What I will be talking about is how I think generative grammar approaches syntactic universals, and I would like to start by saying that I think the topic of linguistic or syntactic universals is actually fairly odd. A legitimate reaction upon mention of this topic could be, what else? That is, basically what we are really interested in is explanation, and not so much in statements like there is something or other, but rather for all X..., such and such happens. That is, laws, or universals.

I think that it is useful to start with an article by a psychologist in the 1930s called Kurt Lewin, who was concerned with scientific explanations in particular and tried to distinguish between two ways of going about thinking about the laws in physics, biology, and other sciences (Lewin 1935). I think that his reflections carry over to cognitive science. In particular, Lewin distinguished between Aristotelian and Galilean explanations. Aristotelian laws or explanations have the following characteristics: they are recurrent, that is statistically significant; they specifically (though not always) target functions, that is they have a functionalist flavor to them; they also allow for exceptions, organized exceptions or not, but at least they allow for exceptions; and finally they have to do with observables of various kinds. Lewin contrasts these sorts of laws or universals with what he calls Galilean laws, which are very different in all respects from Aristotelian laws. In particular, they are typically formal in character, and they are very abstract mathematically. They allow for no exceptions and they are hidden. That is, if you fail to find overtly the manifestation of
a particular law that you happen to study, this does not mean that it is not universal. It just means that it is hidden and that we have to look at it more closely and we will eventually see that the law actually applies.

I think that the contrast between Aristotelian and Galilean laws is very relevant to the study of language because there are various ways of approaching language universals. One of the ways in which you could approach them is like what Joseph Greenberg did with his various arguments on universals. That is not the kind that I am interested in, and it is not the kind of universals that generative grammar really is interested in. The kind of typological universals that Greenberg discovered might be interesting for discovering the type of hidden universals that generative grammar is interested in, but they are not the end of the enterprise. It is worth noting that Greenberg's universals are really surfacing properties of language that typically can be explained in functionalist terms and allow for a variety of exceptions. That is, they are basically tendencies of various sorts, but that is not the kind of thing that generative grammarians have focused on in the past fifty years.

In fact generativists conceived of universals as basically properties of universal grammar (UG). This is the most general definition of universals that I could give, if you ask me what a language universal or linguistic universal (LU) is for a generative grammarian. But that definition actually depends on the specific understanding of UG, and that has been changing for the past 30-35 years. I should say though that no matter how you characterize UG, its content is defined along Galilean lines. We cannot expect universals to be necessarily found on the surface in all languages. That probably is not the case. Conversely, all languages might have a word for yes and no. (I haven’t checked, but say it’s true.) I don’t think we would include this as part of UG, even though it is in all languages. So the understanding of universals that we have as generative grammarians is based on a theory of language that has, as I said, been changing for the past 30-35 years in many ways that do not, I think, make some people very happy as consumers because, to anticipate the conclusion that I will be reaching, the list of universals that we will reach as syntacticians or grammarians will be very refined and abstract, and not directly useful to, for example, the study of language acquisition. We should not be discouraged by that fact. This is a natural result of pursuing a naturalistic approach to language.

What I would like to stress first of all is that the study of syntactic or linguistic universals has run through various stages in generative grammar. In particular, one of the first advances that we were able to make in the understanding of linguistic universals was the distinction that Chomsky (1986b) introduced between I-language and E-language. As soon as you make that distinction, you really have the distinction between I-universals and E-universals. E-univer-
sals are the type of thing that for instance Greenberg universals could be. I-universals would be something like, for example, some deep computational principles of a very abstract sort that are only manifested in very refined and rarified phenomena. It is not something that you can observe by just walking around with a tape recorder or anything of the sort. In fact I think the study of I-universals in this sense started with “Conditions on Transformations” (Chomsky 1973), or if you want, with the discovery of the A-over-A principle – that is, an attempt to try to factor out what the abstract computational principles are, based on a fairly refined empirical view of language. It is true that “Conditions on Transformations” wouldn’t have been possible before Ross’s (1967) investigation of islands. It was only once you reached that very detailed empirical picture that you could try to extract from it this very abstract rule, so Galilean in nature. And so it will be, I think, with other universals.

I think that the stage of the principles and parameters (P&P) approach constitutes a serious attempt to come up with more of those universals, once you have a very good empirical map. That is, once you have attained very good descriptive adequacy, you can try to find and formulate those abstract universals. Things changed, I think, with the advance of the minimalist program, and in particular more recently with the distinction that Hauser, Chomsky, and Fitch (2002) have introduced between the narrow faculty of language (FLN) and the broad faculty of language (FLB). This further distinction basically narrows down the domain of what we take to be language, to be specifically linguistic, and that of course has a direct influence on what we take LU to be. That is, if by LU we mean specific universals for language, then we are going to be looking at a very narrow field, a very narrow set, that is FLN. And there, what we expect to find will be basically abstract general principles such as minimal search, or various refinements of relativized minimality, cyclicity, etc.

Once we reached that stage, then people began to see that perhaps those universals are not specifically linguistic, but might be generic or general principles of efficient computations belonging to third-factor properties, for example. But these would be the kind of LU that may actually be at the core of FLN. Remember that, as Chomsky has discussed recently,¹ there are basically two ways of approaching UG – from above, or from below. And these two approaches will conspire, ideally, in yielding the sources of LU, but for a while we will get a very different picture depending on which perspective we take. Notice, by the way, that if some of these LU are part of third-factor properties, then they may not be genetically encoded, for example. They may be part of general physics or chemical properties, not directly encoded in the

¹ Chomsky (2006).
genome. In this case, the study of LU dissociates itself from genetic nativism (the most common way of understanding the "innateness hypothesis").

