Parameters and markedness in the acquisition of syntax.

G.Drachman,
University of Salzburg.

1. Four assumptions.

Assume to begin with that every element of a sentence must be accounted for (the 'full interpretation' of Chomsky 1984) or licensed, in acquisition as in the end-state grammar.

Assume secondly that a grammar is not a set of rules, but rather a set of just such licensing principles, again in acquisition as in the end-state. Some examples of principles in this sense are:

The Projection Principle	Theta Theory
X-Bar theory	Control theory
Government theory	Case theory
Binding theory	Subjacency.

Assume thirdly that all principles are, at least outside phylogeny, impenetrable, i.e., not influencable from the outside [cf. Pylyshyn 1980, the notion 'autonomy' in Chomsky generally, and its generalisation in Fodor 1983.

Assume finally that, although the principles are 'impenetrable' as such, certain of these principles have parameters of variation associated with them, again at each stage of acquisition.

Note that while a parameter might become frozen into a principle in the development of the species (the phylogenetic question), we do not suppose that a principle as such may become subject to variation; e.g., we don't expect variants of the Projection Principle or c-command to arise through historical change -- though see Section 4.4 for doubts.

Some examples of parameters are:

- Order of Head-Complement structures
- Adjacency-strength, for government
- What qualifies as a proper governor wrt Extraction from Subject position?
- Bounding Nodes, for Subjacency
- Whether S' Pied-Pipes or not (German & English, vs.Dutch)
- Whether INFL is in S (English) or VP (German)
- Whether Lx has an opaque or a transparent VP (for Theta-Binding)
- Whether the R (the affix-hopping) rule applies optionally in the syntax, or only in the P-component.

Our principled goal is now that, if we activated a process called 'Do anything to (e.g., Insert, delete, coindex, substitute or adjoin) any constituent', the Principles and Parameter-settings should exclude all illicit output-sentences, language for language.

2. Types of Parameter.

2.1. The two kinds of parameter.

A parameter is a dimension of a principle for which overt evidence is available, and on which there might therefore be variation. And we expect a parameter to have some default (or, unmarked) value.

Yet some parameters (e.g., Head-Complement direction) have no default value. We assume that, unless indeed these prove to be complexes of more than one parameter, such parameters are 'open'; and the choice between alternative 'values' (rather than 'plus vs. minus') is language-specific.

Conversely, the principle behind a parameter itself is as I said 'impenetrable' in the sense intended, and can undergo no variation. Of course, the best way to view this distinction is in terms of 'natural laws' (for the principles) and 'conventions' (for the parameters).

2.2. Assigning parameters to their types.

We turn now to the question, which of the putative parameters of grammar are of which type, noting first that parameters naturally involve very diverse material: thus (e.g.,) some parameters delimit Categories (Categories for Wh-movement, the domain of V-max), others have to do with orientation (directionality of government or Theta-role assignment) or adjacency (strictness of adjacency for Case-assignment), and yet others have to do with rule-application levels (Wh-movement, Chomsky's 1981 R-rule). Take a few clear cases first.

Which parameters have default settings that may have to be adjusted? Candidates here are Bounding nodes for Subjacency ; here we might have an example of 'the value on a default parameter' being identified by a set, say NP and S (though cf below Sec.D.2.1). Also, there is adjacency for government/Case assignment (whose default value is 'strict' adjacency); and whether Prepositions govern structurally like verbs or not (here the default value is probably that they do not).

Candidates for truly 'open' parameters might be:

- Directionality of Head+Complements
- Which maximal projections undergo alpha-movement
- Whether Wh-movement obtains in the syntax or only at LF
- Whether the R-rule applies in the syntax or only at PF
- Whether Lx has Subject clitics or not.

Conversely, to recapitulate, candidates for 'true universals' (our 'principles') are c-command, Subjacency as a principle of locality, the Binding principles, May's 1977 Q-rule for adjunction in LF, the Theta Criterion, and the Extended Projection Principle.

2.3. General on Triggers.

2.3.1. Definition.

The principles of grammar are absolute, as we saw; but their dependencies, the default and the open parameters, we defined as sensitive to

the environing language. A trigger in grammar development is a stimulus (group) (a) activating a schema (principle) or (b) setting the value .PA of a parameter. Under (b), a trigger thus allows for either the setting of an open parameter, or the changing or re-setting of a default parameter. Of course, these processes are covert and subsumable under Piercean 'abduction'.

