
doi: [http://dx.doi.org/10.1515/TLR.2006.004](http://dx.doi.org/10.1515/TLR.2006.004)

Collection Permanent Link: [http://hdl.handle.net/10822/707727](http://hdl.handle.net/10822/707727)

© 2006 Walter de Gruyter GmbH.

This material is made available online with the permission of the author, and in accordance with publisher policies. No further reproduction or distribution of this copy is permitted by electronic transmission or any other means.
Minimizing government: Deletion as cliticization

DAVID LIGHTFOOT

Abstract

The rigors of the Minimalist Program eliminate movement operations, traces, and 1980s-style conditions of government. This article emphasizes poverty-of-stimulus problems, reviews lexical government effects, and captures them (and more) by treating certain deletion as cliticization. This analysis carves up the grammatical world differently, linking former "movement" conditions with phonological reductions, VP ellipsis, gapped verbs, some agreement relations, etc. The productivity of the analysis supports the rigors of the Minimalist Program.

1. Introduction

1.1. Three phases

Throughout the fifty-year history of generative grammar, grammatical theory and innateness claims have been keyed almost entirely to language acquisition. Linguists have been impressed with poverty-of-stimulus problems: children come to have a system (a "grammar") which ranges over an infinitude of data and is not fully determined by their initial experiences. Consequently we
have postulated that they must have certain information independently of experience, available generally to the species innately and enabling them to develop the kinds of grammars they have on the basis of their rudimentary experience in the first few years of life. Linguists have worked with the explanatory schema of (1).

(1)  
   a. Primary Linguistic Data (UG → grammar)  
   b. Triggering experience (genotype → phenotype)

Children are exposed to primary data. The primary data are the kinds of things that any child could be expected to hear; they do not include negative data about what does not occur, nor information about paraphrase relations, entailments, and certainly not comparative data from other languages. That kind of “secondary data” is available to linguists and features in books, articles and argumentation generally, but is not available to young children. On the basis of exposure only to primary data and endowed with certain properties encoded somehow in the genetic material (Universal Grammar), children acquire a mature grammar, part of their phenotype, which varies depending on whether their initial experiences were in Tokyo or Toronto. This grammar is their language organ, in the terminology of Anderson and Lightfoot (2002), and it develops in accordance with the prescriptions of UG.

One can identify three phases in the kinds of claims made about the fundamental nature of UG. In the period of roughly 1955-1970, much energy was devoted to increasing the expressive power of grammars beyond that of the then-familiar phrase structure grammars, so that they might be adequate for the analysis of natural languages. UG was taken to provide machinery for children acquiring their language system and the machinery incorporated distinct levels of representation and derivational operations mapping one to another (the key claim of Chomsky 1957) and a lexicon (the major technical innovation of Chomsky 1965). Then in the period from the late 1960s to roughly 1990, the focus was on developing constraints on derivations, beginning with the A-over-A constraint of the early 1960s, through the conditions of Ross (1967), Chomsky (1973), and then the development of Government and Binding models and parameters of variation. Throughout that period, Ockham’s methodological principle (entities may not be multiplied beyond necessity) minimized the elements invoked in analyses, but after about 1990 substantive economy principles were introduced under the Minimalist Program (Chomsky 1995) and elements of grammars needed to be motivated by interface requirements, as we see below. All of this influences the form of the innateness claims and they have differed significantly in each of these three periods.

Here I don Minimalist spectacles to review the 1980s notion of government, arguing that it can be dispensed with. I propose one simple principle of UG,
which interacts with seven specific devices in the grammars of English speakers; that interaction yields a complex range of phenomena, which is shown to have a simple explanation. The principle of UG is that to delete something is to cliticize it. That is how certain deletion works and it is specified innately. The effects of government and much more follow from this simple idea.

The analysis requires more extensive discussion than a journal article allows, but this brief treatment suffices to make a discussable case while a monograph is prepared.

1.2. Poverty of stimulus

Before we examine the economies of the Minimalist Program, let us briefly consider the logic of the enterprise, often not sufficiently attended to, and isolate a simple, now classic poverty-of-stimulus problem. Children hear forms like (2a) with the verb sometimes pronounced in its full form *is* and sometimes reduced to *’s*. Hence they might deduce that there is a simple operation reducing the verb (2b). Likewise for other verbal forms (2c).

(2) a. Kim *is’*s happy
   b. *is* → *’s*
   c. am, are, has, have, had, will, would, shall
   d. Kim is happier than Tim *is*
   e. I wonder what the problem *is* (cf. The problem’s twofold)
   f. I wonder what that *is* up there (cf. That’s a fan up there)
   g. I wonder where the concert *is* on Wednesday (cf. The concert’s on Wednesday)
   h. *is* it here?

The problem comes when we observe that not all such verbal forms reduce, for example the boldface items in (2d–h). These data, negative data concerning contexts where *is* does not reduce, are not available to children directly, and that is the poverty-of-stimulus problem: the stimulus appears to be too poor to determine all the properties of the mature system. Children hear some instances of the reduced *’s*, but somehow come to know much more, namely that *is* may be reduced in (2a) but not in the boldface contexts of (2d–h). 1 Furthermore,
Crain and Thornton (1998) show that children have this distinction as soon as they can be tested. If the external stimulus does not provide the relevant information, it must come from elsewhere, internally. So UG is supposed to provide the relevant information but the question then arises in what form. The poverty-of-stimulus problem shows that internal information is needed, but we must construct hypotheses about how the information is formulated, about the form of the information.

We return to this particular problem in a moment, but let us now introduce the notion of government, once a central part of linguistic theorizing but then eliminated on principled grounds.

1.3. Government

In its earliest formulation (Chomsky 1981), government was a sisterhood relation between a head and its complement (3a), or, in the case of an Inflection (I) element and Determiner (D),2 a relation between a head and its Specifier (3b), plus Exceptional Case Marking (3c). In (3) the boldface head governs the italicized phrase. Already there was a significant range of structures covered by the notion.

---

2. The analysis of determiners projecting to a Determiner Phrase is more recent than 1981 (Abney 1987; Horrocks and Stavrou 1987). They take a noun phrase as their complement and govern a Specifier. So a book has the structure of DP[N[book]] and Kim’s book about syntax has ‘s governing (and assigning Case to) Kim as well as NP[book about syntax]:

(i)

\[
\begin{array}{c}
\text{Spec} \\
\text{Kim} \\
\end{array}
\begin{array}{c}
\text{DP} \\
\end{array}
\begin{array}{c}
\text{D} \\
\end{array}
\begin{array}{c}
\text{NP} \\
\end{array}
\begin{array}{c}
\text{PP} \\
\end{array}
\begin{array}{c}
\text{book about syntax} \\
\end{array}
\end{array}
\]

I give the structure of ECM structures like I expect her to win as (3c), but there may be a CP above the IP indicated, depending on the analysis of expressions like Who do you expect her to see?
Government had a pervasive role in grammars: it limited the assignment of thematic roles and Cases, determined the binding domains for the Binding Theory (called Governing Categories in those days), and conditioned morphological operations and cliticization (Aoun and Lightfoot 1984).