The refinements that we have seen in the study of language and LU will force us to reconsider the nature of variation. In this sense, one very good and productive way of studying universals is actually studying variation. Here again, recent advances in the minimalist program have been quite significant because the notion of parameter that we have currently is very different from the notion of parameter that we had, say, in the 1980s. In the 1980s we had a very rich understanding of parameters, including a fair amount of so-called macroparameters of the type that Mark Baker (2001) discussed in his *Atoms of Language*. We no longer have those macroparameters in the theory, simply because we don’t have the principles on which those macroparameters were defined. However, we still have the effects of macroparameters. For example, there is something like a polysynthetic language, but I don’t think we have a polysynthetic parameter, or rather I don’t think we have the room for a polysynthetic macroparameter in FLN. How to accommodate macroparametric effects in a minimalist view of grammar is a challenge for the near future. But it is a positive challenge. That is, maybe this new view of grammar is actually a good one, as I’ll attempt to illustrate through just one example. Take headedness as a parameter. We used to have a very rich structure for P&P, and one of those parameters was basically one that took care of whether complements were to the left or to the right of their heads in a given language. Now the minimalist take on UG no longer has room for such a parameter, but instead tells us that if you have a simple operation like Merge that combines alpha and beta, there are basically two ways in which you can linearize that group (either alpha comes before beta, or after). You must linearize A-B, due to the physical constraints imposed on speech, and there are two ways of doing it. Notice that there you have an effect, since you have a choice between two possibilities depending on the language, but it is no longer the case that we have to look for a parameter in the theory that encodes that. It may just be that by virtue of the physics of speech, once you combine alpha and beta, you have to linearize that set by going one way (alpha before beta) or the other way. I think that this offers new perspectives for studying parameters because LUs are different depending on your theory of language.

Now let me briefly conclude by saying that in a sense, the linguistic progress that we have seen over the past thirty years has taken us closer to a study of LU that is truly Galilean in nature. But that actually should raise a couple of flags, if language is just part of our biological world, and linguistics therefore part of

---

2 As argued below by Luigi Rizzi (see pages 211–219 below).
biology, because biologists are typically, and by tradition, not very interested in universals in the Galilean sense; they are more interested in the Aristotelian kind of universals and tendencies. Gould, Lewontin, and others were fond of noticing two facts about biologists. First, they love details, they love diversity, the same way philologists love details. I certainly don’t like diversity for its own sake. I am interested in general principles and only use the details to the extent that they can inform the study of general principles. Secondly, biologists don’t usually think that there are biological laws of the kind that you find in physics, just because the world of biology is much messier than physics. But here I think linguistics has an advantage, because in a very short history (roughly fifty years) we have been able to isolate invariance amidst diversity, and this is what I was thinking of when discussing I-language vs. E-language, or FLN vs. FLB. One of the things that we have been able to do is make the study of language the study of very simple systems. By narrowing down and deepening our understanding of language we can actually exclude things that belong to details and focus on things where we can discover very deep and comprehensive principles that will be just like what you can find in Galilean laws. That is, they will be exceptionless, abstract, invariant, and hidden.

Janet Dean Fodor

For me, being asked to talk for ten minutes about universals is a bit like being asked to talk for ten minutes on the economy of northern Minnesota in 1890. That is to say, I don’t know much about Minnesota and I don’t know many universals either. But that’s fine, because it allows me to take a very selfish perspective on the subject. I am a psycholinguist and as such it’s not my job to discover linguistic universals, but to consume them. I work on language acquisition, and it is very important when we are trying to understand language acquisition to assess how much children already know when they begin the task of acquiring their target language from their linguistic input. So what matters to me is not just that something is universal, but the idea that if it is universal, it can be innate. And in fact it probably is – how else did it get to be universal? So I will assume here that universals are innate, that they are there at the beginning of the acquisition process, and that they can guide acquisition, increasing its accuracy and its efficiency. Language acquisition is very difficult and needs all

3 I am grateful, as always, to my friend Marcel den Dikken who has exercised some quality control on my claims about syntax in this written version of my round table presentation.
4 For evidence that some innate knowledge becomes accessible only later in child development see Wexler (1999).
the guidance UG can give it.\(^5\) What I will do here is to highlight universals in relation to syntax acquisition. I am going to be walking into the universals store with my shopping bag, and explaining what I would like to buy for my language acquisition model, and why.

A very important point that is often overlooked is that universals (embodied in innate knowledge) play a role not only when learners are trying to find a grammar to fit the sentences they have heard, but at the very moment they perceive an input sentence and assign a mental representation to it. They have to represent it to themselves in some way or other, and it had better be the right way, because if they don’t represent it correctly there is no chance that they will arrive at the correct grammar. So innate knowledge has its first impact on the acquisition process in guiding how children perceive the sentences in the sample of the language they are exposed to. They have to be able to recognize nouns and verbs and phrases and the heads of phrases; they have to know when a constituent has been moved; they have to be able to detect empty categories, even though empty categories (phonologically null elements) are not audible; and so forth. And that is why they need a lot of help, even before they begin constructing a grammar or setting parameters. I want to emphasize that this is so even if acquisition consists in setting parameters. In the P&P model we like to think that an input sentence (a trigger) just switches the relevant parameter to the appropriate value. But for someone who doesn’t know what the linguistic composition and structure of that sentence is, it won’t set any parameters, or it won’t set them right. So if children get their representations right, that’s a very good first step, because it will greatly limit the range of grammars that they need to contemplate as candidates for licensing the input they receive.

Learners need to know what sorts of phenomena to expect – what sorts of elements and patterns they are likely to encounter out there in this language world that is all around them. As one example, consider clitics. Children have to be alert to the possibility that they might bump into a clitic. Imagine a child who has begun to recognize that certain noises compose sentences that contain verbs and objects, and that objects consist of a noun with a possible determiner and that they normally follow (let’s say) the verb, and so on. This child shouldn’t be too amazed if, instead of the direct object she was expecting at the usual place in a sentence, she finds a little morpheme that seems to be attached to the beginning of the verb – in other words, a clitic. Infants need to be pre-prepared for clitics, because if they weren’t it could take them a long time to catch on to what those little morphemes are and how they work. You could imagine a

\(^5\) See Chapter 17 for discussion of how difficult it is to model what small children are doing when they are picking up the syntax of their language.
world of natural languages that didn’t have any clitics, but our world of natural languages does, and infants pick them up extremely early: they are among the earliest things that they get right (Blasco Aznar 2002). So it seems that somehow they are pretuned to clitics, and to the ways in which a clitic might behave. Sometimes a clitic can co-occur with a full DP object (usually it doesn’t, but it can); and there can be indirect object clitics, and locative clitics and reflexive clitics and partitive clitics; and sometimes multiple clitics have to come in a certain order before the verb, and learners should watch out for whether that order is determined by an array of properties that includes person as well as case. None of these differences from phrasal arguments seem to take children by surprise.