2.3.2. Simplicity?

How simple (conversely, complex) can a trigger be? And on the other hand, how complex a pattern of elements-in-schemata or constraints can a single perhaps very simple trigger release?

In ethology, a trigger may be very simple; eg., in the case of birds the imprinting trigger is the first moving object seen after hatching. On the other hand, a complex sequence of stimuli and events must obtain before the triple-spined stickleback will lay her eggs.

So far as language development is concerned, Chomsky 1982 gives an example of what seems a very simple trigger. He says '..if children get information that something is a reciprocal then that ought to put into play a whole range of constraints as to whether and how it can be interpreted and construed'.

Notice that Chomsky is in effect defending a kind of a 'single-trigger' or 'unified onset' account of the activation of Binding Theory (the relevant constraint here) against Matthei's 1979 claim of piecemeal development. The whole Binding complex, then, depends for Chomsky on as simple a trigger as possible.

And we must perhaps talk, further, of indirect triggering: thus the presence of an otherwise unjustified Resumptive Pronoun in nursery-sentences such as:

'Who you you believe the story that he killed the dragon?'
'Who do you wonder why he killed the dragon?'

in serving to circumvent Subjacency, automatically also triggers the appropriate Bounding nodes.

2.4. Acquisition.

2.4.1. What is acquisition now?

In the present model, language-development does not consist of the cumulative acquisition of diverse rules of grammar, whether of Phrase-structure or Transformational. Rather, it consists very largely in the setting of just those interactive parameters across modules of the grammar. The process is largely data-driven, i.e., it takes place at least partly under the influence of the relevant environmental 'triggers', including heard and attended-to data.

Thus a given putative parameter either a) is 'open', so that a 'first guess' during acquisition may or may not be correct; if incorrect, the guess must be corrected, OR b) has a Universal (or, default) setting; in that case it must be reset only if disconfirmed, i.e., if the environment language has a marked (or, non-default) value for it, OR c) is not a parameter but a

principle; and a principle is simply a law.

2.4.2. Are all the stages of acquisition 'natural'?

It has been held (e.g., in White 1981) that, since the child never contravenes essential properties of language, his grammar will at every developmental stage represent a possible human language. Notice now that, if the above outline is valid, the presence of 'open parameters' perhaps does give early first language acquisition a unique status; to the extent that human languages do not seem to allow parameters to remain 'open', White is wrong. Earliest acquisition represents a partly unnatural language.

Note further that early first languages are also unnatural in a second respect; they are liable to cognitive constraints of a purely developmental kind (cf Rizzi on pro-drop acquisition, under 3.3.1.2. below).

3. Markedness in default parameters.

3.1. Unmarked as [+] or [-].

It is of course not the case that every UG principle has associated with it parameters that must be set [+] or [-] for each grammar. Furthermore, it is not the case that the default values of parameters are randomly assigned, as we seemed to imply above, in assuming that the unmarked value of a default parameter could be either [+] or [-].

Suppose we now assume, perhaps on grounds of economy, that the markedness of default parameters applies homogeneously, i.e., that all parameters have the same default value for the initial state of the acquisition device. There are now two possible scenarios, viz., the one with all Unmarked parameter values [+], the other with them all [-].

3.2. Homogeneous Unmarked values

Scenario 1. Suppose the Unmarked values are all [-]. Consider first the clear cases.

(a) [-] Preposition Stranding, since this implies that Prepositions govern structurally, as verbs do (Hornstein and Weinberg 1981) or that the language licenses reconstruction of V-P so that the Case & Theta-role assigning properties of P are transferred to V (Rouveret & Vergnaud 1980).

(b) [-] Presence of pronominal clitics

(c) [-] Subject pro-drop, thus allowing for the marked use in English Casual speech. This seems to hold equally of the interpretations a) the R-rule may obtain in syntax for pro-drop languages, or b) a head (INFL) may in Lx be licensed to give Case to NP-Subject position.

However, take now Subjacency as a more extended example. In the classical treatments (e.g., Chomsky 1981), there are two major components to Subjacency. The constraint to neighbourhood itself is presumably a law (in the sense here assumed). On the other hand, the so-called Bounding Nodes are parametrised; the possibilities ranging from S' thro S, NP, to PP (but not VP). English supposedly has S,NP as Bounding; while Italian has S' and NP, but

not S.

On the present interpretation, the learning model makes all the onset variables [-], so that none of the Categories is a Bounding Node in early acquisition.

Note that, by contrast to the supposition that the setting of Bounding nodes is achieved en bloc (the 'set' solution in Section 2.2.2. above), we will assume here that this setting applies to each Bounding node individually.