Soon government was elaborated further. Lasnik and Saito (1984) reformulated the Empty Category Principle (ECP), which required traces to be “properly governed”, a sub-case of government. Proper government involved one of two forms, lexical government (4a) and antecedent government (4b).

(4) 
\( a \) properly governs \( b \) 
iff either (a) \( a \) governs \( b \) and \( a \) is a lexical category (\( V \), \( N \) or \( A \), but not \( I \) or \( P \)) (lexical government) 
or (b) \( a \) is a phrasal category \( c \)-commanding and locally coindexed with \( b \) (antecedent government)

Antecedent government was a curious extension to the theory, because the structures bore little resemblance to those of (3). However, Lasnik and Saito showed implicitly that their ECP solved poverty-of-stimulus problems (a striking feature of the generative literature is how little explicit discussion there is of poverty-of-stimulus problems). The information of (4) may not have been particularly elegant but it had some desirable empirical effects. In (5a) \( \text{who}_1 \) antecedent-governed the intermediate \( e_i \) but did not govern it lexically; similarly the lower \( \text{who} \) in (5d), when moved covertly at LF, antecedent-governs its trace, which is not lexically governed. (5b) was ill-formed because the deleted complementizer, \( e \), failed to be either lexically or antecedent-governed (cf 5c).

(5) 
a. \( \text{Who}_0 \) did you say \( [e_i e [\text{Kim saw } e_i]] \)?
b. \( \ast \text{Who}_0 \) did you say last week in Chicago \( [e_i e [\text{Kim saw } e_i]] \)?
c. \( \text{?Who}_0 \) did you say last week in Chicago \( [e_i e [\text{Kim saw } e_i]] \)?
d. who wondered [who left]?

Not pretty, but precise and adequate for the particular problems discussed.3

3. Hornstein and Lightfoot (1991) elaborate a number of the many complications entailed by Lasnik and Saito’s analysis. Lightfoot (1992) also details the complicating history of government.
Aoun et al. (1987) argued that antecedent government (4b), requiring a trace to be locally coindexed with a c-commanding phrasal category, looked identical to Principle A of the binding theory and that a binding theory generalized to include wh-traces as anaphors, as developed by Aoun (1985), and holding at the level of LF, made antecedent government superfluous. They adopted the model of (6), where government was simplified from (4) to be a condition on traces at PF, where they needed to be lexically governed (4a).

(6) D-structure
    S-structure
    PF
        ei is lexically governed
    LF
        Generalized Binding

Even that more constrained notion of government is minimalistically suspect, going beyond the head-complement and head-specifier relations of bare phrase structure, but it is more tractable than Lasnik and Saito’s version, and I will take it as a starting point. Lasnik and Saito’s (1984) notion of antecedent government can be subsumed under Aoun et al.’s (1987) notion of Generalized Binding, and it looks as if the effects of Principle A and the separate Control Theory of the GB models of the 1980s can be treated productively under a Minimalist analysis along the lines of Hornstein (2000).

Hornstein construes obligatory control (7) and binding (8) phenomena as resulting from movement, which he construes as copying and deletion (see below). So, if we dispense with any notion of D-structure, in (7a), we can treat the subject of *to win* as copied to the subject of *try* and then deleted. That accounts for the fact that what we used to think of as “PRO” must have an antecedent (7b), which must be local (7c) and c-commanding (7d), subject to a “sloppy” interpretation in ellipsis (7e), and does not permit split antecedents (7f). *John* ends up with two Θ-roles, one checked by *win* and one by *try*.

(7) a. John [tried [John to win]]
    b. PRO must have an antecedent
        Kimi tried [PROi to wash herselfi]
        *Kimj tried [PROj to wash oneselfj]
    c. The antecedent must be local
        Kayi thought [that Kimj failed [PROj/*i to wash herselfj]]
    d. The antecedent must c-command the PRO
        Kim’sj motherj tried [PROj/*i to win]
e. Obligatory control PRO requires a sloppy interpretation in ellipsis
   John expected [PRO to win] and Bill did, too.

f. Obligatory control PRO does not have split antecedents
   *Jim_t told Kim_j [PRO_{i,j} to leave together]

Anaphors work in a similar way (8). Here two cases are checked, accusative by himself and nominative by John. See Hornstein (2000) for the details and for the impressive empirical range of the proposals (and Culicover and Jackendoff 2001 and Landau 2001 for alternative analyses).

(8)

a. John_{i} [hurt himself_{j}]
b. The anaphor must have an antecedent
   *John arrested herself.
c. The antecedent must be local
   *John thinks that Mary arrested himself.
d. The antecedent must c-command the anaphor
   *John’s mother arrested himself.
e. The anaphor requires a sloppy reading under ellipsis
   John likes himself and Bill does, too.
f. Anaphors do not allow split antecedents
   *John told Mary about themselves.

1.4. Minimalism

Hornstein’s analysis of binding and control assumes Chomsky’s (1995) bottom-up approach to bare phrase structure. Elements drawn from the lexicon are merged into structures one-by-one. So the verb visit may be merged with the noun London to yield a VP (9a). Then the inflectional element will is merged with that VP to yield an IP (9b), and the (pro)noun you is merged with that IP to yield another IP (9c).

4. In this article I ignore the internal subject hypothesis, whereby a simple object-extraction like What did Sally buy? has a structure (i). I have nothing to say about the deletion of the lower what and Sally in such structures. Similarly, the lower did needs to delete, and, as noted below in the main text, that deletion falls under our cliticization analysis. The article shows that a number of instances of deletion are subject to poverty-of-stimulus problems suggesting a cliticization analysis. Other instances of apparent deletion are not subject to comparable poverty-of-stimulus problems and do not fall under the cliticization treatment. In some cases we understand why, and in some cases we do not, and so work remains to be done on why the cliticization analysis works in some cases and not in some others. The poverty-of-stimulus problems identified are real and require at least the information postulated here, but the analysis requires further elaboration, of course. I leave for future work some aspects of the question of why the analysis works in certain contexts but not in oth-
(9)  
  a. \( VP[\text{visit}_\text{N}London] \)  
  b. \( IP[\text{will} VP[\text{visit}_\text{N}London]] \)  
  c. \( IP[\text{you IP[\text{will} VP[\text{visit}_\text{N}London]]}] \)  

An expression like (10a) is built bottom-up in the same way. At a certain point the IP you did buy what is built. Buy merges with what to yield a VP; then did is merged with the VP to yield a IP, and then you is merged with the IP to yield another IP, as in (9). Then did is copied and merged, and what undergoes the same process. In each case, the copied element is later deleted in the original position from which it was copied, as indicated by the strike-through. Under this approach, there is no primitive operation of movement as such, but a copied element may be merged and then subsequently deleted. This is crucial for everything that follows.