However, even more than being ready for what they might encounter in language, children need to have expectations about what they are not going to encounter. This is very important for limiting the vast number of potential hypotheses that they might otherwise entertain. Even in constrained linguistic theories which admit only a finite class of possible grammars, that still amounts to a lot of grammars for children to test against their language sample. We don’t want them to waste their time on hypotheses that could not be true. Let’s consider an example of movement, such as:

(1) Which of the babies at the daycare center shall we teach ASL?

There is a missing (i.e., phonologically null) indirect object between teach and ASL, and an overt indirect object (which of the babies at the daycare center) at the front of the sentence, not in its canonical position. Let’s suppose a learner has put two and two together and has recognized this as a case of movement: the indirect object has moved to the front of the sentence. Now why has it moved to the front? Please imagine that this is the first time that you have ever encountered a sentence with overt movement (you are a very small child), and you think perhaps the phrase was moved because it is a plural phrase, or because it is an animate phrase, or because it is a focus phrase, or because it is a very long phrase – or, maybe, because it is a wh-phrase. Some of these are real possibilities that a learner must take seriously: in Hungarian questions, a wh-phrase is fronted because it is a focus; in Japanese a wh-phrase can be fronted by scrambling, motivated by length or by its relation to prior discourse. But other ideas about what motivated this movement are nothing but a waste of time; an infant without innate assistance from UG might hypothesize them and then would have much work to do later, to establish that they’re incorrect and start hypothesizing again. So it helps a great deal to know in advance what couldn’t be the case. To help us think this through, I’m going to make up my own universal principle: in natural language, there is no such thing as a process of
fronting plural noun phrases. That is to say: a plural noun phrase may happen
to be fronted, but not because it's plural; number is not a motivating factor for
movement. Maybe I'm wrong, but let's pretend for the moment that this is a
guaranteed universal. Then it is good for children to know it, because that
makes one less hypothesis they will have to explore.

Similar points apply at all stages of learning. Imagine now a child who has
correctly hypothesized that the noun phrase in our English example was fronted
qua wh-phrase, not because it is plural, etc. He still needs to know how far he
can generalize from this one instance, how broad he should assume this wh-
fronting phenomenon to be. Do all wh-phrases front in this language? Or is it
only [+ animate] wh-phrases that do, or only non-pronominal wh-phrases, or
wh-phrases with oblique case, etc.? I'll assume here that part of the innate
knowledge that children have is that wh-movement is sometimes sensitive to
case; there are languages in which nominative but not accusative arguments can
move in relative clauses. But I'm supposing that wh-movement is never sensi-
tive to number. So if a child hears a question with a singular fronted wh-phrase,
he can safely assume that it is equally acceptable to have plural fronted wh-
phrases, and vice versa: number is not even a conditioning factor on movement
(at least, on A-bar movement). This is another fact that is very useful to know; it
eliminates another hypothesis the child would otherwise have wasted time on.
Note that it's a quite specific fact. There are other phenomena which are
constrained by number. Obviously, anything involving number agreement is
bound to be, but also some unexpected things. For example, the construction:

(2) How tall a man is John?

has no plural counterpart. You can't say:

(3) *How tall men are John and Bill?

That's not English. Nor is:

(4) *How tall two men are John and Bill?

where it's clear that the movement of how tall isn't vacuous. So there is an odd
little bit of number sensitivity here. A wh-adjunct like how tall can be fronted
within its DP (which is then fronted in the clause), but that process is sensitive,
it seems, to singular vs. plural. There are also phenomena that, unlike wh-
movement, are sensitive to whether a constituent is pronominal. In some

---

6 This is one interpretation of the Keenan–Comrie hierarchy (Keenan and Comrie 1977).
7 Pesetsky (1987) notes that what conditions phenomena such as superiority effects in wh-
constructions is discourse-linking, not pronominality, even though the two may be related.
Scandinavian languages, for example, scrambling treats pronouns differently from non-pronominal elements. So here too, there's specific information that a learner would benefit from knowing in advance.

The general point is that if learners didn't have innate knowledge about which properties can and cannot condition wh-movement or any other linguistic phenomenon, then they would have to check out all the possibilities just in case. Many of you have probably read Steven Pinker's first book on language acquisition.\(^8\) It is a very fat book, because what Steve was trying to do in it was to show how a child would set about checking all the possible hypotheses about which features condition a linguistic phenomenon. One of several examples he worked on was the English double NP dative construction, comparing acceptable and unacceptable instances such as:

(5) I gave Susan the book.

(6) *I donated the library a book.

The second example can only be expressed as *I donated a book to the library*, with a prepositional phrase. Which verbs permit the double NP? It takes an enormous number of pages to explain how the child would check out, one by one, all the possible features and feature combinations that might govern the extent of the double NP pattern. According to what was being proposed at that time, the key features were that the verb had to be monosyllabic (or to be of Germanic, not Romance origin; or to be prosodically one foot), and its semantics had to be such that the indirect object became the possessor of the direct object of the event described in the sentence. Pinker noted that the range of potential constraints on lexical alternations is large and heterogeneous, and you can imagine how far down in the child's priority list this particular combination of constraints would be. Clearly it would take a substantial amount of testing (as Pinker illustrates in detail) to discover which are the properties that matter in any particular case. Worse still: in the absence of innate guidance, a learner could imagine that there might be equally idiosyncratic phonological and semantic conditions on any linguistic pattern observed in the input. There would be no way to find out without trying. To be on the safe side, therefore, the child would have to go through the whole laborious procedure of checking and testing in every case – even for phenomena to which no such conditions apply at all.

Surely this is not what children do. But if they don’t, then it seems they must have advance knowledge of what sorts of conditions might be relevant where (e.g., no language requires the verb of a relative clause to be monosyllabic).

