similarity to Bertolo et al, but different in some crucial detail; One can ask about which is closer to the nature of UG.

b. Some examples of similarities: headedness, V to I, V2 (but done differently, as 2 separate parameters in SF), existence of adverbs to show (via word order) whether various kinds of verb raising took place, etc.

c. **Some of the assumptions SF make seem odd from a linguistic theory point of view, and one suspects that the reason is that SF have gone out of their way to put in parameters that have give-away evidence for setting to the correct value, parameters that are much more transparent than the real parameters (we'll discuss examples).**

d. SF specify a "default" value for every parameter. Sometimes because they think it might help learning. Sometimes on general considerations (no movement is default compared to movement (though they have to give this up later for one parameter)), sometimes just because it makes their computations uniform and simpler.)

****

(13) The parameters (see Table 1 in SF):

1. Subject Position (is it <Spec,I, I> or <I, Spec,I>?). One might question in fact, there are any languages in which subject is on right of I. (GW included the SF possibility, but didn't find too many languages in which subjects are on right. Perhaps they get there via some other process?). Default: initial subject (I think that's what they mean)

2. Headedness in IP, NegP, VP, PP: Is the head of these on right or left of the complement? (assumption is they are all the same in one language, at least that's what I *think* the assumption is).

3. Headedness in CP: Does C precede or follow it's (IP) complement? Here we seem to have good examples of both values, e.g. English (where C is on left) versus Japanese (where C is on right).

4. Preposition Stranding: Is stranding possible or not? (I think if it is possible it's optional; at any rate, that's what it should be (Default: No stranding.

5. Topic Marking. This is simply the question of whether a topic is morphologically marked. In their notation, it's marked if there is *wa*

(Japanese topic marking). Note that *wa* itself is optional on the surface in Japanese, even if the DP *is* a topic. But in SF, I think topic marking is obligatory if it exists. This is a case, possibly, of (12c) above – namely, an incorrect assumption was made, possibly to make the learning problem simpler. After all, if *wa* is obligatory, then it's easy to see whether a DP is a topic from the surface input. Presumably, SF could have assumed semantic interpretation at least sometimes gives this information, but they were trying to make everything very simply learnable from surface data.

6. Null Subject. Default value, *no*, a null-subject isn't possible. They don't distinguish different kinds of null-subjects, or partial null-subject languages, or the languages where pleonastics are null, but not referential DP's, or discourse omission, etc. But fair enough, they're trying a simple set of parameters.

7. Null Topic.

8. Wh-Movement. Does a wh phrase move or not on surface, e.g. Japanese *no*, English *yes*? The question is, is there *obligatory* movement of a wh-phrase. I think they put it this way because, of course, a wh-phrase could move via some other process (at least that's what I think they have in mind, not all that clear, somebody could check out the web link, where more detail given they say). Default value: obligatory.

9. Affix Hopping. A bit of an old-fashioned term, but I think they mean this: The verb and tense have to "get together." So either V→ I, *or* I→ V (I lowering) (that is, Affix Hopping) must hold. If that's true, I wonder why they need this as a separate parameter; it would just be a UG property that V and I must "get together." So English has +Affix Hopping. Default Value: NO.

10.VtoI Movement. Default Value is No. (English has default, French has non-default).

11.ItoC Movement. Does I move to C? (They assume as matter of UG that a V can't move directly to C, it has to move through I first. So V2 languages are *yes*, English is *no* on VtoI as well as ItoC). (Doesn't account for aux sometimes winding up in C in English, not part of their system). Default value *No* (they try to make "no movement" default when they can, along with what they say is a general linguistic property, as has been argued). NOTE 2 important properties:

(i) V2 is broken down into "independent" parameters., ItoC Movement and Obligatory Topic (cf: parameter 13). So they don't in any way link a DO winding up in Spec, C with movement of a finite verb to C in V2 languages, it's just an accident of 2 separate parameters. More on this later.