(10)  
  a. What did you buy?  
  b. [\text{what}_i \ [\text{did}_i \ IP[\text{you did} \ buy \ what_i]]]  

I also adopt the proposals of Nunes (1995, 2004), namely that deletion of the copied element follows from the linearization of chains. Linearization is a PF operation that converts a syntactic structure into a sequence of X’ items in consonance with Kayne’s (1994) Linear Correspondence Axiom (LCA). The two \textit{whats} in (10b) are non-distinct and therefore must both precede and be preceded by \textit{buy}, etc. That is a contradiction (\textit{what} cannot both precede and be preceded by \textit{buy}), and it is the failure to yield a linear order that renders ers. Many possibilities come to mind for structures like (i), but I will not resolve them here.
the structure ill-formed, unless one of the what is deleted; and it must be the lower what that is deleted for reasons of the binding theory. Consequently, the fact that there are no chains at PF with more than one overtly realized link entails that the lower what in (10b) must be deleted. Nunes offers a rich analysis, noting exceptional cases where multiple wh-items are pronounced; there he shows that it is only intermediate wh-items that may be pronounced and, indeed, that they have clitic-like qualities (see Nunes 2004: 38–43 for discussion).

So far, so good, but now the question is: how is what is deleted? The question used to be why items could move from one position but not from others, but now the question is quite different, and we cannot simply recast the lexical government restriction on traces (4a); there is no movement and there are no traces. In fact, traces would be undefinable under the bare phrase structure approach: phrases are built with lexical items being merged into structures. Nonetheless, we still need to capture the effects of the lexical government restriction. So let us first review some of those effects and see how we might capture them with the economy and elegance that the Minimalist Program encourages.

2. Analysis

2.1. Head-complement relations

In fact, let us formulate things consistently as poverty-of-stimulus problems. A child might hear the sentences of (11a–c) pronounced with or without the complementizer that. Such experiences would license an operation (11d) whereby that may delete or be unpronounced; the operation is learnable on exposure to sentences like those corresponding to (11a–c). French, Dutch and German children have no such comparable experiences and hence no grounds to postulate a comparable device in their grammars; nothing like (11d) would be triggered.

(11)  a. Peter said [that/0 Kay had left already].
    b. The book [that/0 Kay wrote] arrived.
    c. It was obvious [that/0 Kay left].
    d. That → 0

However, a linguist may observe that the generalization (11d) breaks down at certain points and that may not be null in the contexts of (12). The crucial data here are negative data, data about what does not occur, and hence not available to children.\(^5\) UG must be playing some role.

---

5. This assumes, of course, that the child is not a conservative learner, postulating a grammar
In the early days of the GB model, the deletion of *that* was conditioned by lexical government, but if we don’t appeal to government, we can see that *that* is deleted if its clause is the complement of an overt, adjacent word. In (12a, b) the clause is the complement of *said* and *book* respectively, neither adjacent. In (12c), the clause is the complement of nothing. In (12d) it is the complement of *believes*, which is not adjacent, and in (12e) it is the complement of a verb that is not overt. In (12f) the lower complementizer may not be null, because its clause is not the complement of *said*.

The same condition holds for what we used to think of as traces of wh-movement. English-speaking children learn that wh-elements are displaced, i.e., pronounced in a position other than where they are understood, on hearing and understanding a sentence like *What did Jay see?* In (13a, b) the lowest *ei* is the complement of the adjacent verb, and the intermediate *ei* in (13b) is also adjacent to *say* and in its complement.

---

6. I assume here that restrictive relative clauses are complements to nouns, distinguishing (11b, 12b), and below (Note 15) I will take nonrestrictive relatives to be non-complements, i.e., adjuncts. This recalls analyses from the early days of generative grammar (for a survey, see Stockwell et al. 1973, Ch. 7), but need not be taken literally. We need a syntactic distinction between restrictive and nonrestrictive relative clauses, and restrictive relatives have some properties of complement clauses. What I am calling complement structures may be captured through Kayne’s (1994) raising analysis of restrictive relatives, where a restrictive relative headed by *that* is the complement of *D* and the ‘head’ raises out of the relative clause. For a good discussion of relative clauses and the problems they pose for modern theories of phrase structure, see Borsley (1997).

7. Bošković and Lasnik (2003), adopting ideas from Pesetsky (1991), an unpublished extension of Pesetsky (1995), treat null complementizers as resulting from a PF affixation operation. For them, affixation requires adjacency, but the data of (12) show that head-complement relations are crucially involved.
Assuming the Minimalist structures of (13a', b'), \textit{who} is deleted when it is the complement of an adjacent, overt word or in its complement. If that is the condition, it will predict, with no further learning, that (14a) is ill-formed, because the boldface \textit{who} is undeletable; it is in a clause which is the complement of \textit{apparent}, but it is not adjacent (henceforth boldface indicates an undeletable element). The lowest \textit{who} is the complement of the adjacent, overt \textit{seen}, hence deletable. Also, if \textit{yesterday in Chicago} were not present in (14a), then \textit{who would} be in the complement of and adjacent to the overt \textit{apparent}, hence deletable; this yields the well-formed (14b).

(14) a. *Who, was it apparent yesterday in Chicago CP[\textit{who} that [Kay had seen \textit{who}]]?  
   *Who was it apparent yesterday in Chicago that Kay had seen?  
   b. Who was it apparent that Kay had seen?

We solve the poverty-of-stimulus problem posed by (14a): in this way, children learn simply that \textit{wh}-items may be displaced and the UG condition causes the derivation of (14a) to crash with no further learning: Nunes’ requirements force deletion but deletion is not possible.

Now when considering some other contexts where lexical government was invoked in Aoun et al. (1987), we see more evidence that items may be deleted if they are the complement or in the complement of an overt, adjacent word. So \textit{which man} is deletable in the leftmost conjunct of (15c) (the complement of the adjacent \textit{introduce}) but not the boldface \textit{which woman} in the rightmost conjunct (the complement of a non-overt verb). So the corresponding sentence is ill-formed. Similarly in (15d, e, and g), the boldface element fails to meet the condition for deletion because the relevant verb is not overt. These structures involve \textit{wh}- movement (15c, d), readily learnable as noted above, heavy DP shift (15c, g), learnable on exposure to simple expressions like \textit{John gave to Ray his favorite racket}, and gapping (15c, d, e, g), learnable on exposure to something like (15b, f); the UG principle then solves the poverty-of-stimulus problems of (15c, d, e, g).

8. Notice that \textit{Which man did Jay introduce to Ray and Jim to Tim?}, analogous to the ill-formed (15c), is well-formed. Here only one \textit{wh}- phrase is overt and it moves across-the-board. One way of thinking of this is that across-the-board movement takes place on a three-dimensional structure before the two clauses are linearized; at that point \textit{which man} is the complement of \textit{introduce} (Williams 1978).