---

I do not know precisely how UG prepares children for acquisition challenges such as these. But that is what I am shopping for. I want to know how UG could alert children in advance to what is likely to happen in their target language, what could happen, and what definitely could not. A learner who overlooked a conditioning feature on a rule would overgeneralize the rule. And it is not just rules that are the problem; the same is true in a parameter-setting system if it offers competing generalizations over the same input examples. Overgeneralization can cause incurable errors for learners who lack systematic negative evidence. It follows that learners should never overlook a conditioning feature. But we have also concluded that they can’t afford to check out every potential feature for every linguistic phenomenon they encounter. Concrete knowledge of what can and cannot happen in natural languages at this level of detail would thus be very valuable indeed for learners. Yet linguists interested in universals and innateness mostly don’t map out facts at this level of detail. Why not? Perhaps just because these undramatic facts are boring compared with bigger generalizations. To be able to propose a broad structural universal is much more exciting. But another reason could be that these facts about what can be relevant where in a grammar don’t seem to qualify as true universals—perhaps not even as parameterized universals unless parameters are more finely cut and numerous than is standardly assumed.9 Therefore it appears that we may need a different concept, an additional concept, of what sorts of linguistic knowledge might be innate in children, over and above truly universal properties of languages. To the extent that there are absolute universals, that’s splendid for acquisition theory; it clearly contributes to explaining how children can converge so rapidly on their target language. No learning is needed at all for fully universal facts. But it may be that there are also “soft” universals; that is, universal tendencies that tolerate exceptions though at a cost. This would be a system of markedness, which gives the child some sort of idea of what to expect in the default case but also indicates what can happen though it is a little less likely, or is a lot less likely, or is very unlikely indeed.

There certainly has been work on syntactic markedness. Noam has written about it in several of his books, including in his discussions of the P&P model,10 but not a great deal of research on markedness has actually been done in this framework.11 We don’t have a well-worked-out system of markedness principles that are agreed on. Some linguists are leery of the whole notion. Markedness can be very slippery as a linguistic concept. What are the criteria for something being marked or unmarked? What sort of evidence for it is valid?

9 See Kayne (1996).
10 Chapter 1 of Chomsky (1981) and chapter 3 of Chomsky (1986b).
11 For discussion of syntactic markedness within Optimality Theory see Bresnan (2000) and references there.
(Is it relevant how many languages have the unmarked form? Is the direction of language change more compelling? Or tolerance of neutralization, or ease of processing, etc.?12) On the other hand, if we could manage to build a markedness theory, it would provide just what is needed to reduce labor costs for learners. It can chart the whole terrain of possible languages, with all potential details prefigured in outline to guide learners’ hypotheses. Perhaps this is extreme, but my picture is that all of the things that can happen in a natural language are mapped out innately, either as absolute principles with parameters, or with built-in markedness scales that represent in quite fine detail the ways in which languages can differ.13 What learners have to do is to find out how far out their target language is on each of the various markedness scales. They start at the default end, of course, and if they find that that isn’t adequate for their language sample they shift outward to a more marked position that does fit the facts.14

To illustrate how this would work, let’s consider which verbs are most likely to bridge long-distance extraction, such as wh-movement out of a subordinate clause. In some languages no verbs do: there is no long-distance extraction at all. In languages that do have long-distance extraction, the bridge verbs will certainly include verbs like say and think. English allows movement of a wh-element over the verb say in an example like:

(7) Who did you say that Mary was waving to?

In some languages, such as Polish, that’s about as far as it goes; there is movement across say but not across consider or imagine. In English the latter are acceptable bridge verbs, and perhaps also regret, but we draw the line at resent and mumble. It seems that there is a universal list of more-likely and less-likely bridge verbs, and different languages choose different stopping points along it – although we may hope that it is not a mere list, but reflects a coherent semantic or focus-theoretic scale of some sort.15 If children were innately equipped with this scale, Polish learners could acquire extraction over say without overgeneralizing it to imagine, and English learners could acquire extraction over say and imagine without overgeneralizing it to resent. A different scale seems to control which verbs permit the passive. It’s not the same set

---

12 See Chapter 1 of Klein (1993).
13 Chomsky (1981: 8) writes: “outside the domain of core grammar we do not expect to find chaos. Marked structures have to be learned on the basis of slender evidence too, so there should be further structure to the system outside of core grammar. We might expect that the structure of these further systems relates to the theory of core grammar by such devices as relaxing certain conditions of core grammar...”.
14 See the “tidemark” model in Fodor (1992).
15 See Erteschik-Shir (1997).
in every language, but it also doesn't differ arbitrarily. In all languages the verbs most likely to passivize are action verbs like *push* or *kill*. Languages differ with respect to whether they can passivize perception verbs. We can do so in English, for example:

(8) The boy was seen by the policeman

but many languages cannot; perception verbs are evidently further out than action verbs on the markedness scale for passive. Further out still are verbs of possession and spatial relation. Another example concerns the contexts in which binding-principle exceptions are possible, such as local binding of pronouns. This is extremely unlikely in direct object position, but less unlikely for oblique arguments of the verb; the more oblique an argument is, the less tightly the binding theory seems to hold. Thus a learner can fairly safely ignore the possibility of binding exceptions in some contexts, and yet know to keep an eye out for them in other contexts.¹⁶

My conclusion is that if we insist on absolute universals only, we will forgo a great deal of wisdom that all of us possess, as linguists, concerning the “personality” of natural language. We have to assume, I think, that children have that knowledge too, because otherwise they couldn’t do the formidable job they do in acquiring their language. So here is my plea, my consumer’s request to the “pure” (theoretical and descriptive) linguists who work on universals: Please tell us everything that is known about the sorts of patterns that recur in natural languages, even if it is unexciting, even if it is squishy rather than absolute, even if it has the “scalar” quality that I’ve suggested, so that we can pack it all into our learning models. They will work a whole lot better if we can do that. If we bring these facts out into the open, not just the rather small number of absolute universals, and the parameters that allow for broad strokes of cross-language variation, but all the many partial and minor trends, we will thereby strengthen the innateness hypothesis for language acquisition. I should add one comment on that last point, however. For my purposes, my selfish consumer purposes, it doesn’t matter at all whether the universal trends are specific to language or whether they are general cognitive tendencies. They may be narrowly language-bound in origin, or very general psychological or biological propensities. It would be of great interest to know which is the case. Certainly we should look to see whether some of the curious trends I have cited can be derived from more general underpinnings, linguistic or otherwise. But as long as they exist, whatever their source, they will do what’s needed for psycholinguistics to explain why it doesn’t take a child a lifetime to learn a language.