But in the absence of an auxiliary theory, it is impossible to reconcile the Bounding node settings with a default value of [-], because of the implausible implication that young children may freely violate Subjacency.

Scenario 2. Homogeneously, the Unmarked values are [+].

Consider now the case for [+] as the default value for all parameters. Notice the plausibility of this value wrt the problem of the Bounding nodes for Subjacency; for this value reasonably guarantees that no violations can occur. Positive evidence for a revision to a [-] value for, say, S as Bounding node would thus come primarily from the occurrence of sentences otherwise in violation of Subjacency wrt the node S, as in Italian.

But of course the homogeneous application of [+] as the default value in turn leads to contradictions; thus, assuming [+] for Preposition Stranding implies that, say, all French or German beginner language learners will produce such strandings, in fact illicit in their languages, and in fact unattested in early-acquisition studies for those languages.

3.3. 'Natural' default values.

3.3.1. The Subset Principle.

The 'homogeneous markedness' hypothesis having failed, we shall instead try to apply the learning-theoretic 'subset' principle, to the problem of defining markedness for default parameters. Conceptual parallels can indeed be found in the debate of the 70's on rule-ordering in phonology, viz., in work of Sanders (1970) and Koutsoudas, Sanders and Noll (1974). But the 'subset' principle is in its present form due to Berwick 1982 ; in effect, it says 'choose the most constraining grammar possible'.

We may now understand the setting of the various parameters relatively, viz., by interpreting the Subset Principle as follows: the unmarked values must be chosen so that they automatically allow the minimum of outputs. This seems to impose on us the following three-way assignment:

3.3.1.1. [+] defaults.

'Constraint' parameters like the Bounding Nodes for Subjacency must all be set [+] to guarantee minimal outputs; should Lx in fact allow more than these minimal outputs (as, e.g., S' but not S is a Bounding node in Italian), the positive evidence triggers the reversal to [-] for the node S.

Considering the 'destructive' nature of constraints, there might be a parallel to early phonology, where a cumulation of natural processes results in extreme poverty of outputs. Ceteris paribus, this would suggest that all

possibly-cumulative syntactic constraints are automatically active, with all parameters at [+] at acquisition-onset; the Unmarked values for Bounding nodes are thus all 'plus'.

Compare Subject-pro-drop in casual speech, perhaps truly a parallel, as I demonstrated in Drachman 1975. But it may be that only in end-state casual speech does syntax show cumulative 'destruction' of the kind exemplified in developmental phonology; for it is characteristic of beginner speakers that they keep morphemes intact, even at the expense of morphophonemic alternations, and acquire casual-speech rules only later.

Thus we might hold that the supposed parallel with early phonology is spurious, for example insofar as the developmental constraints in phonology and syntax are of quite different kinds. Thus the way to more plentiful phonological outputs lies either in reversing the ordering between feeding-pairs of processes, or in suppressing individual processes; but the parameters of Subjacency only subserve a law in setting its boundary conditions for particular languages. The resetting of a [+] to a [-] might well parallel the Stampean suppression of a destructive process; but on the other hand there seems to be no syntactic analog to re-ordering of processes, a basic characteristic of developmental phonology.

Quite apart from these considerations, there is the matter of "heard and attended-to" triggers; after all, there is hardly a phonological analog to the distinction between 'dominant data-type' vs 'exotic data-type' that we shall invoke immediately below for syntax. Notice in particular that we can for phonology establish whether the stored representation of a given segment is intact, even in the absence of a distinctive output for that segment; consider common cases of the type 'bat' vs. 'bad'

/baet/ --> [baet], but /baed/ --> [bae:t]

with 'displaced' contrast.

3.3.1.2. [-] defaults.

On the other hand, take the parameters representing optional properties, such as that involved in the licensing of Preposition Stranding in English or Object pro-drop as a syntactically active process in Italian, or the presence or absence of clitics in Lx: these must initially be set at [-], so that only positive evidence will activate them.

Note that according to the 'minimal outputs' criterion, we are driven to assuming a default value of [-] for Subject pro-drop, since the [+] value would extend the set of potential outputs. The claim in Hyams 1983 (based on English, Italian and German data) that the default value here must be [+], must necessarily now be reinterpreted. I take now three alternative reinterpretations, each appealing to a different auxiliary hypothesis.

Rizzi's assumption (1986: fn 27, pg.526) concerns the abstract possibility that initial access is constrained by severe working memory limitations that involve the dropping of various grammatical morphemes (including pronouns) from the initial linguistic representations.