The data of (15) involve gapped verbs but pseudogapping sentences are equally bad with an auxiliary present (e.g., *Which man did Jay introduce to Ray and which woman did Jim to Tim? analogous to (15c), *Jay wondered what Kay gave to Ray and what did Jim to Tim analogous to (15d)). Pseudogapping structures are often analyzed very differently (see Lasnik 1999: Ch.7, who treats them as VP ellipsis), and I avoid them here. Translations of (15) are grammatical in languages with V-to-I copying, and we shall address that in Section 2.6.
(15)  a. Jay introduced Kay to Ray and Jim introduced Kim to Tim.
    b. Jay introduced Kay to Ray and Jim ve Kim to Tim.
    c. *Which man, did Jay introduce which man, to Ray and which woman, Jim ve which woman, to Tim?
       *Which man did Jay introduce to Ray and which woman Jim to Tim?
       *Jay wondered what Kay gave to Ray and what Jim to Tim.
    e. *Jay admired [his uncle from Paramus] greatly [his uncle from Paramus], but Jim ve [his uncle from New York] only moderately [his uncle from New York].
       *Jay admired greatly his uncle from Paramus but Jim only moderately his uncle from New York.
    f. Jay gave his favorite racket to Ray and Jim ve his favorite plant to Tim.
    g. *Jay gave [his favorite racket], to Ray [his favorite racket], and Jim ve [his favorite plant], to Tim [his favorite plant].
       *Jay gave to Ray his favorite racket and Jim to Tim his favorite plant.

The same condition explains why a complementizer may not be null if it occurs to the right of a gapped (non-overt) verb (16b); nor does one find a deleted copy in that same position (the boldface who in 16c).

(16)  a. Jay thought Kay hit Ray and Jim ve CP[that Kim hit Tim].
    b. *Jay thought Kay hit Ray and Jim ve CP[0 Kim hit Tim].
    c. *Who, did Jay think Kay hit who, and who, Jim ve CP[who (that) [Kim hit who]]?
       *Who did Jay think Kay hit and who Jim (that) Kim hit?

So children exposed to some form of English have plenty of evidence that a that complementizer is deletable (11d), that wh-phrases may be displaced (copied), and that heavy DPs may be copied to the end of a clause (15e, g), but they also know without evidence that complementizers and copies may not be deleted unless they are the complement or in the complement of an adjacent, overt word. And the data of (11–16) suggest that this is the information that UG needs to provide, and head-complement relations are crucial. The convergence of that information with the grammar-specific devices which delete a that complementizer and allow a wh-phrase and a heavy DP to be copied, yields the distinctions we have noted and solves the poverty-of-stimulus problems. However, again, postulating that information at the level of UG leaves open the form that it must take, and we turn now to that matter.
2.2. Implementation

We know that elements may cliticize to the left and become an indissoluble part of their host. That happens with the reduced is that we discussed earlier in (2).9 When is reduces, its pronunciation is determined by the last segment of the word to which it attaches (17a): voiceless if the last segment is voiceless, voiced if the last segment is voiced, and syllabic if the last segment is a sibilant or affricate. Precisely the same is true of the plural marker (17b), the possessive (17c), and the third person singular ending on a verb (17d).

(17) a. Pat’s happy, Doug’s happy and Alice’s here.
b. Cats, dogs and chalices.
c. Pat’s dog, Doug’s cat and Alice’s crocodile.
d. Commits, digs and misses.

Children understand Pat’s happy as ‘Pat is happy,’ Pat being the subject of the phrase ‘is happy.’ However, is is pronounced indissolubly with Pat, and children parse what they hear as (18a), i.e. with reduced is attached to the noun, with normal pronunciation applying. (18a) expresses a cue (18b), a piece of structure that serves to determine the shape of the emerging grammar, showing particularly that elements may be cliticized (Lightfoot 1999, 2006). So from hearing and understanding an expression like Pat’s happy, children learn that is may be reduced and absorbed into the preceding word.

(18) a. NPat+’s
b. noun+clitic

Similar factors affect the pronunciation of to in reduced wanna: the to cliticizes to want, to form an indissoluble word, but here the cliticization affects an element of the IP that is the complement of an adjacent want, reminiscent of the phenomena discussed in the last section. In (19b, c) to does not meet this condition and is not reducible: speakers of most forms of English would not say *Who do you wanna visit Rio? or *I don’t wanna win games to be our only goal (with the exception of speakers of the “liberal” dialects discussed by Postal and Pullum 1982, who would say the former). In (19b) the intervening wh-word blocks adjacency of want and to and in (19c) the lower IP is not the complement of the adjacent want.10

---

9. The phonological cliticization affecting is in (17a) Pat’s happy, where is is not the complement of Pat, is a different kind of cliticization from the one subsuming the deletions discussed above. But see the discussion of V-to-I movement below (Section 2.6).

10. In (19b) the lower who deletes but that does not allow to to incorporate into want. Compare the derivation of (19a), which has an abstract structure: Where do you do want [where [you to go]]: here the lowest where incorporates into go and the intermediate where and
Where do you want to go?
Who do you want to visit Rio?
I don’t want to win games to be our only goal.

So our complement condition affects the formation of wanna reductions and we will analogize the deletion phenomena discussed so far to that form of cliticization, recognizing that there are many forms of cliticization (we shall discuss another form in Section 2.6).

If we draw (17) together with (20), we now find something interesting: copies do not delete if they are to the right of a cliticized verb. In (20), the copied wh-phrases may be deleted if is is in its full form, but not if it is reduced; the corresponding sentences with ’s do not occur.11

This suggests again that a deleted copy is incorporated into the element of which it is the complement. In (20), if is cliticizes on to the subject noun and becomes part of that noun, it no longer heads a phrase of which what/where is the complement and no incorporation is possible, hence no deletion if deletion is incorporation.

That idea enables us to capture another subtle and interesting distinction. The sentence (21a) is ambiguous: it may mean that Mary is dancing in New York or just that she is in New York, but working on Wall Street and not dancing. The former interpretation has a structure with a deleted verb (21b) (but perhaps an instance of VP ellipsis; see below). If deleted verbs are incorporated, there must be an appropriate host. There is an appropriate host in (21b), where the deleted verb incorporates into a full verb, but not in (21c), where the boldface verb has no appropriate host; dancing isn’t the complement of Mary’s, therefore no incorporation is possible and no licensing of a deleted verb. Consequently

you successively into want and this does permit to to cliticize on to want to yield Where do you wanna go? This long-noted distinction (Lightfoot 1976) has been keyed to Case-marking (Jaeggli 1980), in a not altogether satisfying way, and I have nothing new to suggest here. Pullum (1997) offers a different, “morpholexical” analysis for to-contraction, treating wanna etc. as a derived lexeme subcategorized for ‘an immediately following bare infinitival subject-less complement’ (1997: 95), which would also make the needed distinction.