(ii) We know that there are languages in which V → I is *no*, but they are verb second languages, so V→C is *yes*. Any of the mainland Scandinavian languages: Danish, Norwegian, Swedish. We know this because these are V2 in root clauses (so V→C) but, in embedded clauses, the finite verb does *not* raise around negation and adverbs, so *not* V→I. Importantly, SF use *only* degree-0 data as input (Wexler and Hamburger 1973 terminology), that is, sentences with no embedded sentences, simple clauses. So this kind of data would never be encountered by the learning system and thus it would never learn the parameter. They don't say, I think, why they made this decision (I seem to recall that Bertolo et al used degree 0 and degree 1 data, for this kind of reason), but probably they would justify it by simplicity of the computation, just for getting off ground. Fair enough, though it misses much. Is there any way of learning the value of V→I parameter in Danish, etc., in degree 0 sentences? Some linguists have argued that the parameter value is related to richness of agreement (Danish has no agreement at all), and that might be an out. Or one would need at least degree 1 data. For reasons I don't understand (because he would know the facts of European comparative syntax), Lightfoot has argued that degree 0 data is sufficient (perhaps supplemented by something from lower C system?). Perhaps SF have those kind of arguments in mind, but they obviously are problematic.

12. Q-Inversion (ItoC in questions). This parameter set to *yes* means that ItoC takes place in questions. Obviously if parameter 11, general ItoC, is set to *yes*, Q-Inversion takes place anyway. But stating this second parameter allows for cases where general ItoC is *no*, but ItoC for questions takes place. I think SF would consider English to be *yes* on Q-Inversion for English, even though it's the aux that moves to C. They say they mean the finite verb. They say that dummy *do* simply counts as a finite verb in their system. So nothing about their system tells us why dummy *do must* be inserted (or however we talk about it in syntax) when there is a question and no other aux. So their system would accept, say, *Eats John ice cream?* Ok, they're going for simplicity again, this isn't a syntactic theory, just a proto-proto system, just pointing out what they have. Default value: *no*. (Goes along with *no movement* as default).**On other hand, this looks like a case of forcing a parameter into a mode such that it can be set, but thereby losing descriptive adequacy and the correct generalization.**

13. Optional (versus obligatory) Topic. What SF mean is that a "Topic" in this sense is the movement of a phrase to Spec, C. In their system, all languages have topics, can move a phrase to Spec, C. But in some languages, this existence of a moved Topic is obligatory. So English is *no* on this

parameter; it can have Topics, *Beans, I like*, but German and other V2 languages are *yes* on this parameter; the moved phrased in Spec, C is what SF mean by Topic. Probably they mean that if a phrase is marked with *wa*, then this would be the Topic that moves to Spec, C. (Default: Obligatory. **This is a counter-example to the "no movement" as default**, perhaps, but perhaps it gives a default that's a Subset (smaller set of good derivations), helping with Subset Principle considerations. That is, "obligatory" often gives a smaller set of derivations than "optional." They discuss it in some way like this.

***Note*: This system does not in any way capture standard linguistic generalizations**. For example, nothing links "Obligatory Topic" and VtoC, contrary to what is usually taken to be fact. E.g. in embedded clauses in V2 languages like German, there is no topic possible and no movement of the verb to C. Languages that have "embedded V2" have both embedded topicalization and VtoC. One might say that the SF system of parameters is syntactically obtuse. **Desideratum: A system of parameter-setting that captures syntactic generalizations**.

****

(14) A *locally valid* trigger (p.95) means a sentence pattern that can *only* be generated by a particular parameter-setting for a particular parameter, when other parameters are set in a particular way. A *globally valid* trigger means a sentence pattern that can only be generated by a particular parameter-setting for a particular parameter *no matter how* the other parameter values are set.

(15) A *locally available trigger* for value v of Parameter P is a locally valid trigger for P (v) that is generated by some but not all languages that have value v of P. [I.e., if this sentence pattern occurs, you know how to set the parameter. But it doesn't occur in *every* language in which that parameter value holds, so you might sometimes need *another* trigger]. A *globally available trigger* for value v of Parameter P is a globally valid trigger for value v of P that occurs in every language whose grammar has value v of P.

(16) SF's *globally valid trigger* is what GW called a *global trigger*. Similarly, SF's *locally valid trigger* what GW called a *local trigger*.