¹⁶ See J. D. Fodor (2001).
Lila Gleitman

I would like to back up a little and point the conversation toward the case of the child learning the meaning of a word—a theme which came up in Noam Chomsky’s discussion earlier in this conference, and also, in a very different way, in Wolfram Hinzen’s talk about arguments and adjuncts. Here’s the problem. It’s obvious that in deciding on the meaning of a new word, we rely at least in part on the extralinguistic situation, the context in which the word is being uttered. What’s obvious, though, is only that this is so. What is not obvious and, rather, lies almost altogether beyond our current understanding is how this is so, or even how it could be so. The information that children—or any learners—get from the world about the meaning of a new word is often flimsy, certainly variable, and not infrequently downright misleading. This is perhaps most poignant in the case of verbs and their licensed argument structures. I got interested in this problem about thirty years ago when Barbara Landau and I studied language acquisition in a congenitally blind child (Landau and Gleitman 1985). We were very startled to discover that the first verb in this child’s vocabulary, at two years old or maybe even slightly younger, was see, and her usage seemed much like our own from the start, referring to a perceptual accomplishment. That is, this child never seemed to have confused look or see with touch, even though, given her perceptual capacities, she herself necessarily touched as a condition for seeing. This case dramatizes the fact that while it is true that situational context commonly fits the intended interpretation, most of the explanatory burden for understanding learning rests on the infant’s ability to represent that context “in the right way.” In this instance, the contexts of the teacher/speaker (the sighted adult community) and the learner aren’t even the same ones. In this brief discussion I want to illustrate the issues by showing you some findings from Peter Gordon (2003) demonstrating prelinguistic infants’ remarkable capacities and inclinations in regard to the meaningful interpretation of events.

In Gordon’s experiments, infants of about 10 months of age (who as yet utter no words) are shown videos depicting what to adults would be giving or hugging events. In the former case, a boy and a girl are shown approaching each other; one hands a stuffed bear to the other, and then they part. In the latter video, the two approach each other, embrace, and then part. The clever part of this manipulation is that in the hugging scene as well as in the giving scene one of the two actors is holding the stuffed bear. So crucially there are three entities involved in a motion event, in both cases. The only difference between the two events is that only in the give scene is this toy transferred from one participant’s

17 See Chapters 2 and 9 above.
Fig. 14.1. Habituation effects for argument versus adjunct: This figure graphs habituation in infants who are watching either a scene depicting giving or hugging (panel a). When a toy animal that one character is carrying is subsequently removed from the video, dishabituation is observed for the giving video but not for the hugging video (panel b).

Source: Courtesy of P. Gordon, 2003

grasp to the other's. Gordon recycled these videos so that infants saw them again and again, leading to habituation (measured as the infant spending less and less time looking at the video at all, but rather turning away). Any individual baby in this experiment saw only the give scene or only the hug scene. Once babies were habituated, they viewed new scenes that were identical to the originals except that the toy was now absent.

As you see in Fig. 14.1, babies dishabituated (started visually attending again) in response to the new (toyless) give scenes but not to the new (toyless) hug scenes. Gordon also tracked the babies' eye movements to various scene elements during the course of the events. What is shown in the next two Figures is the proportion of time that the babies visually attended to the three entities – the boy, the girl, the toy – as the event unfolded in time, specifically, before, during, and after the two actors interacted.

For the give scene (Fig. 14.2) visual attention is heavily attracted to the toy as the actors encounter each other; and when the toy is removed the infants persist in looking at the actors' hands – where the toy used to be – as though searching for it. In contrast, they did not seem to notice the toy very much when it was there in the hug scene, as Fig. 14.3 shows.

No more did they seem to notice when it magically disappeared. That is, they hardly looked toward the hand of the hugger who previously had held it, nor provided other measurable signs that they were thinking, “Whatever happened
Eye Tracking for GIVE video

![Graphs showing visual attention to argument change](image)

Fig. 14.2. Visual attention to argument change: This figure shows eye-tracking records for infants to the toy animal in the give scene as the characters approach, contact each other, and depart (panel 1) and the persistence or enhancement of visual attention when the toy (that which is given) subsequently disappears (panel 2).

Source: Courtesy of P. Gordon, 2003

Eye Tracking for HUG video

![Graphs showing visual attention to adjunct change](image)

Fig. 14.3. Visual attention to adjunct change: Visual attention is diffuse across the characters in the hug scene (panel 1) but shifts to the hugger (the boy) and huggee (the girl) when the toy disappears. The toy itself is largely ignored (panel 2).

Source: Courtesy of P. Gordon, 2003

to that delightful stuffed animal?" Apparently, the babies' implicit supposition was that, even though stuffed bears are of great interest in everyday life, hugging events are not "relevantly" changed as a function of whether one of the huggers is holding one of them during the performance of this act. But an act of giving is demolished if the potential gift does not change hands. Bears are no more than adjuncts to hugging but they can be arguments of giving.

In one sense these charming findings are unsurprising. Of course it would have to be the case that infants could recognize these entities and represent their roles differently as a condition for acquiring hug and give. But we are very much lacking in any detailed knowledge of the conditions or procedures that underlie
evocation of these representations for the sake of word learning. How does an infant—or for that matter an adult—select relevant representations from those made available by inspection of the world that accompanies speech acts? I believe that many developmental psychologists breezily beg or at least trivialize the questions and puzzles here by suggesting that word learning is at bottom demystified merely by alluding to the reference world. Of course it is right that in significantly many cases there is plenty of information around. The issue that Noam Chomsky has sometimes termed the “poverty of the stimulus” problem isn’t always, or perhaps even usually, that there isn’t any potential information. On the contrary, the problem is usually that there’s enough information to drown in—sometimes I have even called this the “richness of the stimulus” problem. To understand word learning at all we have to get a lot more specific about how the relevance problem in word learning is solved with such laser-like accuracy by mere babes. To return to the present example, how does one know enough to ignore a bear held aloft while hugging?18

Some useful directions of research, inspired by Gordon’s work, try to extend and generalize his procedures for older children and adults by using a change-blindness paradigm. Notice in Fig. 14.4, which shows three temporal points

---

18 In Chapter 16 I discuss some first steps that I and many colleagues have tried to take in these regards.
within events, that the animal changes into another at the time of interaction. Pilot findings suggest that this change is more noticeable for giving than for hugging (Trueswell et al., in progress).