11. I adopt the analysis of Chomsky (1977), under which the gap in comparatives like (20a) is a wh- phrase.
Minimizing government

(21d) unambiguously means that Mary is in New York, occupation unspecified, because there is no empty, understood verb. Again, it is inconceivable that children learn such distinctions purely on the basis of external evidence.

(21)

a. Max is dancing in London and Mary is in New York.
b. Max is dancing in London and Mary is dancing in New York.
c. *Max is dancing in London and Mary’s dancing in New York.
d. Max is dancing in London and Mary’s in New York.

Further distinctions follow from this analysis. English allows ellipsed VPs and children have plenty of evidence to that effect (22a). In (22a) the empty VP is the complement of did. In fact, there must be an overt head to license the null VP, suggesting that empty VPs occur only where they are incorporated into a host like null complementizers and deleted copies.12 (22b) is not an instance of VP ellipsis, but rather of Stripping (Sag 1976). (22c) is ill-formed because part of the VP remains, for Naples, and there is no null VP. In the ungrammatical structures of (22d, e) the null VP is separated from its potential host, hence failure to incorporate. An ellipsed VP may occur in a subordinate clause (22f), to the left of its antecedent (22g), in a separate sentence from its antecedent (22h), or within a complex DP (22i), when its antecedent is contained in a relative clause (22j), or even without any overt antecedent (22k), but it must be the complement of an overt head immediately to the left.

(22)

a. Max left on Wednesday but Mary did VPe as well.
b. Max introduced Fred to Phil, and Reuben as well.
c. *Max left for Rio but Mary didn’t VPe for Naples.
d. They denied reading it, although they all had VPe.

vs.

t. They denied reading it, although they had all VPe.

e. They denied reading it, although they often/certainly had VPe.

vs.

t. They denied reading it, although they had often/certainly VPe.

12. Lobeck (1990, 1995) and Zagona (1988) offer good discussions of VP ellipsis and invoke the Empty Category Principle (4). As noted below in the main text, VP ellipsis, unlike gapping and sluicing, seems to be unusual crosslinguistically. Potsdam (1997) observes the distinction between (ia) and (ib), where not licenses an ellipsed VP, and a referee notes the distinction between (iia) and (iib), where to licenses an ellipsed VP.

(i) a. *It is possible to eat this fruit, and we recommend that you VPe.
b. It is possible to eat this fruit, and we recommend that you not VPe.

(ii) a. *Kim began signing a song before Jim began VPe.
b. Kim began to sing a song before Jim began to VPe.
David Lightfoot

f. Max left for Rio, although Mary didn’t vpe.
g. Although Max couldn’t vpe, Mary was able to leave for Rio.
h. Susan went to Rio.
   Yes, but Jane didn’t vpe.
i. The man who speaks French knows Dp[the woman who doesn’t vpe].
j. People who appear to support mavericks generally don’t vpe.
k. Don’t vpe!

This suggests that, like null complementizers, deleted copies and the deleted verb of (21b), an ellipsed VP incorporates to the left, to an adjacent, overt item of which it is the complement (23). Where that is not possible, there is no ellipsis, as indicated by the contrasts of (22d, e).

(23) Max could visit Rio and Susan INFL could + vpe.

That, in turn, now explains the non-occurrence of (24a), noted by Zagona (1988): the ellipsed VP needs an appropriate host, a full phonological word, of which it is the complement, as in (24b); in (24a) has has become part of the noun John and no longer heads a phrase of which vpe is the complement.

(24) a. *I haven’t seen that movie, but John’s vpe.
    b. I haven’t seen that movie, but John has + vpe.

Here I am dealing with the distribution of ellipsed VPs, not their interpretation. Their interpretation raises many interesting issues which are independent of this account.

2.3. Incorporation

So copies are deleted in PF in order to satisfy linearization requirements and our analysis takes deletion to be an instance of incorporation, which allows the analysis to generalize to certain null elements that are not copies. In (25a) the complement incorporates into the adjacent see, and in (25b) Jay is in the complement of expected, adjacent to it, and accordingly incorporates into it.

(25) a. Who did Jay see?
    b. Jay was expected [Jay to win]

13. A referee notes the grammaticality of (ia), which might reflect the fact that verbs may move with their clitics (ib).

(i) a. … but why couldn’t you vpe?
    b. Jean l’embrasse souvent e.
The analysis appeals to head-complement relations and adjacency.

This analysis captures many other distinctions. For example, English-speakers’ grammars typically have an operation whereby a “heavy” DP is displaced to the right, as we have noted. Under our approach, that now means copying and reducing the copied element to silence by absorbing it clitic-like into a host. In (26a) the copied element is the complement of introduced, hence incorporated successfully; in (26b) it is in the complement of the adjacent expect; but in (26c) the element that needs to delete is neither the complement nor contained in the complement of anything, and the derivation is ill-formed and crashes.

(26) a. I introduced [all the students from Brazil]i to Mary [all the students from Brazil]i
   I introduced to Mary all the students from Brazil.

b. I expect [[all the students from Brazil]i to be at the party][all the students from Brazil]i
   I expect to be at the party all the students from Brazil.

c. *[all the students from Brazil]i are unhappy] [all the students from Brazil]i
   *Are unhappy all the students from Brazil.

Our UG principle, that deletion of this kind is incorporation, solves the poverty-of-stimulus problem of (26c): children simply learn that heavy DPs may be copied to the right and the UG condition accounts for the non-occurrence of (26c) with no further learning or experience needed.

Consider now “objective genitives” like (27). An expression like Jay’s picture is three-ways ambiguous: Jay may be the owner of the picture, the painter, or the person portrayed. The latter reading is the so-called objective genitive and is usually analyzed as in (27), where Jay is copied from the “object” position to the Specifier of the DP. The operation is specific to grammars of English-speakers and does not occur in French, for example. This much is learnable: children hear expressions like Jay’s picture in contexts where it is clear that Jay is pictured.

(27) DP[Jayi’s NP[picture Jayi]]

A curious thing is that comparable expressions like The picture of Jay’s, The picture is Jay’s, and The picture which is Jay’s (28) show only a two-way ambiguity, where Jay may be the owner or the painter but not the person portrayed. This is yet another poverty-of-stimulus problem, because it is inconceivable that children are systematically supplied with evidence that the objective interpretation is not available in these cases. We now have an explanation for this, because the structure of these expressions would need to be as in (28).
(28)  a. *the picture of DP[Jay’s NP[e Jay]]  
   (The picture of Jay’s)  
 b. *the picture is DP[Jay’s NP[e Jay]]  
   (The picture is Jay’s)  
 c. *the picture which is DP[Jay’s NP[e Jay]]  
   (The picture which is Jay’s)  

A preposition like of is always followed by a DP, a possessive like Jay’s occurs only as the Specifier and head of a DP (see Note 2), and Ds always have a NP complement, even if the noun is empty, as it is here (where it is understood as ‘picture’). Now we can see why the structures are ill-formed: the copied Jay has no host to incorporate into, hence boldface and the derivation crashes. Jay is the complement of the adjacent noun, but that noun is not overt, hence not a viable host.