(17) GW didn't have the notion of "available." In the Appendix, SF discuss why they need such a notion, because they are trying to find a set of cues that would work; it doesn't crucially matter whether the cue for value v of

Parameter P has one globally valid cue or a larger set of locally valid cues that together will allow the parameter to always be set to v. (SF seem to need a definition here that they don't give, something like "a set T of locally available triggers for value v of parameter P *spans* P if for every language L in which P (v) occurs [i.e. no matter what the settings for the other parameters] , there is a member of T that is a locally available trigger for P(v) in L [i.e. with the other parameter settings. Or L is perhaps better replaced by the parameter values, I'm simply taking the sequence of parameter values as the "language.")

(18) Example: in CoLAG (the system here), an object of a preposition separated from the preposition is a globally valid cue for preposition-stranding. E.g. any sentence (actually the pattern) like *Who did you give the book to* will instantiate this pattern (since *to* and its object *who* are separated) and be a globally valid cue for preposition stranding. This is probably the best case of a cue in CoLAG (it seems to me). I couldn't think of any clear counter-examples. I asked Norvin Richards (pc) if he knows of any language that does *not* have preposition stranding but *does* have a sentence in which the preposition and object are separated, and in a 10 minute discussion, no clear cases emerged, so it might be right. One would think of rightward processes that might occur to the object of a preposition, though no stranding in, say, wh-questions. E.g. I thought of complex NP-shift, but of course that applies to the PP, but not to the object of a P: **I gave the book to t on Tuesday [the man who had asked to borrow it*]. Norvin thought of pseudo-passives (*John was taken advantage of*), but so far as he knows, any language with such a construction also has preposition-stranding (in, say, wh-questions). Anyway, to date, this might actually be a good globally available cue. Any ideas that it's wrong?

(19) I-triggers: a piece of computation of the p-m that is the trigger. Somehow a "grammatical" characterization of the parameter.

(20) **Example (8) of SF, from idea of Lightfoot, Spec, CP with a phrase XP in it is the characterization of the positive value of the V2 parameter:**

Problem : Of course this can't be right in general since, e.g. a wh-phrase in Spec, CP (or a topic, or whatever) doesn't imply the general V2 parameter. A child learning English who used this cue would come up with a syntax

that included *obligatory topic* (like e.g. German), obviously wrong. Perhaps one simply needs more parameters and more refined parameters.

(21) **Then a sentence encountered is the E-language instantiation of this I-language characterization.** One could assume that learners have as part of UG not only the I-trigger definition but also the E-trigger definitions; this seems redundant and not very insightful. As SF say, a more ambitious undertaking would be a computational analysis done by the learner. Can this be done? They discuss it at end, inconclusively at best, it seems to me. One major issue is the large computational undertaking that would be demanded of a kid for them to calculate whether a sentence (or sentence pattern) can be generated only by languages with a particular value of a certain parameter. (Classic trigger models (cf: GW) don't demand any of this, that's one of their attractions. The idea there is to have a learner that has very little to do; the GW argument was that this is what the P and P approach in linguistic theory says should happen.) In this paper SF simply provide examples of E-triggers. So, at least for simplicity, let's think of a trigger as an E-trigger, specified as part of UG for the learner. That's what's done in CoLAG, anyway.

(22) **SF take an E-trigger as a "sentence pattern".** GW did this too, but there's a crucial difference. GW's sentence pattern constituted the whole sentence e.g. OVS (object verb subject), whereas SF say a trigger can be part of a sentence. We'll see examples. SF argue that a sentence pattern (e.g. something about a particular part of a sentence (everything else be a variable or something) is somehow more powerful). E.g. one can say something like, *ka* (explain later, this is Q-marker in C) occurs in sentence. If it were a sentence, one would have to list all possible sentence with *ka* (or all possible sentence patterns with *ka*). So at the least, there may be a compactness argument for parts of sentence as the triggers. May not be a crucial issue.