More generally, observation of the reference world, while informative for word learning, seems hardly ever to be sufficient unless the category encoded is of a basic-level object (cf. Rosch 1978). In other cases, a mosaic of conspiring cues – each of them inadequate or even obfuscating by itself – from the situation and from the surrounding speech event are exploited by learners young and old to converge almost errorlessly on the lexicon of the native tongue.

Language invariance and variation

Luigi Rizzi

In this short presentation, I would like to focus on how linguists deal with the problem of invariance and variation in natural language. If you describe and compare languages, you observe that some properties are constant and other properties vary across languages. Then the question is how we can express what is universal and what are the observed patterns of variation. The theoretical entities that are used to address this issue are the concepts of Universal Grammar and particular grammars. These concepts have undergone significant development in the last twenty-five years or so. Let us briefly go through these developments. The “traditional” approach for me, the one that I studied when I first entered the field, is the Extended Standard Theory of the early and mid-seventies. The approach is really focused on the concept of particular grammar. A particular grammar is a set of precise formal rules that are related to constructions. So the particular grammar of English, for example, is a set of rules about the form of, let’s say, active sentences, passive sentences, questions, imperatives, relatives, and so on. This set of rules somehow represents, in an intrinsic manner, the knowledge of the language that the speaker has intuitively. In addition to particular grammars there is a general entity, Universal Grammar (UG), which in the framework of Extended Standard Theory would be considered a kind of grammar metatheory: if a particular grammar is a theory of a language, UG is a theory of the theory of the language. So UG specified, in this way of looking at things, the format of grammatical rules – that is, what the ingredients are that you may expect to find in the rules of specific languages. And then there were certain general conditions on rule application, like Chomsky’s A-over-A Principle, principles expressing empirical generalizations like Island Constraints, and so forth.

There was a theory of language acquisition that went with this framework, more or less explicitly, according to which the language acquisition process is
actually a process of rule induction. That is to say, the child, equipped with the notions of UG, has to figure out on the basis of experience what the properties are of the particular rule system pertaining to the language he is exposed to. So there is a process of rule induction, the determination of a particular rule system on the basis of experience.

There were a number of problems with this approach. One had to do with the difficulty of basing comparative syntax on this way of looking at things. What happened was that linguists would write a formal grammar concerning a particular language, and then when they started analyzing the next language, basically they had to start from scratch and write another system of rules that was in part related to the previous one, but it was truly difficult to pull out the properties that the two systems had in common. That was something that I experienced very directly because my first attempt to do syntactic research was basically to adapt to Italian what Richard Kayne had done about French. I came up with a system of formal rules for certain Italian constructions that had a sort of family resemblance to the rules that Kayne had proposed for French, but it was really hard to factor out the common properties (Kayne 1975).

Then, one major problem with this approach had to do with the acquisition model, because there weren't clear ideas on how rule induction would work.

Things changed around the late 1970s with Chomsky’s lectures in Pisa (Chomsky 1981), which gave rise to his 1981 book *Lectures on Government and Binding*, articulating the principles and parameters approach, based on very different ideas. The key notion really became UG, which was construed as an integral component of particular grammars: UG was conceived of as a system of principles which contain some parameters, some choice points expressing the possible cross-linguistic variation; particular grammars could be seen as UG with parameters fixed or set in particular ways. This went with a particular model of language acquisition. Acquiring a language meant essentially setting the parameters on the basis of experience. This is not a trivial task, as a number of people including Janet Fodor, for instance, have observed. In a number of cases the evidence available to a child may be ambiguous between different parametric values, there are complex interactions between parameters, etc. Still, in spite of such problems, parameter setting is a much more workable concept than the obscure notion of rule induction was. And so language acquisition studies blossomed once this model was introduced, and modern comparative syntax really started. For the first time there was a technical

---

19 On the origins of parameter theory see also Baker (2001), and the introductory chapter of Chomsky (2003).
language that could be used to express in a concise and precise way what languages have in common and where languages differ.

Let me just mention for our non-linguist friends a couple of examples. One fundamental parameter has to do with basic word order properties. In some languages, VO languages, the verb precedes the object, as in English, for example, *love Mary*, or in French *aime Marie*. Other languages have OV, Object Verb order: Latin is one case, Japanese is another. If we are to deal with these properties we need at least a principle and a parameter. The principle is Merge, the fundamental structure-building procedure:

(1) Merge: \( \ldots A \ldots B \ldots \rightarrow [A B] \)

It basically says “take two elements, A and B, string them together, and you will have formed a new linguistic entity, [AB] in this case.” But then we need some kind of parameter to account for the difference between, let’s say, English and Japanese, having to do with linear order. In some languages the head (the verb) precedes the complement, while in other languages the head follows the complement:

(2) Head precedes/follows complement

This simple ordering parameter has pervasive consequences in languages which consistently order heads and complements one way or the other. So, two examples like the English sentence (3a) and its Japanese counterpart (3b) differ rather dramatically in order and structure, as illustrated by the two trees (4a) and (4b):

(3) a. John has said that Mary can meet Bill
    b. John-wa [Mary-ga Bill-ni a-cru- to] itte-aru
       John-TOP [Mary-NOM Bill-DAT meet-can- that] said-has

English expressions have a fundamentally right-branching structure, Japanese expressions a fundamentally left-branching structure, not the perfect mirror image because certain ordering properties (such as the order subject–predicate) remain constant, but almost the mirror image.