(29) reflects another distinction provided by our account. (29a) is well-formed and involves no deletion of a copied element, whereas (29b) involves two instances of DP-copying and deletion (to yield the passive constructions). The leftmost instance is well-formed, because the copied Jay is in the complement of the adjacent known and therefore deletes; however, in the rightmost conjunct, the copied he has no overt host to incorporate into and therefore cannot delete, hence boldface and leading the derivation to crash.

(29)  a. It is known that Jay left but it isn’t ve that he went to the movies.  
 b. *Jay i is known [Jayi to have left] but he i isn’t ve [he i to have gone to the movies]  
   *Jay is known to have left but he isn’t to have gone to the movies.

And there is more: it is well-known that an expression like They were too angry to hold the meeting is ambiguous, meaning either that they were too angry for them or for some unspecified person to organize the meeting; the ambiguity lies in who was going to organize the meeting (Chomsky 1986: 33). The former reading has the structure of (30a), where they is copied and deleted; the CP is the complement of angry and they is in that complement and adjacent to angry, hence incorporated. The other reading has arbitrary PRO as the subject of hold (30b): nothing is copied and that would not be possible because the clause is an adjunct to angry, not a complement (represented here by boldface brackets).

(30)  a. They, were too angry CP[they, to hold the meeting]  
   They were too angry to hold the meeting.  
 b. They were too angry CP[PROarb to hold the meeting]  
 c. Which meeting, were theyj too angry CP[which meeting, [theyj to hold which meeting,]]
Which meeting were they too angry to hold? 
d. *Which meeting were they too angry to hold [which meeting [PROarb]]

However, the corresponding question Which meeting were they too angry to hold? is unambiguous and has only the anaphoric reading under which they are going to hold the meeting (30c). It lacks the meaning of an arbitrary subject for hold and (30d) is ill-formed. Where the reading is anaphoric (30c), the clause is the complement of angry and therefore which meeting in that complement can be incorporated into angry and thus deleted. Likewise, the copied they deletes successively in (30a) (see the next section). However, where the subject of hold is the non-obligatory control, arbitrary PROarb, the clause is an adjunct to angry and therefore the boldface which meeting cannot be incorporated and is undeletable.

This treatment of extraction licensed by incorporation requirements on elements at the front of a complement suggests an explanation for the observation by Chung (1983) that the effects of the Sentential Subject Constraint (Ross 1967) are systematically absent from VSO languages. This would be expected under the approach adopted here if the relevant structure has the inflected verb fronted as in (31).

(31) \[I+V_{CP}[I+V_{IP}[\ldots I+V_{IP}[\ldots]]]]

2.4. Clitics versus affixes

So far we have been talking about deletion sites as involving incorporation into a host, as if the deleted item were some kind of clitic. Indeed, it is profitable to view the incorporated items as clitics.

Zwicky and Pullum (1983) distinguish between clitics and affixes, and their distinction permits some further understanding. Specifically, they argue that the English reduced negative n’t is an affix: so in our terms isn’t, for example, is formed in the lexicon and merged directly into syntactic structure. That distinguishes between (32b), where isn’t is merged with here to form a constituent, and the ill-formed (32c).

(32) a. John’s not here.
b. John isn’t here.
c. *John’sn’t here.

Two of Zwicky and Pullum’s criteria for their distinction are given in (33); (F) says that affixes may not attach to material already containing clitics, hence the non-occurrence of (32c).
E. Syntactic rules can affect affixed words, but cannot affect clitic groups.

F. Clitics can attach to material already containing clitics, but affixes cannot.

This allows us to distinguish between the structures of (34): (E) allows a syntactic copying operation (what we used to think of as displacement or movement) to affect couldn’t, an affixed form, but not could’ve, where ’ve is cliticized on to could (hence the grammaticality of the corresponding Could Kim’ve seen that? but not *Could Kimn’t see that?).

(34) a. Couldn’t Kim see that?
   b. *Could’ve Kim seen that?

If n’t is an affix, then phonologically reduced verbs (’s, ’ve, wanna, etc.), ellipsed VPs, null complementizers, gapped verbs, and deleted copies are clitics. If clitics may attach to material already containing clitics (33F), we allow (35a–d) but not (35e), which has an affix attached to could’ve, in violation of (33F).14

(35) a. Kim visited NY and Jim could’ve VPe.
   b. Kim visited NY but Jim couldn’t VPe.
   c. Kim visited NY but Jim couldn’t’ve VPe.
   d. I’d’ve visited NY.
   e. *Jim could’ve n’t seen it.

There is a vast literature on clitics and many distinctions are drawn; indeed Arnold Zwicky has argued in his later work that there are no clitics (introduction to Nevis et al. 1994). I have drawn selectively from that literature in arguing that the deletion sites discussed so far are clitics along the lines of the to adjoined to want in wanna. That helps us to understand, for example, why successive incorporation is sometimes possible. Below I will suggest that reduced is is a different kind of clitic, which is the analog for V-to-I movement

14. Notice that (35a) is well-formed but (24a) is not. In (35a) the null VP is the complement of the adjacent could’ve but in (24a), it is not the complement of John’s. This also accounts for the distinction in (i) below.

(i) a. Kim canceled her subscription, and I would’ve VPe, too.
   b. *Kim canceled her subscription, and I’d’ve VPe, too.

A null VP following the reduced ’ve in (ia) is the complement of would’ve, but a null VP following the reduced ’ve in (ib) is not the complement of I’d’ve, where the reduced auxiliaries have been cliticized to the subject DP and are no longer in the position of I with a VP complement.
and perhaps for gapping. However, it may be that the incorporation analysis of deletion (i.e. old-style traces and ellipsis sites) is correct but that the incorporated elements are not clitics; the claims are logically distinct.

2.5. Subjects

Thinking of the deletion of copied phrases as cliticization enables us to understand old puzzles about the fixed subject condition (Bresnan 1972) and the “that-trace” effect of the 1970s (Chomsky and Lasnik 1977), later subsumed under the agreement relations of Rizzi (1990). It also enables us to learn more about the cliticization operation. In general, subjects resist displacement; when they are copied into a displaced position, odd things happen.

In (36a) the complementizer contained in the clausal complement to thought cliticizes to thought straightforwardly and may be deleted (unpronounced). In (36b) the lowest who cliticizes to saw, the intermediate who to think and the complementizer cliticizes to think+who, again successively like the clitics of (35) and straightforwardly.

(36) a. I thought [that/0 Ray saw Fay]
b. Who0, did you think [who, that/0 Ray saw who]?
c. *Who0, did you think [who, that who saw who]?
   *Who do you think that saw Fay?