(22a**)** ***Problem*****: How *does* the child that hasn't set the parameters yet come up with a piece of grammar like "such and such a phrase is in Spec, C?"** The child needs an E-trigger (a sentence pattern) that shows clearly that a such and such a phrase in Spec, C. Lightfoot suggests (i) as the appropriate sentence pattern ("E-trigger" for SF) for (20):

(22b) **A sentence-initial non-subject immediately followed by a finite verb**

(22c) SF don't comment on this proposed E-trigger in detail, but point out that Gibson and Wexler expressed skepticism that such could be found in general). (20c) sounds good. After all, in V2 languages like German, such patterns exist. **But the devilish problem of the lack of unambiguous triggers once again raises its ugly head. Suppose that the language is an SOV language that allows null subjects, e.g. Korean or Japanese. Then OV (Object followed by finite verb) will be a very common form of input. The child using Lightfoot's attempted E-trigger will conclude that the object is in Spec, V and that Korean and Japanese are V2 languages (with V moving *left* to a left-branching C). Such a conclusion massively generates the wrong language, and is massively wrong, of course, on grammatical grounds, descriptive adequacy.**

(22d) The problem is the *lack of unambiguous cues* (SF's *valid triggers*). The effect of this little piece of sentence (initial object followed by finite verb) occurs in V2 languages, but can be mimicked by languages that *don't* move V to a left-ward C, nor raise a phrase to the Spec of that C. Rather the language can do this by omitting a subject and having the object generated before V. **This game of assuming that a parameter's value can be easily detected via an obvious sentence that "seems" to be a key to it is a "just so" story in exactly the sense of Lewontin commenting on how the theory of evolution (Darwin's principles) are used to derive cognitive (or other) traits.** In this case, it's particularly easy to see how most proposals couldn't possibly work, at least for any linguist minimally familiar with comparative syntax.

(22e) **What's the intuition about why this is wrong? Hypothesis/suggestion: Just maybe if we find a way to order (probably partially order) the parameters so that the learning system is not allowed to set parameter X until evidence for the value of parameter Y has set Y, then we can avoid (some of?) these problems. If, for example, the learner had unambiguously decided that the language was SVO or SOV, then the status of V2 might be much clearer in the above example. Will this work in general? We don't know, but it looks promising for at least some examples. What's crucial is to build real, as adequate as possible parametric systems and test whether there are ways to set them. It's trivial to construct arbitrary systems of parameters that one selects for their abilities to set (e.g. just construct 20 parameters, all of them easily able to be set unambiguously; they won't look like syntactic systems of parameters).**

COMPUTATIONS

(23) The point of SF's paper is to do computations that allow them to find triggers of particular kinds. It's a complicated enough system that the computer search is probably essential. Some brief results and comments:

(24) Table 1, p. 104 puts a check mark in for each parameter and each value if there is such a spanning set for that value of that parameter. In other words, if there is a check mark, there is a set of locally available cues that will allow the learner to deterministically and correctly set that value of that parameter, no matter what language (set of parameter values) she is learning.

(25) If value v of a parameter P doesn't have a basis (a set of locally available triggers for all languages in which v of P occurs), then Table 1 lists the percentage of languages in CoLAG which are missing such a set of locally available triggers for value v of P.

(26) The first 5 parameters have a basis for both values of the parameter.

(27) The next 5 parameters have a basis for every marked value, though a basis is missing for the default value. From49% to 100% of the languages are missing a basis for the default values. However, one could argue (as SF would like to do) that this is no problem because the learner starts with this default value. Note that this result could be taken to show that "default value" is important for the theory of parameter-setting, for learnability reasons (a conclusion GW already came to on the basis of their kind of triggering model).

(28) The last 3 languages are the true problems, because a basis is missing not only for the default values, but also for the marked values. It is these 3 parameters that SF spend most of the paper discussing.

(29) Step 3: Irrelevance. Compute which languages lack a basis only because they are surface-indistinguishable from another language. Since only weak equivalence is demanded of the learner in this attempt, we eliminate the weakly these cases from the set of problematic cases. (Note:

the lack of interpretive information leads us to this position. I think that SF are aware of the weakness of missing interpretive information).

(30) Step 4: Local validity. For the 3 problem parameters which have no unambiguous E-triggers, look for locally valid triggers that are conditioned by other parameter values. This occupies a great deal of their discussion. Let's see later if we can get any examples. If not, try to read through some examples.