We have broad parameters of this sort, having to do with the ways in which Merge works, and parameters on the other basic operations. The other fundamental operation is Move, so there are parameters on movement. Some languages have properties like Verb Second having to do with the fact that the inflected verb always occupies the second position (German, for instance, has this property), and the parameter basically amounts to the fact that there are two slots in the left periphery of these languages which must be filled by movement, one by the inflected verb and the other by any constituent. A third
(4) a.  
```
(4) a.  
N  T
John  
T  V
has  
V  C
said
C  T
that
N  T
Mary  
T  V
can  
V  N
meet  Bill
```

b.  
```
(4) b.  
N  T
John-wa  
V  T
C  V  -aru
itte-
T  C
to
N  T
Mary-ga  
V  T
-eru-
N  V
Bill-ni a-
```
kind of parameter has to do with Spell-out. There are certain elements that can
or must be left unpronounced in particular configurations in some languages.
One classical case is the Null Subject parameter: subject pronouns can be left
unpronounced in languages like Italian, Spanish, etc. You can say things like
parlo italiano (‘(l) speak Italian’) for instance, and this property relates in a non-
trivial manner to other properties of the language (Rizzi 1982 and much
subsequent work).

So the question that arose at some point, after a few years of development of
these ideas, was how to express the format of these parameters. Is it the case
that anything can be parameterized, or is there a specific locus for parameters?
The first idea on the locus for parameters was that parameters were expressed
directly in the structure of principles. This was probably suggested by the fact
that the first parameter that was discussed in the late seventies had to do with
a particular locality principle, Subjacency, the parameterization involving the
choice of the nodes that would count as bounding nodes, or barriers for locality
(the S/S' parameter) (Rizzi 1978). On the basis of this case, it was assumed for
some time that maybe parameters were generally expressed in principles, and
that could be the general format. Among other things, this assumption gave a
certain idea on the important question of how many parameters one should
expect in UG. As the UG principles were assumed to be reduced in number,
if parameters were expressed in the structure of principles one could expect an
equally reduced number of parameters.

This view was abandoned fairly quickly, for a number of reasons. One
reason was that some principles turned out not to be parameterized. There
are certain things that don’t vary at all, certain principles do not allow for any
sort of variation. In no language, as far as we know, does a structure like the
following

(5) He thinks that John is crazy

allow for coreference between He and John (principle C of the Binding Theory).
That seems to be a general, invariable property of referential dependencies, and
many other principles seemed to work like that.

The second reason was that some macroparameters, big parameters initially
assumed to characterize basic cross-linguistic differences, turned out to require
reanalysis into clusters of smaller parameters. One case in point was the so-
called Configurationality parameter. Some languages have a much freer word
order than other languages. Originally it was thought that there was a major
parameter dividing languages with free word order vs. languages without free
word order, essentially. But it quickly turned out that there are different degrees
of free word order: some languages are freer in the positioning of the subjects,
others are freer in the reordering of the complements (scrambling), etc. You have a continuum—not in a technical sense, but in the informal sense that there are different degrees of freedom, so that the big "non-configurationality" parameters really needed to be divided into smaller parameters.

The third reason was that some parametric values turned out to be intimately related to specific lexical items. For instance, consider the Long-Distance Anaphor parameter—the fact that certain reflexives roughly corresponding to English *himself* in some languages allow for an antecedent that is not in the same local clause (in Icelandic, for example). This turned out to be the specific property of certain lexical items: if the language has such special lexical items, that is, anaphors of a certain kind, then these anaphors work long-distance. So, we are not looking at a global property of the grammatical system, but simply at the presence or absence of a certain kind of item in the lexicon. These considerations led to the general view that parameters are not specified in the structure of principles, but rather are properties specified in the lexicon of the language. In fact, assuming the fundamental distinction between the contentive lexicon (nouns, verbs, adjectives, elements endowed with descriptive content), and the functional lexicon (determiners, tense, mood, aspect specifications, auxiliaries, complementizers, etc.), parameters could be seen as specifications in the functional lexicon. So, a reasonable format for parameters would look like the following:

(6) \( H \) has \( F \)

where \( H \) is a functional head, and \( F \) is a feature determining the possibility of one of the major operations, either Merge or Move or Spell-out, essentially. This is the general format of parameters that seems to be justified. This view implies important differences with the view expressing the parameters in the principles. For instance, the order of magnitude of parameters is now related not to the number of principles, but to the size of the functional lexicon.

If you take certain approaches, like the cartographic approach (Belletti 2004; Cinque 1999, 2002; Rizzi 2004), assuming very rich functional structures, the implication is that there can be a very rich system of parameters. Much recent work on the cartography of the left periphery of the clause has led to the identification of a rich system of functional heads corresponding to the \( C \) (complementizer) domain, a system delimited by Force and Finiteness and hosting positions for Focus, different kinds of Topics, preposed adverbials, operators for the various \( A' \) constructions, etc. (see various papers in Belletti 2004, Rizzi 2004). And the cartography of the IP structure has uncovered a very detailed functional system for the clausal structure, with dedicated heads of Modality, Mood, Tense, Aspect, and Voice; similar conclusions hold for the
structure of major phrases, DPs, etc. (Cinque 1999 and various references in Belletti 2004 and Rizzi 2004). Putting together the theory of parameters, some minimalist assumptions on linguistic computations, and cartography, we end up with something like the following typology of parameters:

(7) For H a functional head, H has F, where F is a feature determining H’s properties with respect to the major computational processes of Merge, Move, and Spell-out. For instance:

Merge parameters:  
- what category does H select?  
- to the left or to the right?

Move parameters:  
- does H attract a lower head?  
- does H attract a lower phrase to its Spec?

Spell-out parameters:  
- is H overt or null?  
- does H license a null dependent?

So we have parameters determining the capacity of a functional head to undergo merge: what categories does it select; and does it take complements to the left or to the right?20 And perhaps even more fundamental properties, such as: does the language use that particular functional head? It may be the case that (certain) heads of the cartographic hierarchy may be “turned on” or “turned off” in particular languages.

Then we have Move parameters. Heads function as attractors: they may attract a lower head which incorporates into the attractor, or a phrase which moves to the attractor’s specifier. So, does the tense marker attract the lexical verb, as it does in the Romance languages but not in English or most varieties of Continental Scandinavian? Does a head of the complementizer system attract the inflected verb, as in V-2 languages? And does the head attract some phrase to its specifier position, as the C head in V-2?

And then we have Spell-out parameters, having to do with the phonetic realization of the elements involved. Is a particular head overt or not? For instance, the topic head is realized in some languages (one particular use of Japanese wa seems to be analyzable along these lines), but not in others (e.g., in Romance Clitic Left Dislocation). And does a head license null dependents? For instance, does the verbal inflection license a null subject? That is one of a number of possible ways of looking at the null subject parameter in current terms.