By contrast, Hummer 1986, surveying pro-drop data specifically for German, claims that the data are in fact irrelevant to the problem of the default

value for pro-drop; rather, he maintains, configurational relations have not yet come into play at the stage at which 'pro-drop' first arises. Hummer's auxiliary hypothesis is thus that early German shows merely pre-syntactic Topic-loss.

Can we reinterpret the data without reintroducing a pre-syntactic development stage (Cf Marantz 1981)? Since both English and German show Casual speech pro-drop (for English cf Drachman 1975), we might well assume that the data in question simply results from overgeneralisation from Casual Speech. However, as Hummer points out (personal communication), early pro-drop in German even occurs in sentences with Object fronting (i.e., in non-sentence-initial position), while Casual pro-drop does not: thus this third alternative also proves less than secure.

3.3.2. On 'Dominant Data-types'.

An interesting problem arises in connection with the 'open' parameters, viz., that the empirical data seem sometimes to contradict a prediction that follows from our standpoint on constraining the grammar. To illustrate this, compare Preposition Stranding (hereafter P.S.) with Object pro-drop.

Take P.S. first. Since it makes for further outputs, we are bound to say that the default value for P.S. is [-]. Similarly, many languages lack syntactically active Object pro-drop, so its appearance in Italian must be marked; its default value is thus again [-]. But while the prediction for Object pro-drop holds up (viz., children do not produce Object pro-drop without overt inputs) the prediction for P.S. seems to be empirically false, for young children do not (as would be expected) commonly produce questions with Pied Piping of Preposition-Phrases, as in

'In which cupboard did you put my teddy-bear, Mummy?'

But in fact it is unreasonable to expect necessarily to witness the data for the (nevertheless present) unmarked value for P.S. To distinguish the two cases, we introduce the notion 'dominant data-type'. By this we mean that certain data types (e.g., simple questions out of Preposition Phrases) occur in the child's heard and understood input so early and so often that the parameter-value is already set before the relevant outputs are attempted. By contrast, the data for Object pro-drop is so exotic ('this leads to conclude the following', or 'Good music reconciles with oneself') that one predicts a quite late switching of the parameter value, so that early child utterances of Italian children should show the (unmarked) non-application.

4. Doubts on some basic assumptions.

4.1. On negative evidence.

To revert to the possible interpretation of the setting for Subjacency as [-] rather than [+]. Suppose it were [-]. Then to answer the question, why (at the relevant point in development) Subjacency violations do not occur, we might well question the putative principle concerning 'positive evidence only'.

It may be that, at least for problems whose solution is not urgent for the

beginner language-learner, there comes a point at which the continued absence of a certain type of structure is indeed appraised and acted upon (Cf. Chomsky 1981:9, and 16 fn.9). We will thus talk of 'significantly-absent data-type', as a kind of converse to the 'dominant data-type' situation mentioned above. Subjacency might, a-fortiori, be a case in point; we need only suppose that the appraisal has occurred before the point at which the relevant complexity of utterances is otherwise available.

A further indirect form of data relevant to the child's setting of parameters might be the occurrence of 'rescue' strategies, e.g., the otherwise-unjustified insertion of Resumptive pronouns. Thus, consider the relevance for Subjacency of story-teller questions to children like:

'Who do you wonder why she had to praise the emperor's clothes?'
'Who do you believe the story that the giant nearly killed him?'

4.2. On overgeneralisation.

Maybe the child does not necessarily choose the most constraining grammar, as is suggested by the presence of overgeneralisations in each domain; e.g.,

a) in morphology, as with 'went-ed', 'see-d'.
b) in word-semantics, as with 'mommy' applied to any woman.
c) Object pro-drop, sponsored by Subject pro-drop in Greek.
d) Perhaps Casual-speech pro-drop helps to trigger the pro-drop parameter in English acquisition, while Casual-speech Topic-loss does the same for German pro-drop during primary acquisition, as mentioned above.
e) Similarly, SOV order in German acquisition may be partly sponsored by sentences with Modals, including Imperatives, as in :

'Du sollst Dein Wurst essen!' 'Eat your sausage!'

All these candidates for 'overgeneralisation' have perhaps rather varied status. Thus, on the one hand the whole issue is perhaps moot wrt word-semantics. On the other hand, while it is perhaps true that the quantitatively most prominent area of over-generalisation is that of morphology, still, if this proves the case, it calls for an explanation!