(31) Step 5: Compactness. How to compactly characterize the triggers, i.e. as a simple piece of a pattern. Done "by hand." That is, look at the E-triggers and see if we can identify a "pattern" that characterizes them.

(33) **Ok, what *do* the E-triggers look like? And do they make sense? And do they seem *robust* (in AI sense)? That is, do they look as if they will scale up well as the parametric space is expanded. That's a crucial problem. Do the triggers work well (if they do) only because of an artificially curtailed space, or a wrongly defined space. This leaves us much room to investigate.**

(34) *TABLE 4* (P. 140) GIVES EXAMPLES OF TRIGGERS (GLOBALLY VALID, THAT IS, UNAMBIGUOUS, TRIGGERS) FOR THE NONDEFAULT VALUE OF EACH PARAMETER AND AN ATTEMPT AT DEFINING THE I-TRIGGERS. IT'S A CRUCIAL PLACE TO LOOK FOR HOW THE RESULTS ARE SUPPOSED TO WORK.

(35) Crucially, as SF say on p. 139, except for where it's otherwise stated, the E-triggers in Table 4 are *more* than just globally valid triggers. In addition, they are globally available, so that one of the triggers leads to deterministic learning of *every* language in CoLAG.

(36) Easy example, as discussed, preposition standing as E trigger, "P and O3 both present but not adjacent". OI is CoLAG's name for the object of a preposition. O1 is the direct object, O2 is the indirect object. We need to use these grammatical relations separately in the learning system, in the system of defining sentence "patterns", the triggers.

(37) NOTE: SF, following GW, assume that the learner can calculate (before having set the parameter values), the grammatical relations, that is, what

grammatical role each DP plays (along with knowing the grammatical categories, etc.). GW justify this by the assumption that semantic and syntactic interpretation has helped the child determine what the subject of the sentence is, the direct object, etc. How so? Well, perhaps agreement with the verb helped for the subject, perhaps some interpretive properties, etc. The object might show up in a certain position as the patient in a position. This isn't formalized; but the idea is that the learner already has this. I don't know of any learning system that hasn't made this type of assumption.

(38) One E-Trigger for Subject Final is "S[ubject] follows Verb[-FIN]". I think that works because a non-finite verb doesn't raise in CoLAG (**of course, they do in some languages, but perhaps not high enough to cause a problem for this E-trigger?**) and that means that for a subject to follow the non-finite verb, it had to be generated there; the verb couldn't have raised to the left around it. Are there counter-examples? Non-finite verbs that raise to left of the subject? **Are there no languages that are, say, SVO underlyingly, that raise a non-finite verb in some constructions?**

(43') Avendo Maria accettato di aiutarci, potremo risolvere il problema
'Having Maria accepted to help us, we can resolve the problem'
(Hyams (1987) following Rizzi (1982); Several other examples are given).

(43'') Of course this non-finite verb that precedes the subject is in an embedded clause, and CoLAG only looks at degree-0 sentences. But if we got a bit more general in coverage (and we would need to in order to set parameters involved in this infinitival constructions), **the subject would sometimes occur before the non-finite verb and sometimes occur after it in the language. What does this make of the E-trigger? Doesn't the learner then take Italian to be Subject Final?**

(44) **A 2nd E-Trigger for Subject Final is *S follows a non-sentence-initial Direct Object.*** In Co-LAG, the only leftward movements of a direct object are to Spec, C, which is always the initial position, when it is filled. So if the direct object is not in initial position, it is not in Spec, C, and therefore it is in its non-moved original position. Thus if the Subject follows such a non-initial direct object, the subject must have been generated to the right of the direct object because there is no rightward movement of the subject.

Therefore the subject must be generated on the right of Ibar; the value of the parameter is *Subject Final*.