This is the general picture that many people assume at present. Now, as there are many more parameters than we originally thought, it turns out that the different parametric choices will enter into various complex kinds of interactions,

---

20 In the approach of Kayne (1994), the head-complement ordering property is in fact restated as a movement parameter.
generating many possible configurations of properties, so that the superficial
diversity to be expected is great. Nevertheless, the deductive interactions between
principles and parameters still are quite tight, so that there are many logical
possibilities that are excluded even in a system which has a richer parametric
specification of the kind I am describing.

I would like to conclude with a brief discussion of the reanalysis that
Guglielmo Cinque (2005) proposed of one of the universals that Joseph Green-
berg (1963) had identified in his very important work in the sixties. Greenberg
had observed that if you look at a variety of languages, you notice that certain
elements that enter into the structure of the nominal expressions can vary in
order, although there are limits to order variation. If we limit our attention to
cases in which the noun is either at the beginning or at the end of the string of
modifiers, we basically find three types. One type is realized by English and by
the Germanic languages in general, where the order is demonstrative, numeral,
adjective, noun (Dem Num Adj N) giving something like:

(8) these three nice books

One also finds quite a few other languages in which the order is the mirror
image: N Adj Num Dem. Thai has that property, so a noun phrase in Thai has
the order

(9) books nice three these

– an exact mirror image to English. Then, by restricting our attention to cases in
which the noun is either final or initial, a third case that is found, instantiated by
the African language Kikuyu, is N Dem Num Adj, like English except for the
fact that N is at the beginning of the string:

(10) books these three nice

Apparently, we never find the fourth logical possibility given this pattern, that is to
say, a language which would be like Thai, with a mirror-image order of adjective,
numeral, and demonstrative, but with the noun in final position (*Adj Num Dem
N):

(11) *Nice three these books

Now Guglielmo Cinque (2005) has shown that this systematic gap can be derived
from very reasonable computational principles. Just in a very simplified manner,
what we can say is that we can take the Germanic order as the basic order. So (8) –
demonstrative, numeral, adjective, noun – is the initial, first-merge order. Other
orders can then be derived by Move, but movement is always driven by movement
of the noun, so that the noun may move alone, and then you get a structure like
(10), with the same order of elements as in English except that the noun has moved stepwise to initial position. Or you have another possible instance of movement, which some linguists have called Snowballing Movement. The noun moves stepwise, but at each step it pied-pipes the whole structure it has moved to, a procedure which ends up producing the mirror-image effect. In this case, you start with something like the English order, you move the noun to the left of the adjective, and now you take the newly-created constituent, noun plus adjective, to the specifier of the numeral, and so on. If you repeat this movement a number of times, you obtain the exact mirror image of the Germanic order. But there are no other possibilities. Particularly, one cannot get the order in (11) because the noun is in final position in this case, which indicates that the noun has not moved, but noun movement is the engine of the whole process, so in the absence of noun movement the order cannot be subverted. In this case there is simply no way to get the reversal of the order with respect to the basic order. Cinque shows that the gap observed by Greenberg is not an exception, it follows from reasonable principles of linguistic computation. Following this model, it may be possible to give principled explanations to much important empirical work within the typological tradition.

In conclusion: there are more parameters than previously assumed, because parameters are properties of functional heads, and the inventory of functional heads is rich, particularly if the cartographic view is correct. Still, deductive interactions between principles and parameters are tight, and therefore the attested patterns of variation are only a fraction of the logical possibilities.

**General Discussion**

**HIGGINBOTHAM:** In relation to Luigi's point (after Cinque), you can easily derive the fact that you can say *these three nice books* but not *books nice three these* just from compositionality – you know, just from a hierarchy. It's not clear to me that we need anything else.

**CHOMSKY:** Part of the sequence just comes, independently of precedence and c-command, from the composition (presumably D and NP). So the D is going to remain outside anyhow, and then what is left is just the relation between *three* and *nice*. And here there seems like a fairly clear semantic property. I mean, *nice books* are a kind of books, but *three books* aren't a kind of books. There is an old paper by Tom Bever from years ago on adjectives,\(^{21}\) where he tried to argue, with some plausibility, I think, that there is a kind of squishiness in adjectives and some of them are more noun-like. For instance, *red*...
can be a color, whereas *nice* can’t be a something, and he argued that the more noun-like ones tend to be closer to the noun. So these kinds of considerations could be the answer to the *three nice* order, in which case you’d get the ordering.

RIZZI: Okay, so suppose you can derive the hierarchy from the needs of semantic compositionality and some related factors, as Jim and Noam suggest. This gives the Germanic order *These three nice books*. What about the other permissible orders? And the impossible one? Take the mirror image order *Books nice three these*: this could also be a direct reflection of compositionality on external merge, but here the syntactic assumptions you make become crucial. Suppose we adopt Kayne’s antisymmetry, which rules out a structure like [[[books] nice] three] these]: then, within Kayne’s system there must be a computational procedure (snowballing movement) deriving this order from the basic order. Consider now the order *Books these three nice*: here, basically under anybody’s assumptions, you need movement of N (or NP) to derive this particular order. And then you must make sure that the movement computation, which is needed anyhow, does not overgenerate, and can’t give rise to the unattested order *Adj Num Dem N*, a fact that Cinque plausibly tries to derive from the assumption that only N can move in this configuration (possibly pied-piping some other material), so if N doesn’t move, there is no way to alter the basic order Dem Num Adj N. So, Cinque’s point is that under reasonable assumptions on the fundamental hierarchy of projections in nominal expressions and on possible movement processes, one can derive the typological facts. This approach looks very plausible to me.

Then the question arises which is raised by your remarks: where does the initial hierarchy come from? Here I think it is entirely plausible that the hierarchy is grounded in semantics, that the requirements of compositional semantics impose certain orders and are inconsistent with others. The cartographic endeavor tries to determine what the functional hierarchies are for different kinds of expressions across languages, what varies and what remains constant. As far as I can tell, this is fully compatible with the attempt to trace the observed hierarchies to the interpretive considerations raised by Chomsky and Higginbotham. In fact, in my opinion, the cartographic projects and results invite such efforts to provide “further explanations” in terms of interface requirement.