4.3. On the supposed independence of parameters.

Where principles, or the values of their parameters intersect there will, just as in the phonology of casual speech, be cumulative effects. And the chances of such interaction being strong is greater if all the variables are contained in one module: Cf. Borer's 'inflectional' model, containing Case-relations, inflectional relations and Theta-role assignment (1984:15).

Further, parameters associated with the same principle might well show hierarchical properties; I think of the relation between NP, PP, S, and S' as potential Bounding Nodes for Subjacency.

However, if the values of parameters associated with different principles correlate rather than apply independently, then one of the thus correlated parameters might prove to be redundant.

4.4. On the status of certain supposed 'laws'.
Doubts about the Absolute (i.e., law-like) status of particular principles have been expressed. For example, cf. 4.4.1-3 below.

4.4.1. C-command.

C-command is parametrised in Chomsky 1981:166, with 'strong' command for trace government, but 'weak' command being relevant to trace binding. Still, one cannot imagine that (e.g.) the functions of the two variants might be reversed in some language.

4.4.2. Projection Principle.

Here we mention Hale and the non-configurational version of the Extended Projection Principle, taken up in Pesetsky 1982 wrt Russian Subjectless sentences (cf. Drachman 1986). Cf here the notion, developed in Rizzi 1986 wrt so-called 'Object-pro', that an argument may be 'missing' if its Theta-role is 'saturated' in the lexicon.

4.4.3. Theta Criterion.

The Theta Criterion (one A-position may acquire (only) one role) is seriously questioned in Jackendoff 1972, a position upheld in his 1986. Cf. Chomsky's proposal (1981:139, fn.14) to disarm this position.

References.

Berwick, R.C. (1982). Locality principles and the acquisition of syntactic knowledge. PhD dissertation, MIT.

Borer, H. (1984). Parametric Syntax: Case Studies in Semitic and Romance Languages. Dordrecht: Foris.

Chomsky, N. (1981). Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, N. (1982). Some Concepts and Consequences of the Theory of Government and Binding. Cambridge: MIT Press.

Chomsky, N. (1984). Changing Perspectives on Knowledge and Use of Language. MS

Drachman, G. (1975). The syntax of Casual Speech. In Salzburger Beitraege zur Linguistik (ed. by G. Drachman), 273-296. Tuebingen: Verlag G.Narr.

Drachman, G. (1986). On Subject Extraction. Paper for the 7th. Annual Linguistics Meeting, University of Thessaloniki, 1986.

Fodor, J. (1983). The Modularity of Mind. Cambridge: MIT Press.

Hornstein, N. and A. Weinberg (1981). Case Theory and Preposition Stranding. Linguistic Inquiry 12.1.55-91.

Hummer, P. (1986). Null-Subjekte in der (deutschen) Kinder-sprache. Paper read at the All-Austria Linguistics Meeting, Klagenfurth. 1986.

Hyams, N. M. (1983). The Acquisition of Parametrised Grammars. PhD dissertation, CUNY.

Jackendoff, R. (1972). Semantic Interpretation in Generative Grammar. Cambridge: MIT Press.

Jackendoff, R. (1986). The Role of Thematic Relations in Linguistic Theory. MS.

Koutsoudas, A., G.A. Sanders, and C.Noll (1974). On the Application of Phonological Rules. Language 50.1-28.

Marantz, A. (1981). On the Nature of Grammatical Relations. Unpublished MIT PhD dissertation.

Matthei, E.H. (1981). Children's Interpretations of Sentences Containing Reciprocals. In Language Acquisition and Linguistic Theory (ed. by S.L.Tavakolian), 97-115. Cambridge: MIT Press.

May, R.C. (1977). The Grammar of Quantification. PhD dissertation, MIT. IULC Bloomington.

Pesetsky, D. (1982). Complementiser-trace Phenomena and the Nominative Island Condition. Linguistic Review. 1.3.297-344.

Pylyshyn, Z.W. (1980). Computation and Cognition: Issues in the Foundation of Cognitive Science. The Behavioral and Brain Sciences 3.1.111-169.

Rizzi, L. (1986). Null Objects in Italian and the Theory of pro. Linguistic Inquiry 17.3.501-557.

Rouveret, A. and J-R. Vergnaud (1980). Specifying Reference to the Subject. Linguistic Inquiry. 11.1.97-202.

Sanders, G.A. (1970). Precedence Relations in Language. Foundations of Language.

White, L. (1980). Grammatical Theory and Language Acquisition. PhD dissertation, McGill University. IULC Bloomington.