(45**) Does this E-Trigger scale up to real syntax? How about VOS sentences in many Romance languages, e.g. Italian? The object is in non-initial position. The subject follows this object. Thus Italian has Subject Final, via this E-trigger.**

(46) One standard analysis of Italian, has the subject generated on the left of Ibar. The verb in Italian raises high. A plausible syntax is that the verb raises to a fairly high functional category (e.g. AGR) and the subject only raises to the Spec of a lower category (perhaps TNS). This would give us VSO. There have been arguments for leftward movement of the object, over S, but below V, in Italian. This would generate VOS, while the subject was generated on the left of I. [There were very old analyses, in 70's?, maybe early 80's?, of Italian, in which the subject generate someplace on right, and had a kind of pretty much superindexing with the subject on left, but I think these analyses have been abandoned. Of course, subjects to the right of Ibar may not be possible at all in UG, even if one doesn't accept Kayne.] Another possibility is that the object is generated as the left complement of V (SOV) and the verb always raises higher than the object. (NOTE: infinitives raise in Italian, so that we'd get the right orders).

(47) **Or one might argue for another analysis. There are 2 general issues:**

a. is the E-trigger robust so that it works on more correct set of analyses?

b. Is the E-trigger robust enough so that when further parameters are added, the trigger still works?

(48) **Crucial fact: As more parameters are added to a system, E-triggers can only be weakened, not strengthened. Why? Because adding parameters might make a globally valid E-trigger no longer a globally valid E-trigger in the expanded system. So robustness is a serious issue**.

(49) **Headedness in CP. Nondefault Value is Complementizer final. I-Trigger is *C is the right sister of its IP complement.***

(50) Two E-Triggers given. They are "both locally available but jointly sufficient." That means that together the E-triggers provide a basis, span the

space. One trigger allows the parameter to be set in some languages, the other triggers allows the parameter to be set in all the other languages.

(51) **First E-trigger: *overt complementizer* ka *in final position in a question***. *Ka* is taken as the morpheme that in some languages fills C in a question. So if in a question *ka* is in final position, C must be in final position. In CoLAG (maybe in all grammars?) there is no way for C to be in final position via movement of all the phrases to the left of C (in CoLAG at most one phrase may occur to the left of C, in Spec, C. At any rate, there is no process that will move a V to the left of C).

(52) **The second (and only other) E-trigger for Complementizer final: *Aux in final position in a question if no ka is present*.**

(53) **Suppose a language with C on the right has no Q-marker. Since the only other E-trigger for "C on the right of IP" is an aux in final position, what SF must assume is that in *every* language, either *ka* OR a aux is in C. (Or possibly this only has to be assumed for languages with C on the right of IP). Where do languages without V→C but without *ka* stand? *Perhaps* SF are assuming that *ka* might be phonetically invisible in a language when the finite verb isn't filling C. And they would also have to assume that the learner knows this via the fact that they can interpret the sentence (independently of parameters or markings) as a question. NO, that wouldn't work. If *ka* present but phonetically invisible, then the learner wouldn't have access to whether it was on right or left. So to make the trigger work either *ka* (a phonetically audible Q-marker) or a finite aux *must* be in C in every language. Is this true for all languages in which C is on the right of IP? It isn't true for all languages in which C is on left of IP, e.g. English neither has an audible Q-marker in C nor a finite verb in C. (**Since the aux is always finite in CoLAG (I assume), the aux is finite.)

(54) **Is it a universal that in a question, in a language in which C is on right of IP, in a question, C is *either filled by a Q-marker (ka) or the finite verb*. It's certainly not necessarily true for languages with C on the left, e.g. Italian.**

(55) In Italian, "…auxiliaries and modals cannot be inverted with the subject in tensed clauses."

*Ha Gianni mangiato

'Has Gianni eaten?'

(56) What's the good question, and what's in C? Possibly Italian is a candidate for no Aux→C in questions, and yet no Q-marker *ka*.

(57) Possibly many languages in which we don't have ItoC in questions, and yet no *ka* in C. Questions can be marked in other ways, e.g. via intonation or particular devices as in Mandarin.

(58) **So why did SF assume such (if they did? It seems that way). Is this a case of assuming a particular constraint because it makes finding an E-trigger easy? The insistence on 0–degree learning might have contributed to this wrong syntax. Otherwise one might look for the position of a complementizer with respect to an embedded sentence. Would that work in general as a trigger?**

(59) **Another Example of a "non-problematic" parameter: Null-Topic.**

(60) That is, is a null-topic permitted, where a topic means a phrase has moved to spec, CP, so that if a null-topic is permitted it means that a topic in Spec,CP may be phonetically null.