However, in (36c) the intermediate who cliticizes to think, but the lowest (boldface) who apparently cannot cliticize to that, presumably because that does not take complements in the usual sense (despite the name “complementizer,” the following clause does not “complete” the meaning of that in the way that Fay completes the meaning of saw) and is not an appropriate host. Likewise for equivalent complementizers in other languages, as we shall see.

Similarly in (37a) the boldface who may not cliticize to the complementizer how, because how has no complement. Hence the difference with (37b), where each copied element is deleted in the appropriate way, and the sentence is grammatical if not completely felicitous.

(37) a. *Who0, do you wonder [who, how [who, solved the problem]]
   *Who do you wonder how solved the problem?
   b. What0, do you wonder [what, how [John solved what]]
   What did you wonder how John solved?

Interestingly, Jim McCloskey points out that a complementizer (C) can be a host for clitics in some languages, notably German, where the positioning of weak pronoun clitics with respect to C grounds one of the most famous
and most beautiful arguments for the analysis of V2 structures in terms of I-to-C raising’ (personal communication). He notes that English that is not a “proper governor,” nor is a raised auxiliary a proper governor; Luigi Rizzi (1990) showed that fronted auxiliaries do not save heavy DP shift from subject positions (38).

(38) *Will ___ apply for this position, all those PhD students who aren’t yet employed.

However, Rizzi also argues that in German an inflected verb that has raised to C is a proper governor, unlike in English. So C, both a that-like complementizer and a fronted inflected verb, may vary from language to language in its ability to host a deletion site.

Not only do complementizers like that and how not generally host clitics, but neither do prepositions. This explains the well-known observation that generally prepositions do not licence movement sites (French *Qui as-tu parlé avec?, Dutch *Wie heb je met gesproken?, *Who have you spoken with?). In English, prepositions may be stranded but only where they are themselves re-analyzed into a complex verb, as in (39a) (see Hornstein and Weinberg 1981 for discussion of the re-analysis operation); compare the ill-formed (39b,c) where the PP is not the complement of an adjacent verb and consequently may not host the deleted copy.

(39) a. Who did you talk to? 
   *Who did you talk at the meeting to who?
   *Who did you talk at the meeting to?
   *What did you sleep during what?
   *What did you sleep during?

Let’s return to the topic of subjects. We can also understand the distinction between (40a,b). In (40a) the boldface who may not be cliticized to the left-adjacent what because what (in the Spec of CP) does not take a complement. There is no comparable problem in (40b).

(40) a. *This is the student who I wonder [what] who bought
   *This is the student who I wonder what bought.
   This is the sweater which I wonder who bought.

I adapt the analysis of Rizzi (1990) for structures like (41); it is well-formed because the lowest who agrees with the higher who (by co-indexing) and, there-
fore, under this approach, cliticizes to it. The higher who is in the complement of think and cliticizes to the adjacent verb.\textsuperscript{15}

(41) Who did you think CP[\textcolor{red}{whoi} saw Fay]?
    \textit{Who did you think saw Fay?}

In some languages that kind of complementizer-subject agreement is marked overtly. So in French if a subject qui is deleted, that is possible only if there is an agreeing form of the complementizer, qui instead of the usual que (42d). This is the phenomenon treated by the old, mysterious que-to-qui rule.

(42) a. Je crois CP[que Marie a vu Jean].
    ‘I believe that Mary has seen John.’

b. Qui, crois-tu CP[\textcolor{red}{qui} que Marie a \textcolor{red}{qui} vu]?
   Qui crois-tu que Marie a vu?
   ‘Who do you believe Marie has seen?’

c. *Qui, crois-tu CP[\textcolor{red}{qui} que qui a vu Jean]?
   *Qui crois-tu qui a vu Jean?
   d. Qui, crois-tu CP[qui, qui a vu Jean]?
   Qui crois-tu qui a vu Jean?
   ‘Who do you believe has seen John?’

Rizzi (1990) notes something similar happens in West Flemish, where a deleted subject is licensed by a special agreeing form of the complementizer die instead of the usual da (43b).

\footnotesize

15. The analysis predicts correctly that relative pronouns and complementizers are deletable if they introduce restrictive relative clauses, if restrictive relatives are complements of the head noun (as in (i) below), but not in non-restrictive relatives, if they are non-complement, adjunct clauses (ii). In (i) which cliticizes on to book and then that is free to do likewise. As observed in note 6, this used to be a standard analysis (Stockwell et al. 1973); it is no longer so but can be reconstructed through Kayne’s (1994) raising analysis.

(i) This is the book (which) (that) Bill read.
(ii) I met Bill, *(who) Jill had introduced.

As with (ii), in general our analysis leads us to expect a \textbf{wh}-item at the front of an adjunct clause not to be deletable (hence *What did you go to Rome without reading?), and I leave as an unresolved problem why a \textbf{wh}-item appears to be deletable at the front of an adjunct clause particularly in a parasitic gap construction, in the position of \textbf{x} in (iii) \textit{(This is the book which I filed without reading)}.

(iii) This is the book (which) I filed CP[\textcolor{red}{x} without reading].
Other grammars circumvent the difficulty of deleting subjects to the right of a complementizer by using resumptive pronouns. So Swedish requires a resumptive pronoun (det) in the subject position when the complementizer hur is present (44a) (Engdahl 1985); if it were a matter of deleting a copied vilket ord, there would be no host and the derivation would crash.

Similarly, the West African language Vata adopts the same strategy, but here even for local movement in simple, unembedded clauses. Again we see the familiar subject-object asymmetry: an extracted subject is manifested by a resumptive pronoun (45a), but not a deleted object (45b).

2.6. Verb movement

Interestingly, grammars with verb movement behave differently in certain ways (Note 8). Notice first that sentences like (15c), repeated here as (46a), are well-
formed in Italian (46b). Since verbs typically move to the higher Inflection (I) position in such grammars, that suggests that what we used to view as an indexed empty verb (a trace) counts as overt (see also Merchant 2002 for the same conclusion) and may license a deleted wh-phrase (as opposed to the unmoved, gapped verbs in (15c,d,e,g)); under the present view, this means that a deleted copy may license a clitic like a deleted wh-phrase. So French *Que lit elle?* ‘What is she reading?’ has the structure of (46c); the lower que cliticizes to lit, which is in turn deleted after copying. And a verb-second (and object-verb) grammar like that of Dutch speakers has structures like (46d), where in *Wie doodde je?* ‘Who did you kill?’ the deleted wie cliticizes to doodde, which is subsequently deleted.