(61) Default Value: *NO* (that is, null-topics are *not* permitted)

(62) I-Trigger: The non-head daughter of CP has [+NULL]

(63) E-Trigger for Non-Default Value (+Null-Topic): "O2 present but O1 absent", that is an indirect object is in the sentence string presented, but the direct object is not.

(64) In CoLAG the only way for an object to be omitted is for it to be topicalized and then the null-topic possibility happening. The presence of the indirect object must mean for CoLAG that there is a direct object of the verb, so that it has had to be omitted, and thus it must have undergone topicalization and then omission (the latter via +Null-Topic).

(65) Thus the E-trigger in (63) is globally valid; if such a sentence pattern occurs, then the parameter value *must* be +Null-Topic.

(66) Even more strongly, the E-trigger in (63) is *globally available*; any language that has +Null-Topic must generate a sentence with the pattern in

(63), with an indirect object but no direct object. No parameter interaction can take that away.

(67) Is this right in general? Can we have a sentence with an indirect object but no direct object on the surface that doesn't have null-topic? Well, certainly if we include more processes, e.g. incorporation or whatever process accounts for: *I lend to anybody who is needy*. "Indirect objects" (if by that is meant e.g. a recipient or some other thematic role account) can often appear without the direct object.

(67') a. Who did you tell the story (to)?
b. I told Bill

(67b) should not lead to +Null Topic but what will prevent it given the E-trigger in (63)? There are probably large numbers of examples like this; do we really want +Null-Topic to depend on the omission of an indirect object?

(68) Perhaps such examples are too "advanced" or special in some other way and the learner won't understand those enough to affect learning. Or perhaps a learner will never hear them at all. But that should be demonstrated. Perhaps in many languages extremely "simple" sentences will have this pattern (how about object omission *without* topicalization?) **This is another example of how adding more grammatical processes takes away a globally available E-trigger. One might worry that this can happen all too often**.

AN EXAMPLE OF A "PROBLEMATIC" PARAMETER AND A "SOLUTION": A BETWEEN-PARAMETER DEFAULT

(69) Optional (versus obligatory) Topic has no set of E-triggers that span the non-default value, i.e. no set of locally available triggers that will always (in all languages) set the non-default value. See Table 1, last parameter (23% of the languages are missing such a "basis" (in our terminology).

(70) What to do?

(71) Default value: Obligatory Topic. The problem is that when the topic is optional, for 23% of such languages, there is no set of E-triggers that will always set the value correctly to +optional topic.

(72) **The problem is that there is an interaction between the Null Topic parameter and the Optional Topic parameter – see discussion around pp. 110-113. Intuitively if null topics are permitted, hard for learner to figure out if topicalization has taken place and the topic omitted or, on the other hand, if the topicalization hasn't taken place. I won't go through here why the attempted solutions don't seem to work.**

Very Brief Comments on an Alternative to Cues: Reinforce or punish *all* parameters in a randomly chosen set of parameters (Yang) (see my "Grammatical Computation in OI Stage" for more detailed comments

(73) One statement of Yang's Model (Yang 2010), where P is a sequence of probabilities p(i), representing the probability that parameter i is set to value 1.

a. Upon receiving an input sentence **s**, the learner uses P to probabilistically (and thus non-deterministically) generate a composite grammar **G**.

b. If **G** can analyze **s**, reward all the parameter choices in **G**; i.e., increase/decrease
pi if _i has been chosen the value 1/0

c. If **G** fails to analyze **s**, punish all the parameter choices in **G**

(74) In brief, the learner has a probability for each of the 2 values of all parameters. On hearing a sentence, she chooses a grammar according to these probabilities. If the grammar generates the sentence, then the probabilities of all of these parameter values is increases If the grammar fails to generate the sentence, then the probability of all of these parameter values decreases.

(75) Learning Model: Bush Mosteller linear model, pretty much discarded by the early 1970's in learning theory, but it's not clear that the particular model of learning is essential to Yang's theory. (It might be for particular developmental predictions).

(76) We don't know that this particular probabilistic model always converges to the correct grammar. It's claimed that the proof is trivial, but not so clear, I think. Would be good to see a proof.

(77) In particular, a crucial question: to what extent does the model rely on unambiguous cues. Is the existence of unambiguous cues necessary for the model to learn? Is that one of the assumptions that underlie the claim that the model always learns?

(78) I think it must. But somewhat confusing. E.g. Yang (2000), in criticizing cue models writes (p. 12): "In practice, grammars in very small parameter domains (e.g., Gibson & Wexler 1994, Fodor 1998) do have unique unambiguous evidence for their correct identification, though the matters are not clear in realistic parameter spaces."

(79) So this seems to suggest that realistic parameter spaces might not have unambiguous evidence (something that I've just argued for in the SF system).

(80) But in practice, Yang's analyses seem to depend on unambiguous evidence.

(81) From same paper, Yang (2010), p. 14,: "The probabilistic nature of the variational learning (Yang 2002) can take advantage of the parametric space in a different way. Specifically, many parameters may be associated with signatures (Yang 2002, 39). The signature for a parameter refers to sentences that are analyzable only if that parameter takes on the correct value of the target language. Empirically, it is not difficult to find parameters with signatures."

(82) A "signature" as defined above is simply another term for "unambiguous evidence." That is, the sentence is analyzable only of the parameter has the correct value. And of course, that must mean independently of other parameters.

(83) The learning computations or even informal arguments for this system that I can find always seem to rather explicitly depend on "signatures", that is on unambiguous evidence. A major example is verb raising, as in French, as in (84), from Yang 2010, p. 14 (as an example of a signature).

(84) a. Jean voit souvent Marie
John sees often Marie
b. John often sees Marie

(85) Yang, p. 14: "The relative position of finite verbs and adverbs, then, would be the signature for the verb to tense raising parameter: when a child learning French has the parameter to the English option, it is guaranteed to fail upon seeing sentences such as (4a), whereas the English learner cannot analyze (4b) if it has selected the French option either."

(86) Thus the finite verb appearing before the adverb is a "signature" of the positive value of the "verb to tense" parameter. A classical idea about unambiguous evidence. A signature is a piece of unambiguous evidence.

(87) But there is no discussion of *other* parameters. Remember, to have unambiguous evidence in the sense that hearing that evidence *must* lead to correctly setting the parameter means that this must be true no matter what the other parameter values are. There is no discussion of this issue.

(88) What is the "signature". Presumably it is supposed to be:

V(finite) Adverb

(89) Of course it will have to be a particular type of adverb, but let's put that aside. Typically negation is taken to be a prominent "adverb" of this type, e.g. *pas* in French. In many languages negation might appear to the right of the verb although there is no raising of the verb to a left Tense head; the language in fact might have Tense on the right.

(90) More dramatically, consider a mainland Scandinavian language like Swedish. This is an SVO V2 language. But it does not have independent Verb to Tense raising, as witnessed by the NEG V order in embedded clauses (where there is no V2).

(91) So V ADV order in a matrix clauses that arises from V2 is not a signature of independent verb to tense.

(92) Now way to see how both V2 and Verb to Tense parameters will be set via "signatures."

(90) E.g. Japanese or Korean might have a negative morpheme attached to the right of the verb (the verb being in the final position except for its inflections). Shouldn't this constitute a "signature" of Verb to (left) Tense? Wrongly, of course.

(91) Another example of why Tense Adverb is *not* a *signature* for Verb to Tense raising: Korean short form negation

(92) Korean is an SOV language, with the negative morpheme *an* generated on the left of the VP (probably in Specifier of a NegP). (Baek, Hagstrom, among others). The object can raise to the left of *an* (typically thought to be not semantically induced) to yield the typical "short form negation with surface form:

(93) (Subject) Object *an* V, displaying ADV V

(94) So no setting of Verb to Tense as positive on this story. But it is often thought that Korean *does* have Verb to Tense – on right. Setting Verb to Tense depends on other parameters.

(95) We don't see these results in analyses of this model because we haven't seen results of computations in which a *set* of syntactic parameters are learned.

(96) If the set has unambiguous data, this model, like SF, will learn. Without such data? Not so clear.