(46)  
   a. *Which man did Jay introduce which woman to Ray and which woman to Tim?*
   b. Quale ragazzo Gianni ha presentato a Maria e quale ragazza tu a Paolo?
      ‘Which boy has Gianni introduced to Maria and which girl have you to Paolo?’
   c. que i litj [elle i quei]
   d. wie i dooddej [je wie doodde]

The copied verb is deleted in the cases of (46b–d). Similarly in (47a). However, that deletion does not fall under the wanna-style cliticization analysis that we have entertained: lit is not the complement or in the complement of elle in (46c) and has is not (in) the complement of Jay in (47a). Notice, however, that has may cliticize on to Jay in an expression like Jay’s read the book, and as a deleted copy in (47a), even though it is not (in) the complement of Jay. So one might distinguish two kinds of cliticization, one found in wanna forms and the other in forms like Jay’s read the book, each providing a model for deletion: the wanna cliticization provides a model for the deletion of phrasal copies, complementizers, and ellipsed VPs, and reduced has provides a model for deletion of raised verbs. 16 If these deletions result from a phonological operation, as

16. Perhaps the gapped verb in (i) is cliticized in this second fashion, attaching to the subject DP.

(i) Max speaks French and Mary *ve* German.

Gapped verbs have a very different distribution from ellipsed VP, so they do not cliticize in the way that we have analyzed ellipsed VPs; compare (ii) with (22).

(ii)  
   a. *Max speaks French, although Mary *ve* German.
   b. *Jim said that Max speaks French and Kim said that Mary *ve* German.
   c. *Max *ve* French and Mary speaks German.
   d. *The man who speaks French knows to [the woman who *ve* German].
   e. *Max drove to New York and Susan did *ve* to Chicago.
suggested by Chomsky (1995), analogous to the phonological cliticization in Pat’s happy (see Note 9), then (47b and c) would not be generated (the deletion of have in (47b) and be in (47c) yield ill-formed structures) because they involve morphological discrepancies between the “copied” forms, so there can be no copying involved from the position indicated, hence the Head Movement Condition effects.

(47)  
(a. _chasio.[Jay hasi read the book]
   Has Jay read the book?
(b. _chasio.[Jay could havei read the book]
   *Has Jay could read the book?
(c. _cisoi.[Jay have bei-en singing]
   *Is Jay have singing?

3. Conclusion

Innateness claims depend on detailed distinctions between what a child may be supposed to garner from her environment and what needs to be specified internally, or innately. I have argued that English-speakers learn that certain verbs may be phonologically reduced, complementizers may be null, that wh-phrases may be displaced (pronounced in positions other than where they are understood), verbs may be gapped, that heavy DPs may be displaced to the right, VPs may be ellipsed, possessive noun phrases may have objective interpretations. These seven properties are readily learnable from the linguistic environment, and we can point to plausible primary linguistic data (PLD). Such data that all English-speaking children hear include sentences like Kim’s happy (2), manifesting reduction; Peter said Kay had left already (11a), exhibiting a null complementizer; Who did Jay see? (13a) with a displaced wh-phrase; Jay introduced Kay to Ray and Jim Kim to Tim (15b), an example of gapping; Jay gave to Ray his favorite racket (15g), heavy DP shift; Max could visit Rio and Susan could, too (23), an ellipsed VP; and Jay’s picture, meaning ‘picture of Jay’ (27).

The way to think of this, I believe, is that children identify certain structures, cues; some of these structures vary from grammar to grammar and they constitute the parameters of variation. Consider the object-verb/verb-object parameter. Under the cue-based approach, children find either _vp[DP V] or _vp[V DP] structures, very specific information. Children use structures or lose them: a child who encounters object-verb _vp[DP V] structures in the primary data, loses _vp[V DP] structures, which atrophy. Notice that children are reacting to abstract structures, elements of grammar, which are required to understand expressions that they hear (for extensive discussion of cue-based acquisition, see
As far as innateness is concerned, i.e., UG, we began by examining earlier notions of government, postulated to solve poverty-of-stimulus problems but enriched over time and becoming quite complex. There is nothing necessarily wrong with complexity, but notions of economy and elegance have proven successful in other areas, certainly for physicists; therefore, it is worth asking if things can be done differently. We asked whether we could re-think government in Minimalist terms and come up with different analyses, conforming to Minimalist conceptions of bare phrase structure and hypothesizing only elements that are motivated at one of the two interface levels, phonological and logical forms. I have argued that an empty element (a deleted phrasal copy, null complementizer, ellipsed VP, the ellipsed *dancing* in 21b, c) is incorporated or cliticized into an adjacent coindexed phonological head (N, V, I) of which it is (in) the complement. The old ECP effects are Linear Correspondence Axiom (LCA) effects and government is eliminated.

One simple idea at the level of UG interacts with seven grammar-specific devices, all demonstrably learnable, and that interaction yields a complex range of phenomena. This involves carving up the grammatical world differently. The proof of the Minimalist pudding lies in whether it permits productive analyses with good empirical coverage and I offer this analysis in that spirit.

The evidence I have offered has been drawn largely from English and a more extensive treatment will explore effects in other languages. However, if one follows the logic of the poverty of the stimulus argument, much can be learned from detailed analysis of a single language, if one attends carefully to issues of learnability.

We seek a single object: the genetic properties of the language organ (Anderson and Lightfoot 2002). They permit language acquisition to take place in the way that it does and that means that we must examine language variation along the lines of Baker (2001); that yields a wealth of empirical considerations. What we postulate must solve the poverty-of-stimulus problems that we identify and solve them for all languages as well. We also want our ideas to be as elegant and economical as is feasible. In addition, the grammars that our theory of UG permits must meet other demands.

---

17. To illustrate the abstractness of the structures, consider the fact that Dutch has $\verb[DP V]$ structures (i.e., is an object-verb language). If children are also degree-0 learners (Lightfoot 1989), sensitive only to simple structures, then sentences like *Hoe belt hij Peter op?* ‘How does he call Peter up?’ expresses the $\verb[DP V]$ cue: the analysis is *hoe belt hij* $\verb[VP]([Peter op])*, where the separable particle *op* marks the position of the deleted verb (which children do not hear). Cues are identified only in unambiguous structures (Lightfoot 1999, 2006).
To take just one example, they must allow speech comprehension to take place in the way that it does. That means that considerations of parsing might drive proposals. That hasn’t happened much yet, but there is no principled reason why not, and the situation might change. Similarly, evidence drawn from brain imaging or even from brain damage might suggest grammatical properties. In fact, the proposals here look promising for studies of on-line parsing. When a person hears a displaced element, say a *wh*-phrase at the beginning of an expression, she needs to search for the deletion site, the position in which it needs to be understood. The ideas developed here restrict the places that she can look. I cannot examine the consequences of this at this time, but they look potentially useful for parsing studies. In fact, perhaps somebody could have arrived at such proposals from the study of on-line parsing.

One uses what looks like the best evidence available at any given time, but that will vary as research progresses, and consequently the form of our innate-ness claims will vary. There are many basic requirements that our hypotheses must meet, and there is no shortage of empirical constraints, and therefore many angles one may take on what we aim for. In this article I have taken one angle and progressed beyond where government took us: to delete an element is to cliticize it. This is certainly not the end of any story, but a reasonable way to proceed and an improvement on earlier accounts.

References

Minimizing government


